Chapter 3

The basic OLG model: Diamond

There exists two main analytical frameworks for analyzing the basic intertemporal choice, consumption versus saving, and the dynamic implications for the economy as a whole of this choice: the *overlapping-generations* (OLG) approach and the *representative agent* approach. In the first type of models the focus is on (a) the interaction between different generations alive at the same time, and (b) the never-ending entrance of new generations and thereby new decision makers. In contrast, the second type of models views the household sector as consisting of a finite number of infinitely-lived dynasties all alike.[1] This approach, also called the Ramsey approach, will be described in Chapter 10.

The OLG approach is the topic of the present and the next chapter. The *life-cycle* aspect of human behavior is at the center of attention. During lifetime one's educational level, working experience, working capacity, income, and needs change and this is reflected in the individual labor supply and saving behavior. Different OLG models focusing on the aggregate implications of the behavior of coexisting individuals at different stages in their life have shown their usefulness for analysis of public debt, taxation of capital income, financing of social security (pensions), design of educational systems, non-neutrality of money, the possibility of speculative bubbles, etc.

We will here present what is considered the basic macroeconomic OLG model, put forward by the American economist and Nobel Prize laureate Peter A. Diamond (1940-).[2] Saving and dissaving are core elements in life-cycle behavior and hence also in Diamond's model. Before going into the specifics of the model, we will give a brief outline of possible motives for saving.

[1] The interpretation is that the parents take the utility of their descendants fully into account by bequeathing them and so an intergenerational chain arises.

[2] Diamond (1965).

3.1 General motives for saving

Before going into the specifics of Diamond's model, let us briefly consider what may in general motivate people to save:

(a) The *consumption-smoothing motive for saving.* Individuals go through a life cycle where earnings typically have a hump-shaped time pattern; by saving and dissaving the individual then attempts to obtain the desired smoothing of consumption across lifetime. This is the essence of the *life-cycle saving hypothesis* put forward by Nobel laureate Franco Modigliani (1918-2003) and associates in the 1950s. This hypothesis states that consumers plan their saving and dissaving in accordance with anticipated variations in income and needs over lifetime. Because needs vary less over lifetime than income, the time profile of saving tends to be hump-shaped with some dissaving early in life (for instance if studying), positive saving during the years of peak earnings and then dissaving after retirement.

(b) The *precautionary motive for saving.* Income as well as needs may vary due to conditions of *uncertainty*: sudden unemployment, illness, or other kinds of bad luck. By saving, the individual can obtain a buffer against such unwelcome events.

Horioka and Watanabe (1997) find that empirically, the saving motives (a) and (b) are of dominant importance (Japanese data). Yet other motives include:

(c) Saving enables the purchase of *durable consumption goods* and owner-occupied housing as well as repayment of debt.

(d) Saving may be motivated by the *desire to leave bequests* to heirs.

(e) Saving may simply be motivated by the fact that financial wealth may lead to *social prestige* and economic as well as political *power.*

Diamond's OLG model aims at simplicity and concentrates on motive (a). In fact only one aspect of motive (a) is considered, namely the saving for retirement. People live for two periods only, as "young" they work full-time and as "old" they retire and live by their savings.

Now to the details.

3.2 The model framework

The flow of time is divided into successive periods of equal length, taken as the time unit. Given the two-period lifetime of (adult) individuals, the period length is understood to be very long, around, say, 30 years. The main assumptions are:

1. The number of young people in period t, denoted L_t, changes over time according to $L_t = L_0(1+n)^t$, $t = 0, 1, 2, ...$, where n is a constant, $n > -1$. Indivisibility is ignored and so L_t is just considered a positive real number.

2. Only the young work. Each young supplies one unit of labor inelastically. The division of available time between work and leisure is thereby considered exogenous. People have no bequest motive.

3. Output is homogeneous and can be used for consumption as well as investment in physical capital. Physical capital is the only non-human asset in the economy; it is owned by the old and rented out to the firms. Output is the numeraire (unit of account) used in trading. Cash (physical means of payment) is ignored.[3]

4. The economy is closed (no foreign trade).

5. Firms' technology has constant returns to scale.

6. In each period three markets are open, a market for output, a market for labor services, and a market for capital services. Perfect competition rules in all markets. Uncertainty is absent; when a decision is made, its consequences are known.

7. Agents have perfect foresight.

Assumption 7 entails the following. First, the agents are assumed to have "rational expectations" or, with a better name, "model-consistent expectations". This means that forecasts made by the agents coincide with the forecasts that can be calculated on the basis of the model. Second, as there are no stochastic elements in the model (no uncertainty), the forecasts are point estimates rather than probabilistic forecasts. Thereby the model-consistent expectations take the extreme form of *perfect foresight*: the agents agree in their expectations about the future evolution of the economy and ex post this future evolution fully coincides with what was expected.

[3] We may imagine that agents have safe electronic accounts in a fictional central bank allowing costless transfers between accounts.

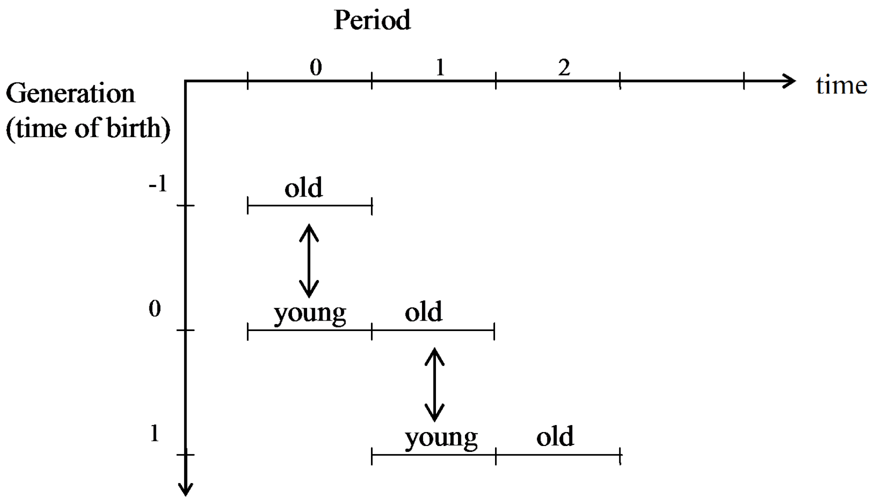


Figure 3.1: The two-period model's time structure.

Of course, this is an unrealistic assumption. The motivation is to simplify in a first approach. The results that emerge will be the outcome of economic mechanisms in isolation from expectational errors. In this sense the model constitutes a "pure" case (benchmark case).

The time structure of the model is illustrated in Fig. 3.1. In every period two generations are alive and interact with each other as indicated by the arrows. The young supply labor and earn a labor income. They consume an endogenous fraction of this income and save the remainder for retirement. Thereby the young offset the dissaving by the old. Possibly, aggregate net saving in the economy will be positive. At the end of the first period the savings by the young are converted into direct ownership of capital goods. In the next period the now old owners of the capital goods rent them out to the firms. We may imagine that the firms are owned by the old, but this ownership is not visible in the equilibrium allocation because pure profits will be nil due to the combination of perfect competition and constant returns to scale.

Let the consumption good (= the output good) be the numeraire, i.e., in any period the consumption good has the price 1. Let $\hat{r}_t$ denote the rental rate for capital in period t; that is, $\hat{r}_t$ is the real price a firm has to pay at the end of period t for the right to use one unit of someone else's physical capital through period t. So the owner of K_t units of physical capital receives a

$$\text{real (net) rate of return on capital } = \frac{\hat{r}_t K_t - \delta K_t}{K_t} = \hat{r}_t - \delta, \tag{3.1}$$

where δ is the rate of physical capital depreciation which is assumed constant, $0 \leq \delta \leq 1$.

Suppose there is also a market for loans. Assume you have lent out one unit of output from the end of period $t-1$ to the end of period t. If the *real interest rate* in the loan market is r_t, then, at the end of period t you should get back $1+r_t$ units of output. In the absence of uncertainty, equilibrium requires that capital and loans give the same rate of return,

$$\hat{r}_t - \delta = r_t. \tag{3.2}$$

This *no-arbitrage* condition indicates how the rental rate for capital and the more everyday concept, the interest rate, would be related in an equilibrium where both the market for capital services and a loan market were active. We shall see, however, that in this model no loan market will be active in an equilibrium. Nevertheless we will follow the tradition and call the right-hand side of (3.2) the *interest rate*.

Table 3.1 provides an overview of the notation. As to our timing convention, notice that any stock variable dated t indicates the amount held at the beginning of period t. That is, the capital stock accumulated by the end of period $t-1$ and available for production in period t is denoted K_t. We therefore write $K_t = (1-\delta)K_{t-1} + I_{t-1}$ and $Y_t = F(K_t, L_t)$, where F is an aggregate production function. In this context it is useful to think of "period t" as running from date t to right before date $t+1$. So period t is the half-open time interval $[t, t+1)$ on the continuous-time axis. Whereas production and consumption take place *in* period t, we imagine that all decisions are made at discrete points in time $t = 0, 1, 2, ...$ ("dates"). We further imagine that receipts for work and lending as well as payment for the consumption in period t occur at the end of the period. These timing conventions are common in discrete-time growth and business cycle theory;[4] they are convenient because they make switching between discrete and continuous time analysis fairly easy.

Table 3.1. List of main variable symbols

[4] In contrast, in the *accounting* and *finance* literature, typically K_t would denote the *end-of-period-t* stock that begins to yield its services *next* period.

Symbol	*Meaning*
L_t	the number of young people in period t
n	generation growth rate
K_t	aggregate capital available in period t
c_{1t}	consumption as young in period t
c_{2t}	consumption as old in period t
w_t	real wage in period t
r_t	real interest rate (from end of per. $t-1$ to end of per. t)
ρ	rate of time preference (impatience)
θ	elasticity of marginal utility
s_t	saving of each young in period t
Y_t	aggregate output in period t
$C_t = c_{1t}L_t + c_{2t}L_{t-1}$	aggregate consumption in period t
$S_t = Y_t - C_t$	aggregate gross saving in period t
$\delta \in [0, 1]$	capital depreciation rate
$K_{t+1} - K_t = I_t - \delta K_t$	aggregate net investment in period t

3.3 The saving by the young

Suppose the preferences of the young can be represented by a lifetime utility function as specified in (3.3). Given w_t and r_{t+1}, the decision problem of the young in period t then is:

$$\begin{aligned} \max_{c_{1t},c_{2t+1}} U(c_{1t}, c_{2t+1}) &= u(c_{1t}) + (1+\rho)^{-1}u(c_{2t+1}) \qquad \text{s.t.} & (3.3)\\ c_{1t} + s_t &= w_t \cdot 1 \qquad (w_t > 0), & (3.4)\\ c_{2t+1} &= (1 + r_{t+1})s_t \qquad (r_{t+1} > -1), & (3.5)\\ c_{1t} &\geq 0, c_{2t+1} \geq 0. & (3.6) \end{aligned}$$

The interpretation of the variables is given in Table 3.1 above. We may think of the "young" as a household consisting of one adult and $1+n$ children whose consumption is included in c_{1t}. Note that "utility" appears at two levels. There is a *lifetime utility function*, U, and a *period utility function*, u.[5] The latter is assumed to be the same in both periods of life (this has little effect on the qualitative results and simplifies the exposition). The period utility function is assumed twice continuously differentiable with $u' > 0$ and $u'' < 0$ (positive, but diminishing marginal utility of consumption). Many popular specifications of u, e.g., $u(c) = \ln c$, have the property that $\lim_{c\to 0} u(c) = -\infty$. Then we *define* $u(0) = -\infty$.

[5] Other names for these two functions are the *intertemporal utility function* and the *subutility function*, respectively.

The interpretation is that $c = 0$ is something the individual will avoid whenever economically possible.

The parameter ρ is called the *rate of time preference.* It acts as a utility discount *rate,* whereas $(1+\rho)^{-1}$ is a utility discount *factor,* cf. Box 3.1. By definition, $\rho > -1$, but $\rho > 0$ is usually assumed. We interpret ρ as reflecting the degree of *impatience* with respect to the "arrival" of utility. When preferences can be represented in this additive way, they are called *time-separable.* In principle, as seen from period t the interest rate appearing in (3.5) should be interpreted as an *expected* real interest rate. But as long as we assume perfect foresight, there is no need that our notation distinguishes between actual and expected magnitudes.

Box 3.1. Discount factors and discount rates

A *discount factor* is a factor by which future benefits or costs, measured in some unit of account, are converted into present equivalents. By a *discount rate* is meant an interest rate applied in the construction of a discount factor. The higher the discount rate the lower the discount factor.

One should bear in mind that a discount rate depends on what is to be discounted. In (3.3) the unit of account is "utility" and ρ acts as a *utility discount rate.* In (3.7) the unit of account is the consumption good and r_{t+1} acts as a *consumption discount rate.* If people also work as old, the right-hand side of (3.7) would read $w_t + (1 + r_{t+1})^{-1}w_{t+1}$ and thus r_{t+1} would act as an *earnings discount rate.* This will be the same as the consumption discount rate if we think of real income measured in consumption units. But if we think of nominal income, that is, income measured in monetary units, there would be a *nominal earnings discount rate*, namely the *nominal* interest rate, which in an economy with inflation will exceed the consumption discount rate. Unfortunately, confusion of different discount rates is not rare.

As the price of the consumption good is 1, it is not visible. The reason that the right-hand side of (3.4) is written $w_t \cdot 1$ is that the inelastic labor supply of the young is normalized to one unit of work.

In (3.5) the interest rate r_{t+1} acts as a (net) rate of return on saving.[6] An interest rate may also be seen as a discount rate relating to consumption over time.

[6]While s_t in (3.4) appears as a *flow* (non-consumed income), in (3.5) s_t appears as a *stock* (the accumulated financial wealth at the end of period t). This notation is legitimate because the magnitude of the two is the same when the time unit is the same as the period length. Indeed, the interpretation of s_t in (3.5) $s_t = s_t \cdot \Delta t = s_t \cdot 1$ units of account.

In real life the gross payoff of individual saving is sometimes nil (for instance if invested in a project that completely fails). But we ignore this possibility and so the discount factor $1/(1 + r_{t+1})$ is always well-defined.

Indeed, by isolating s_t in (3.5) and substituting into (3.4), we may consolidate the two period budget constraints of the individual into *one* budget constraint,

$$c_{1t} + \frac{1}{1+r_{t+1}} c_{2t+1} = w_t. \tag{3.7}$$

This is an *intertemporal budget constraint* and says that the present value from the end of period t, of the planned consumption sequence, must equal the labor income received at the end of period t. The interest rate appears as the discount rate entering the discount factor converting future amounts of consumption into present equivalents. That is why addition on the left-hand side of equation (3.7) makes sense.

Solving the saving problem

To avoid the possibility of corner solutions, we impose the No Fast Assumption

$$\lim_{c \to 0} u'(c) = \infty. \tag{A1}$$

In view of the sizeable period length in the model, this is definitely plausible.

Inserting the two budget constraints into the objective function in (3.3), we get $U(c_{1t}, c_{2t+1}) = u(w_t - s_t) + (1+\rho)^{-1} u((1+r_{t+1})s_t) \equiv \tilde{U}_t(s_t)$, a function of only one decision variable, s_t. According to the non-negativity constraint on consumption in both periods, (3.6), s_t must satisfy $0 \leq s_t \leq w_t$. Maximizing with respect to s_t gives the first-order condition

$$\frac{d\tilde{U}_t}{ds_t} = -u'(w_t - s_t) + (1+\rho)^{-1} u'((1+r_{t+1})s_t)(1+r_{t+1}) = 0. \tag{FOC}$$

The second derivative of $\tilde{U}_t$ is

$$\frac{d^2\tilde{U}_t}{ds_t^2} = u''(w_t - s_t) + (1+\rho)^{-1} u''((1+r_{t+1})s_t)(1+r_{t+1})^2 < 0. \tag{SOC}$$

Hence there can at most be one s_t satisfying (FOC). Moreover, for a positive wage income there always exists such an s_t. Indeed:

LEMMA 1 Let $w_t > 0$ and suppose the No Fast Assumption (A1) applies. Then the saving problem of the young has a unique solution $s_t = s(w_t, r_{t+1})$. The solution is interior, i.e., $0 < s_t < w_t$, and s_t satisfies (FOC).

Proof. Assume (A1). For any $s \in (0, w_t)$, $d\tilde{U}_t(s)/ds > -\infty$. Now consider the endpoints $s = 0$ and $s = w_t$. By (FOC) and (A1),

$$\begin{aligned}
\lim_{s \to 0} \frac{d\tilde{U}_t}{ds} &= -u'(w_t) + (1+\rho)^{-1}(1+r_{t+1}) \lim_{s \to 0} u'((1+r_{t+1})s) = \infty, \\
\lim_{s \to w} \frac{d\tilde{U}_t}{ds} &= -\lim_{s \to w_t} u'(w_t - s) + (1+\rho)^{-1}(1+r_{t+1}) u'((1+r_{t+1})w_t) = -\infty.
\end{aligned}$$

By continuity of $d\tilde{U}_t/ds$ follows that there exists an $s_t \in (0, w_t)$ such that at $s = s_t$, $d\tilde{U}_t/ds = 0$. This is an application of the *intermediate value theorem.* It further follows that (FOC) holds for this s_t. By (SOC), s_t is unique and can therefore be written as an implicit function, $s(w_t, r_{t+1})$, of the exogenous variables in the problem, w_t and r_{t+1}. □

Inserting the solution for s_t into the two period budget constraints, (3.4) and (3.5), we immediately get the optimal consumption levels, c_{1t} and c_{2t+1}.

The simple optimization method we have used here is called the *substitution method*: by substitution of the constraints into the objective function an unconstrained maximization problem is obtained.[7]

The consumption Euler equation

The first-order condition (FOC) can conveniently be written

$$u'(c_{1t}) = (1+\rho)^{-1}u'(c_{2t+1})(1+r_{t+1}). \tag{3.8}$$

This is known as an *Euler equation,* after the Swiss mathematician L. Euler (1707-1783) who was the first to study dynamic optimization problems. In the present context the condition is called a *consumption Euler equation.*

Intuitively, in an optimal plan the marginal utility cost of saving must equal the marginal utility benefit obtained by saving. The marginal utility cost of saving is the opportunity cost (in terms of current utility) of saving one more unit of account in the current period (approximately). This one unit of account is transferred to the next period with interest so as to result in $1 + r_{t+1}$ units of account in that period. An optimal plan requires that the utility cost equals the utility benefit of instead having $1 + r_{t+1}$ units of account in the next period. And this utility benefit is the discounted value of the extra utility that can be obtained next period through the increase in consumption by $1 + r_{t+1}$ units compared with the situation without the saving of the marginal unit.

It may seem odd to attempt an intuitive interpretation this way, that is, in terms of "utility units". The utility concept is just a convenient mathematical device used to represent the assumed *preferences.* Our interpretation is only meant as an as-if interpretation: as if utility were something concrete. An interpretation in terms of concrete *measurable quantities* goes like this. We rewrite (3.8) as

$$\frac{u'(c_{1t})}{(1+\rho)^{-1}u'(c_{2t+1})} = 1 + r_{t+1}. \tag{3.9}$$

The left-hand side measures the *marginal rate of substitution,* MRS, of consumption as old for consumption as young, evaluated at the point (c_1, c_2). MRS is

[7] Alternatively, one could use the Lagrange method.

defined as the increase in period-$(t+1)$ consumption needed to compensate for a marginal decrease in period-t consumption. That is,

$$MRS_{c_2 c_1} = -\frac{dc_{2t+1}}{dc_{1t}}|_{U=\bar{U}} = \frac{u'(c_{1t})}{(1+\rho)^{-1}u'(c_{2t+1})}, \tag{3.10}$$

where we have used implicit differentiation in $U(c_{1t}, c_{2t+1}) = \bar{U}$. The right-hand side of (3.9) indicates the marginal rate of transformation, MRT, which is the rate at which saving allows an agent to shift consumption from period t to period $t+1$ via the market. In an optimal plan MRS must equal MRT.

Even though interpretations in terms of "MRS equals MRT" are more satisfactory, we will often use "as if" interpretations like the one before. They are a convenient short-hand for the more elaborate interpretation.

By the Euler equation (3.8),

$$\rho \lesseqqgtr r_{t+1} \text{ implies } u'(c_{1t}) \gtreqqless u'(c_{2t+1}), \text{ i.e., } c_{1t} \lesseqqgtr c_{2t+1},$$

respectively, in the optimal plan (because $u'' < 0$). That is, absent uncertainty the optimal plan entails either increasing, constant or decreasing consumption over time according to whether the rate of time preference is below, equal to, or above the market interest rate, respectively. For example, when $\rho < r_{t+1,}$ the plan is to start with relatively low consumption in order to take advantage of the relatively high rate of return on saving.

Note that there are infinitely many pairs (c_{1t}, c_{2t+1}) satisfying the Euler equation (3.8). Only when requiring the two period budget constraints, (3.4) and (3.5), satisfied, do we get the unique solution s_t and thereby the unique solution for c_{1t} and c_{2t+1}.

Properties of the saving function

The first-order condition (FOC), where the two budget constraints are inserted, determines the saving as an implicit function of the market prices faced by the young decision maker, i.e., $s_t = s(w_t, r_{t+1})$.

The partial derivatives of this function can be found by applying the *implicit function theorem* on (FOC). A practical procedure is the following. We first interpret $d\tilde{U}_t/ds_t$ in (FOC) as a function, f, of the variables involved, s_t, w_t, and r_{t+1}, i.e.,

$$\frac{d\tilde{U}_t}{ds_t} = -u'(w_t - s_t) + (1+\rho)^{-1}u'((1+r_{t+1})s_t)(1+r_{t+1}) \equiv f(s_t, w_t, r_{t+1}).$$

By (FOC),

$$f(s_t, w_t, r_{t+1}) = 0. \tag{*}$$

The implicit function theorem (see Math tools) now implies that if $\partial f/\partial s_t \neq 0$, then the equation (*) defines s_t as an implicit function of w_t and r_{t+1}, $s_t = s(w_t, r_{t+1})$, with continuous partial derivatives

$$\frac{\partial s_t}{\partial w_t} = -\frac{\partial f/\partial w_t}{D} \quad \text{and} \quad \frac{\partial s_t}{\partial r_{t+1}} = -\frac{\partial f/\partial r_{t+1}}{D},$$

where $D \equiv \partial f/\partial s_t \equiv d^2\tilde{U}_t/ds_t^2 < 0$ by (SOC). We find

$$\begin{aligned}
\frac{\partial f}{\partial w_t} &= -u''(c_{1t}) > 0, \\
\frac{\partial f}{\partial r_{t+1}} &= (1+\rho)^{-1}\left[u'(c_{2t+1}) + u''(c_{2t+1})s_t(1+r_{t+1})\right].
\end{aligned}$$

Consequently, the partial derivatives of the saving function $s_t = s(w_t, r_{t+1})$ are

$$s_w \equiv \frac{\partial s_t}{\partial w_t} = \frac{u''(c_{1t})}{D} > 0 \quad (\text{but } < 1), \tag{3.11}$$

$$s_r \equiv \frac{\partial s_t}{\partial r_{t+1}} = -\frac{(1+\rho)^{-1}[u'(c_{2t+1}) + u''(c_{2t+1})c_{2t+1}]}{D}, \tag{3.12}$$

where in the last expression we have used (3.5).[8]

The role of w_t for saving is straightforward. Indeed, (3.11) shows that $0 < s_w < 1$, which implies that $0 < \partial c_{1t}/\partial w_t < 1$ and $0 < \partial c_{2t}/\partial w_t < 1 + r_{t+1}$. The positive sign of these two derivatives indicate that consumption in each of the periods is a *normal* good (which certainly is plausible since we are talking about the total consumption by the individual in each period).[9]

The sign of s_r in (3.12) is seen to be ambiguous. This ambiguity regarding the role of r_{t+1} for saving reflects that the Slutsky substitution and income effects on consumption as young of a rise in the interest rate are of opposite signs. To

---

[8] A perhaps more straightforward procedure, not requiring full memory of the exact content of the implicit function theorem, is based on "implicit differentiation". First, keeping r_{t+1} fixed, one calculates the total derivative w.r.t. w_t on both sides of (FOC). Next, keeping w_t fixed, one calculates the total derivative with respect to r_{t+1} on both sides of (FOC).

Yet another possible procedure is based on "total differentiation" in terms of *differentials*. Taking the differential w.r.t. s_t, w_t, and r_{t+1} on both sides of (FOC) gives $-u''(c_{1t})(dw_t - ds_t) +$ $+(1+\rho)^{-1}\cdot\{u''(c_{2t+1})\left[(1+r_{t+1})ds_t + s_t dr_{t+1}\right](1+r_{t+1}) + u'(c_{2t+1})dr_{t+1}\} = 0$. By rearranging we find the ratios ds_t/dw_t and ds_t/dr_{t+1}, which will indicate the value of the partial derivatives (3.11) and (3.12).

[9] Recall, a consumption good is called *normal* for given consumer preferences if the demand for it is an increasing function of the consumer's "endowment", sometimes called the "initial resources" or "initial wealth". Since in this model the consumer is born without any financial wealth, the consumer's "endowment" evaluated at the end of period t is simply the value of the labor earnings, i.e., $w_t \cdot 1$.

understand this, it is useful to keep the intertemporal budget constraint, (3.7), in mind. The *substitution effect* on c_{1t} is negative because the higher interest rate makes future consumption cheaper in terms of current consumption. And the *income effect* on c_{1t} is positive because with a higher interest rate, a given budget can buy more consumption in both periods, cf. (3.7).[10] Generally there would be a third Slutsky effect, a *wealth effect* of a rise in the interest rate. But such an effect is ruled out in this model. This is because there is no labor income in the second period of life. Indeed, as indicated by (3.7), the "initial resources" of a member of generation t, evaluated at the end of period t, is simply w_t, which is independent of r_{t+1}. (In contrast, with labor income, say w_{t+1}, in the second period, the "initial resources" would be $w_t + w_{t+1}/(1+r_{t+1})$. This present discounted value of life-time earnings clearly depends negatively on r_{t+1}, and so a negative wealth effect on c_{1t} of a rise in the interest rate would arise.)

Rewriting (3.12) gives

$$s_r = \frac{(1+\rho)^{-1}u'(c_{2t+1})[\theta(c_{2t+1})-1]}{D} \gtreqless 0 \text{ for } \theta(c_{2t+1}) \lesseqgtr 1, \tag{3.13}$$

respectively, where $\theta(c_{2t+1})$ is the absolute *elasticity of marginal utility* of consumption in the second period, that is,

$$\theta(c_{2t+1}) \equiv -\frac{c_{2t+1}}{u'(c_{2t+1})}u''(c_{2t+1}) \approx -\frac{\Delta u'(c_{2t+1})/u'(c_{2t+1})}{\Delta c_{2t+1}/c_{2t+1}} > 0,$$

where the approximation builds upon $u''(c_{2t+1}) \approx \Delta u'(c_{2t+1})/\Delta c_{2t+1}$. The inequalities in (3.13) show that when the absolute elasticity of marginal utility is below one, then the substitution effect on consumption as young of an increase in the interest rate dominates the income effect and saving increases. The opposite is true if the elasticity of marginal utility is above one.

The reason that $\theta(c_{2t+1})$ has this role is that $\theta(c_{2t+1})$ reflects how sensitive marginal utility of c_{2t+1} is to a rise in c_{2t+1}. To see the intuition, consider the case where consumption as young – and thus saving – happens to be unaffected by an increase in the interest rate. Even in this case, consumption as old, c_{2t+1}, is automatically increased (in view of the higher income as old through the higher rate of return on the unchanged saving); and the marginal utility of c_{2t+1} is thus decreased in response to a higher interest rate. The point is that this outcome can only be optimal if the elasticity of marginal utility of c_{2t+1} is of "medium" size. A very high absolute elasticity of marginal utility of c_{2t+1} would result in a sharp decline in marginal utility – so sharp that not much would be lost by dampening the automatic rise in c_{2t+1} and instead increase c_{1t}, thus reducing saving. On the

[10] Economists' jargon for substitution effect and income effect is sometimes *carrot effect* and *hammock effect*, respectively.

other hand, a very low elasticity of marginal utility of c_{2t+1} would result in only a small decline in marginal utility – so small that it is beneficial to take advantage of the higher rate of return and save *more*, thus accepting a first-period utility loss brought about by a lower c_{1t}.

We see from (3.13) that an absolute elasticity of marginal utility equal to exactly one is the case leading to the interest rate being *neutral* vis-a-vis the saving of the young. What is the intuition behind this? Neutrality vis-a-vis the saving of the young of a rise in the interest rate requires that c_{1t} remains unchanged since $c_{1t} = w_t - s_t$. In turn this requires that the marginal utility, $u'(c_{2t+1})$, on the right-hand side of (3.8) falls by the same percentage as $1 + r_{t+1}$ rises. At the same time, the budget (3.5) as old tells us that c_{2t+1} has to rise by the same percentage as $1 + r_{t+1}$ if s_t remains unchanged. Altogether we thus need that $u'(c_{2t+1})$ falls by the same percentage as c_{2t+1} rises. But this requires that the absolute elasticity of $u'(c_{2t+1})$ with respect to c_{2t+1} is exactly one.

The elasticity of marginal utility, also called the marginal utility flexibility, will generally depend on the level of consumption, as implicit in the notation $\theta(c_{2t+1})$. There exists a popular special case, however, where the elasticity of marginal utility is constant.

EXAMPLE 1 *The CRRA utility function.* If we impose the requirement that $u(c)$ should have an absolute elasticity of marginal utility of consumption equal to a constant $\theta > 0$, then one can show (see Appendix A) that the utility function must, up to a positive linear transformation, be of the CRRA form:

$$u(c) = \begin{cases} \frac{c^{1-\theta}-1}{1-\theta}, & \text{when } \theta \neq 1, \\ \ln c, & \text{when } \theta = 1. \end{cases}, \tag{3.14}$$

It may seem odd that when $\theta \neq 1$, we subtract the constant $1/(1-\theta)$ from $c^{1-\theta}/(1-\theta)$. Adding or subtracting a constant from a utility function does not affect the marginal rate of substitution and consequently not behavior. So we could do without this constant, but its occurrence in (3.14) has two formal advantages. One is that in contrast to $c^{1-\theta}/(1-\theta)$, the expression $(c^{1-\theta}-1)/(1-\theta)$ can be interpreted as valid even for $\theta = 1$, namely as identical to $\ln c$. This is because $(c^{1-\theta}-1)/(1-\theta) \to \ln c$ for $\theta \to 1$ (by L'Hôpital's rule for "0/0"). Another advantage is that the kinship between the different members, indexed by θ, of the "CRRA family" becomes more transparent. Indeed, by defining $u(c)$ as in (3.14), all graphs of $u(c)$ will go through the same point as the log function, namely $(1, 0)$, cf. Fig. 3.2. The equation (3.14) thus displays the CRRA utility function in *normalized form*.

The higher is θ, the more "curvature" does the corresponding curve in Fig. 3.2 have. In turn, more "curvature" reflects a higher incentive to smooth consumption across time. The reason is that a large curvature means that the marginal utility

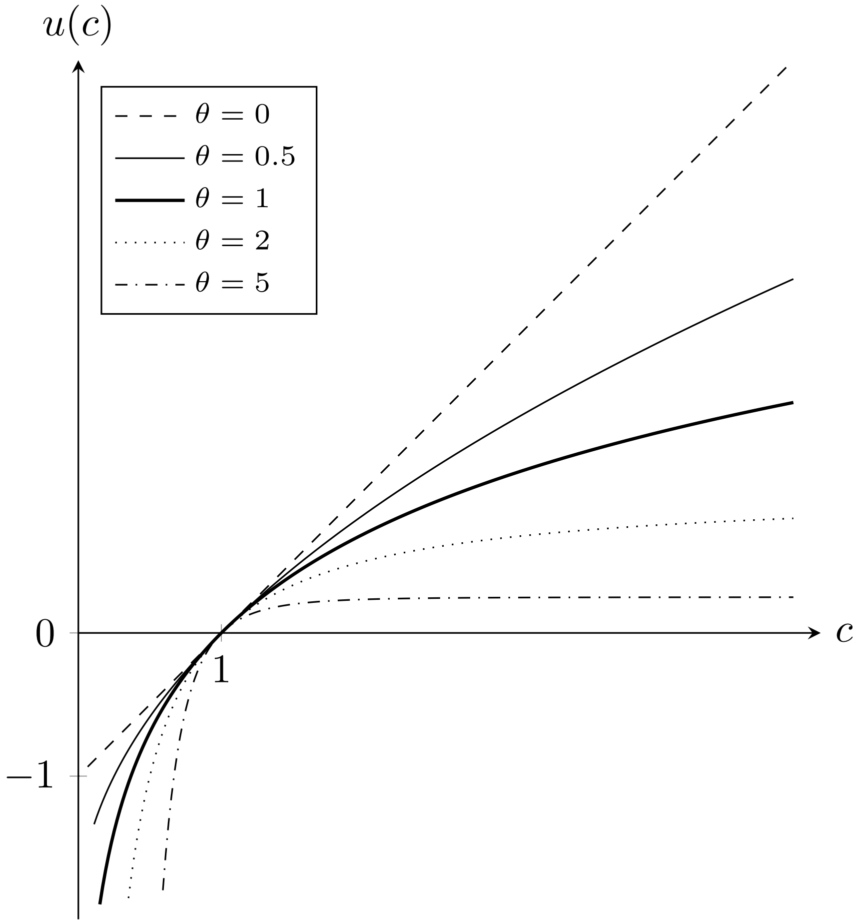


Figure 3.2: The CRRA family of utility functions.

will drop sharply if consumption rises and will increase sharply if consumption falls. Consequently, not much utility is lost by lowering consumption when it is relatively high but there is a lot of utility to be gained by raising it when it is relatively low. So the curvature θ indicates the degree of *aversion towards variation in consumption*. Or we may say that θ indicates the strength of the *preference for consumption smoothing*.[11] □

Suppose the period utility is of CRRA form as given in (3.14). (FOC) then

[11] The name CRRA is a shorthand for *Constant Relative Risk Aversion* and comes from the theory of behavior under uncertainty. Also in that theory does the CRRA function constitute an important benchmark case. And θ is in that context called the *degree of relative risk aversion*.

yields an explicit solution for the saving of the young:

$$s_t = \frac{1}{1 + (1+\rho)(\frac{1+r_{t+1}}{1+\rho})^{\frac{\theta-1}{\theta}}} w_t. \tag{3.15}$$

We see that the signs of $\partial s_t/\partial w_t$ and $\partial s_t/\partial r_{t+1}$ shown in (3.11) and (3.13), respectively, are confirmed. Moreover, the saving of the young is in this special case proportional to income with a factor of proportionality that depends on the interest rate (as long as $\theta \neq 1$). But in the general case the saving-income ratio depends also on the income level.

A major part of the attempts at empirically estimating θ suggests that $\theta > 1$. Based on U.S. data, Hall (1988) provides estimates above 5, while Attanasio and Weber (1993) suggest $1.25 \leq \theta \leq 3.33$. For Japanese data Okubo (2011) suggests $2.5 \leq \theta \leq 5.0$. As these studies relate to much shorter time intervals than the implicit time horizon of about 2×30 years in the Diamond model, we should be cautious. But if the estimates *were* valid also to that model, we should expect the income effect on current consumption of an increase in the interest rate to dominate the substitution effect, thus implying $s_r < 0$ *as long as there is no wealth effect* of a rise in the interest rate.

When the elasticity of marginal utility of consumption is a constant, θ, its inverse, $1/\theta$, equals the *elasticity of intertemporal substitution* in consumption. This concept refers to the willingness to substitute consumption over time when the interest rate changes. Under certain conditions the elasticity of intertemporal substitution reflects the elasticity of the ratio c_{2t+1}/c_{1t} with respect to $1 + r_{t+1}$ when we move along a given indifference curve. The next subsection, which can be omitted in a first reading, goes more into detail with the concept.

Digression: The elasticity of intertemporal substitution*

Consider a two-period consumption problem like the one above. Fig. 3.3 depicts a particular indifference curve, $u(c_1) + (1 + \rho)^{-1}u(c_2) = \bar{U}$. At a given point, (c_1, c_2), on the curve, the marginal rate of substitution of period-2 consumption for period-1 consumption, MRS, is given by

$$MRS = -\frac{dc_2}{dc_1}|_{U=\bar{U}} ,$$

that is, MRS at the point (c_1, c_2) is the absolute value of the slope of the tangent to the indifference curve at that point.[12] Under the "normal" assumption of "strictly convex preferences" (as for instance in the Diamond model), MRS is

[12] When the meaning is clear from the context, to save notation we just write MRS instead of the more precise $MRS_{c_2 c_1}$.

rising along the curve when c_1 decreases (and thereby c_2 increases). Conversely, we can let MRS be the independent variable and consider the corresponding point on the indifference curve, and thereby the ratio c_2/c_1, as a function of MRS. If we raise MRS along the indifference curve, the corresponding value of the ratio c_2/c_1 will also rise.

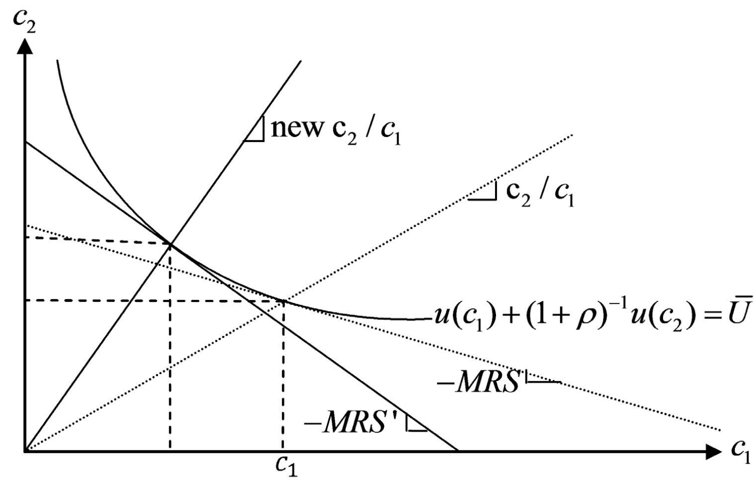


Figure 3.3: Substitution of period 2-consumption for period 1-consumption as MRS increases to MRS'.

The *elasticity of intertemporal substitution in consumption* at a given point is defined as the elasticity of the ratio c_2/c_1 with respect to the marginal rate of substitution of c_2 for c_1, when we move along the indifference curve through the point (c_1, c_2). Letting the elasticity with respect to x of a differentiable function $f(x)$ be denoted $\mathrm{E}\ell_x f(x)$, the elasticity of intertemporal substitution in consumption can be written

$$\mathrm{E}\ell_{MRS}\frac{c_2}{c_1} = \frac{MRS}{c_2/c_1}\frac{d\,(c_2/c_1)}{dMRS}\,|_{U=\bar{U}} \approx \frac{\frac{\Delta(c_2/c_1)}{c_2/c_1}}{\frac{\Delta MRS}{MRS}},$$

where the approximation is valid for a "small" increase, ΔMRS, in MRS.

A more concrete understanding is obtained when we take into account that in the consumer's optimal plan, MRS equals the ratio of the discounted prices of good 1 and good 2, that is, the ratio $1/(1/(1+r))$ given in (3.7). Indeed, from (3.10) and (3.9), omitting the time indices, we have

$$MRS = -\frac{dc_2}{dc_1}\,|_{U=\bar{U}} = \frac{u'(c_1)}{(1+\rho)^{-1}u'(c_2)} = 1 + r. \tag{3.16}$$

Letting $R \equiv 1+r$ and $\sigma(c_1, c_2)$ denote the elasticity of intertemporal substitution, evaluated at the point (c_1, c_2), we then have

$$\sigma(c_1, c_2) = \frac{R}{c_2/c_1}\frac{d\,(c_2/c_1)}{dR}\,|_{U=\bar{U}} \approx \frac{\frac{\Delta(c_2/c_1)}{c_2/c_1}}{\frac{\Delta R}{R}}. \tag{3.17}$$

Consequently, the elasticity of intertemporal substitution can here be interpreted as the approximate percentage increase in the consumption ratio, c_2/c_1, triggered by a one percentage increase in the inverse price ratio, holding the utility level unchanged.[13]

Given $u(c)$, we let $\theta(c)$ be the absolute elasticity of marginal utility of consumption, i.e., $\theta(c) \equiv -cu''(c)/u'(c)$. As shown in Appendix B, we then find the elasticity of intertemporal substitution to be

$$\sigma(c_1, c_2) = \frac{c_2 + Rc_1}{c_2\theta(c_1) + Rc_1\theta(c_2)}. \tag{3.18}$$

We see that if $u(c)$ belongs to the CRRA class and thereby $\theta(c_1) = \theta(c_2) = \theta$, then $\sigma(c_1, c_2) = 1/\theta$. In this case (as well as whenever $c_1 = c_2$) the elasticity of marginal utility and the elasticity of intertemporal substitution are simply the inverse of each other.

3.4 Production

Output is homogeneous and can be used for consumption as well as investment in physical capital. The capital stock is thereby just accumulated non-consumed output. We may imagine a "corn economy" where output is corn, part of which is eaten (flour) while the remainder is accumulated as capital (seed corn in the ground).

The specification of technology and production conditions follows the simple competitive one-sector setup discussed in Chapter 2.4. Although the Diamond model is a long-run model, we shall in this chapter for simplicity ignore technological change.

The representative firm

There is a representative firm with a neoclassical production function and constant returns to scale (CRS). Omitting the time argument t when not needed for

[13] This characterization is equivalent to saying that the elasticity of substitution between two consumption goods indicates the approximate percentage *decrease* in the ratio of the chosen quantities of the goods (when moving along a given indifference curve) induced by a one-percentage *increase* in the *corresponding* price ratio.

clarity, we have

$$Y = F(K, L) = LF(k, 1) \equiv Lf(k), \qquad f(0) \geq 0, f' > 0, f'' < 0, \tag{3.19}$$

where Y is output (GNP) per period, K is capital input, L is labor input, and $k \equiv K/L$ is the capital-labor ratio. The derived function, f, is the production function in intensive form. Capital installation and other adjustment costs are ignored. With $\hat{r}$ denoting the rental rate for capital, (pure) profit is $\Pi \equiv F(K, L) - \hat{r}K - wL$. The firm maximizes Π under perfect competition. This gives, first, $\partial\Pi/\partial K = F_K(K, L) - \hat{r} = 0$, that is,

$$F_K(K, L) = \frac{\partial [Lf(k)]}{\partial K} = f'(k) = \hat{r}. \tag{3.20}$$

Second, $\partial\Pi/\partial L = F_L(K, L) - w = 0$, that is,

$$F_L(K, L) = \frac{\partial [Lf(k)]}{\partial L} = f(k) - kf'(k) = w. \tag{3.21}$$

We may interpret these two conditions as saying that the firm will in every period use capital and labor up to the point where the marginal productivities of K and L, respectively, given the chosen input of the other factor, are equal to the respective factor prices from the market. Such an intuitive formulation does not take us far, however. Indeed, because of CRS there may be infinitely many pairs (K, L), if any, that satisfy (3.20) and (3.21). What we can definitely ascertain, however, is that in view of $f'' < 0$, a $k > 0$ satisfying (3.20) will be *unique*.[14] Let us call it the *desired capital-labor ratio* and recognize that at this stage the separate factor inputs, K and L, are indeterminate. While (3.20) and (3.21) are just first-order conditions for the profit maximizing representative firm, to get further we have to appeal to *equilibrium* in the factor markets.

Factor prices in equilibrium

Let the aggregate demand for capital services and labor services be denoted K^d and L^d, respectively. Clearing in factor markets in period t implies

$$K_t{}^d = K_t, \tag{3.22}$$

$$L_t{}^d = L_t = L_0(1 + n)^t, \tag{3.23}$$

[14]It might seem that k is overdetermined because we have two equations, (3.20) and (3.21), but only one unknown. This reminds us that for *arbitrary* factor prices, $\hat{r}$ and w, there will generally *not* exist a k satisfying both (3.20) and (3.21). But in equilibrium the factor prices faced by the firm are not arbitrary. They are equilibrium prices, i.e., they are adjusted so that (3.20) and (3.21) become consistent.

where K_t is the aggregate supply of capital services and L_t the aggregate supply of labor services. As was called attention to in Chapter 1, unless otherwise specified it is understood that the rate of utilization of each production factor is constant over time and normalized to one. So the quantity K_t will at one and the same time measure both the capital input, a flow, and the available capital stock. Similarly, the quantity L_t will at one and the same time measure both the labor input, a flow, and the size of the labor force as a stock (= the number of young people).

The aggregate input demands, K^d and L^d, are linked through the desired capital-labor ratio, k^d. In equilibrium we have $K_t^d/L_t^d = k_t{}^d = K_t/L_t \equiv k_t$, by the market clearing conditions (3.22) and (3.23). The k in (3.20) and (3.21) can thereby be identified with the ratio of the stock supplies, $k_t \equiv K_t/L_t > 0$, which is a predetermined variable. Interpreted this way, (3.20) and (3.21) *determine* the equilibrium factor prices $\hat{r}_t$ and w_t in each period. In view of the no-arbitrage condition (3.2), the real interest rate satisfies $r_t = \hat{r}_t - \delta$, where δ is the capital depreciation rate, $0 \leq \delta \leq 1$. So in equilibrium we end up with

$$r_t = f'(k_t) - \delta \equiv r(k_t) \qquad (r'(k_t) = f''(k_t) < 0), \tag{3.24}$$

$$w_t = f(k_t) - k_t f'(k_t) \equiv w(k_t) \qquad (w'(k_t) = -k_t f''(k_t) > 0), \tag{3.25}$$

where causality is from the right to the left in the two equations. In line with our general perception of perfect competition, cf. Section 2.4 of Chapter 2, it is understood that the factor prices, $\hat{r}_t$ and w_t, adjust quickly to the market-clearing levels.

Technical Remark. In these formulas it is understood that $L > 0$, but we may allow $K = 0$, i.e., $k = 0$. In case $f'(0)$ is not immediately well-defined, we interpret $f'(0)$ as $\lim_{k\to 0^+} f'(k)$ if this limit exists. If it does not, it must be because we are in a situation where $\lim_{k\to 0^+} f'(k) = \infty$, since $f''(k) < 0$ (an example is the Cobb-Douglas function, $f(k) = Ak^\alpha$, $0 < \alpha < 1$, where $\lim_{k\to 0^+} f'(k)$ $= \lim_{k\to 0^+} A\alpha k^{\alpha-1} = +\infty$). In this situation we simply include $+\infty$ in the range of $r(k)$ and define $r(0) \cdot 0 \equiv \lim_{k\to 0^+}(f'(k) - \delta)k = 0$, where the last equality comes from the general property of a neoclassical CRS production function that $\lim_{k\to 0^+} kf'(k) = 0$, cf. (2.18) of Chapter 2. Letting $r(0) \cdot 0 = 0$ also fits well with intuition since, when $k = 0$, nobody receives capital income anyway. Note that since $\delta \in [0, 1]$, $r(k) > -1$ for all $k \geq 0$. What about $w(0)$? We interpret $w(0)$ as $\lim_{k\to 0} w(k)$. From (2.18) of Chapter 2 we have that $\lim_{k\to 0^+} w(k) = f(0)$ $\equiv F(0, 1) \geq 0$. If capital is essential, $F(0, 1) = 0$. Otherwise, $F(0, 1) > 0$. Finally, since $w' > 0$, we have, for $k > 0$, $w(k) > 0$ as also noted in Chapter 2. □

To fix ideas we have assumed that households (here the old) own the physical capital and rent it out to the firms. In view of perfect competition and constant returns to scale, pure profit is nil in equilibrium. As long as the model ignores uncertainty and capital installation costs, the results will be unaffected if instead we let the firms themselves own the physical capital and finance capital investment by issuing bonds and shares. These bonds and shares would then be accumulated by the households and constitute their financial wealth instead of the capital goods themselves. The equilibrium rate of return, r_t, would be the same.

3.5 The dynamic path of the economy

As in microeconomic general equilibrium theory, it is important to distinguish between the set of technically feasible allocations and an allocation brought about, within this set, by a specific economic institution (the rules of the game). The economic institution assumed by the Diamond model is the private-ownership perfect-competition market institution.

We shall in the next subsections introduce three different concepts concerning allocations over time in this economy. The three concepts are: *technically feasible paths*, *temporary equilibrium*, and *equilibrium path*. These concepts are related to each other in the sense that there is a whole *set* of technically feasible paths, *within which* there may exist a unique equilibrium path, which in turn is a sequence of states that have certain properties, including the temporary equilibrium property.

3.5.1 Technically feasible paths

When we speak of technically feasible paths, the focus is merely upon what is feasible from the point of view of the given technology as such and available initial resources. That is, we disregard the agents' preferences, their choices given the constraints, their interactions in markets, the market forces etc.

The technology is given by $Y_t = F(K_t, L_t)$, and there are two exogenous "initial resources", the labor force, $L_t = L_0(1+n)^t$, and the initial capital stock, K_0. From national income accounting aggregate consumption can be written $C_t \equiv Y_t - S_t = F(K_t, L_t) - S_t$, where S_t denotes aggregate gross saving, and where we have inserted (3.19). In a closed economy aggregate gross saving equals (ex post) aggregate gross investment, $K_{t+1} - K_t + \delta K_t$. So

$$C_t = F(K_t, L_t) - (K_{t+1} - K_t + \delta K_t). \tag{3.26}$$

Let c_t denote aggregate consumption per unit of labor in period t, i.e.,

$$c_t \equiv \frac{C_t}{L_t} = \frac{c_{1t}L_t + c_{2t}L_{t-1}}{L_t} = c_{1t} + \frac{c_{2t}}{1+n}.$$

Combining this with (3.26) and using the definitions of k and $f(k)$, we obtain the dynamic resource constraint of the economy:

$$c_{1t} + \frac{c_{2t}}{1+n} = f(k_t) + (1-\delta)k_t - (1+n)k_{t+1}. \tag{3.27}$$

DEFINITION 1 Let $\bar{k}_0 \geq 0$ be the historically given initial ratio of available capital and labor. Let the path $\{(k_t, c_{1t}, c_{2t})\}_{t=0}^{\infty}$ have nonnegative k_t, c_{1t}, and c_{2t} for all $t = 0, 1, 2, \ldots$. The path is called *technically feasible* if it has $k_0 = \bar{k}_0$ and satisfies (3.27) for all $t = 0, 1, 2, \ldots$.

The next subsections consider how, for given household preferences, the private-ownership institution in combination with competitive markets generates a *selection* within the set of technically feasible paths. A member of this selection (which may but need not have just one member) is called an *equilibrium path.* It constitutes a sequence of states with certain properties, one of which is the temporary equilibrium property.

3.5.2 A temporary equilibrium

It is natural to think of next period's interest rate as an *expected* interest rate that provisionally can deviate from the ex post realized one. We let $r_{t+1}^e > -1$ denote the expected real interest rate of period $t+1$ as seen from period t.

Essentially, by a temporary equilibrium in period t is meant a state where for a *given* r_{t+1}^e, all markets clear in the period. There are three markets, namely two factor markets and a market for produced goods. We have already described the two factor markets. In the market for produced goods the representative firm supplies the amount $Y_t^s = F(K_t^d, L_t^d)$ in period t. The demand side in this market has two components, consumption, C_t, and *gross* investment, I_t. Equilibrium in the goods market requires that demand equals supply, i.e.,

$$C_t + I_t = c_{1t}L_t + c_{2t}L_{t-1} + I_t = Y_t^s = F(K_t^d, L_t^d), \tag{3.28}$$

where consumption by the young and old, c_{1t} and c_{2t}, respectively, were determined in Section 3.

By definition, aggregate gross investment equals aggregate net investment, I_t^N, plus capital depreciation, i.e.,

$$I_t = I_t^N + \delta K_t \equiv I_{1t}^N + I_{2t}^N + \delta K_t \equiv S_{1t}^N + S_{2t}^N + \delta K_t = s_t L_t + (-K_t) + \delta K_t. \tag{3.29}$$

The first equality follows from the definition of net investment and the assumption that capital depreciation equals δK_t. Next comes an identity reflecting that aggregate net investment is the sum of net investment by the young and net investment by the old. In turn, saving in this model is directly an act of acquiring capital goods. So the net investment by the young, I_{1t}^N, and the old, I_{2t}^N, are identical to their net saving, S_{1t}^N and S_{2t}^N, respectively. As we have shown, the net saving by the young in the model equals $s_t L_t$. And the net saving by the old is negative and equals $-K_t$. Indeed, because they have no bequest motive, the old consume all they have and leave nothing as bequests. Hence, the young in any period enter the period with no inheritance, no non-human wealth. Consequently, any non-human wealth existing at the beginning of a period must belong to the old in that period and be the result of their saving as young in the previous period. As K_t constitutes the aggregate non-human wealth in our closed economy at the beginning of period t, we therefore have

$$s_{t-1}L_{t-1} = K_t. \tag{3.30}$$

Recalling that the net saving of any group is by definition the same as the increase in its non-human wealth, the net saving of the old in period t is $-K_t$. Aggregate net saving in the economy is thus $s_t L_t + (-K_t)$, and (3.29) is thereby explained.

DEFINITION 2 Let a given period t have capital stock $K_t \geq 0$, labor supply $L_t > 0$, and hence capital-labor ratio $k_t = K_t/L_t$. Let the expected real interest rate be given as $r_{t+1}^e > -1$. And let the functions $s(w_t, r_{t+1}^e)$, $w(k_t)$, and $r(k_t)$ be defined as in Lemma 1, (3.25), and (3.24), respectively. Then a *temporary equilibrium* in period t is a state $(k_t, c_{1t}, c_{2t}, w_t, r_t)$ of the economy such that (3.22), (3.23), (3.28), and (3.29) hold (i.e., all markets clear) for $c_{1t} = w_t - s_t$, $c_{2t} = (k_t + r(k_t)k_t)(1+n)$, where $w_t = w(k_t) > 0$ and $s_t = s(w_t, r_{t+1}^e)$.

The reason for the requirement $w_t > 0$ in the definition is that if $w_t = 0$, people would have nothing to live on as young and nothing to save from for retirement. The system would not be economically viable in this case. With regard to the equation for c_{2t} in the definition, note that (3.30) gives $s_{t-1} = K_t/L_{t-1} = (K_t/L_t)(L_t/L_{t-1}) = k_t(1+n)$, which is the wealth of each old at the beginning of period t. Substituting into $c_{2t} = (1+r_t)s_{t-1}$, we get $c_{2t} = (1+r_t)k_t(1+n)$, which can also be written $c_{2t} = (k_t + r_t k_t)(1+n)$. This last way of writing c_{2t} has the advantage of being applicable even if $k_t = 0$, cf. Technical Remark in Section 3.4. The remaining conditions for a temporary equilibrium are self-explanatory.

PROPOSITION 1 Suppose the No Fast Assumption (A1) applies. Consider a given period t with a given $k_t \geq 0$. Then for any $r_{t+1}^e > -1$,
(i) if $k_t > 0$, there exists a temporary equilibrium, $(k_t, c_{1t}, c_{2t}, w_t, r_t)$, and c_{1t} and c_{2t} are positive;
(ii) if $k_t = 0$, a temporary equilibrium exists if and only if capital is not essential; in that case, $w_t = w(k_t) = w(0) = f(0) > 0$ and c_{1t} and s_t are positive (while $c_{2t} = 0$);

(iii) whenever a temporary equilibrium exists, it is unique.

Proof. We begin with (iii). That there is at most one temporary equilibrium is immediately obvious since w_t and r_t are functions of the given $k_t : w_t = w(k_t)$ and $r_t = r(k_t)$. And given w_t, r_t, and r^e_{t+1}, c_{1t} and c_{2t} are uniquely determined.

(i) Let $k_t > 0$. Then, by (3.25), $w(k_t) > 0$. We claim that the state $(k_t, c_{1t}, c_{2t}, w_t, r_t)$, with $w_t = w(k_t)$, $r_t = r(k_t)$, $c_{1t} = w(k_t) - s(w(k_t), r^e_{t+1})$, and $c_{2t} = (1 + r(k_t))k_t(1 + n)$, is a temporary equilibrium. Indeed, Section 3.4 showed that the factor prices $w_t = w(k_t)$ and $r_t = r(k_t)$ are consistent with clearing in the factor markets in period t. Given that these markets clear (by price adjustment), it follows by Walras' law (see Appendix C) that also the third market, the goods market, clears in period t. So all criteria in Definition 2 are satisfied. That $c_{1t} > 0$ follows from $w(k_t) > 0$ and the No Fast Assumption (A1), in view of Lemma 1. That $c_{2t} > 0$ follows from $c_{2t} = (1 + r(k_t))k_t(1 + n)$ when $k_t > 0$, since $r(k_t) > -1$ always.

(ii) Let $k_t = 0$. Suppose $f(0) > 0$. Then, by Technical Remark in Section 3.4, $w_t = w(0) = f(0) > 0$ and $c_{1t} = w_t - s(w_t, r^e_{t+1})$ is well-defined, positive, and less than w_t, in view of Lemma 1; so $s_t = s(w_t, r^e_{t+1}) > 0$. The old in period 0 will starve since $c_{2t} = (0 + 0)(1 + n)$, in view of $r(0) \cdot 0 = 0$, cf. Technical Remark in Section 3.4. Even though this is a bad situation for the old, it is consistent with the criteria in Definition 2. On the other hand, if $f(0) = 0$, we get $w_t = f(0) = 0$, which violates one of the criteria in Definition 2. □

Point (ii) of the proposition says that a temporary equilibrium *may* exist even in a period where $k = 0$. The old in this period will starve and not survive. But if capital is not essential, the young get positive labor income out of which they will save a part for their old age and be able to maintain life also next period which will be endowed with positive capital. Then, by our assumptions the economy is viable forever.[15]

Generally, the term "equilibrium" is used to denote a state of "rest", often just "temporary rest". The temporary equilibrium in the present model is an example of a state of "temporary rest" in the following sense: (a) the agents optimize, given their expectations and the constraints they face; and (b) the aggregate demands and supplies in the given period are mutually consistent, i.e., markets clear. The qualification "temporary" is motivated by two features. First, in the next period the conditioning circumstances may be different, possibly as a consequence of the currently chosen actions in the aggregate. Second, the given expectations may turn out wrong.

[15] For simplicity, the model ignores that in practice a certain minimum per capita consumption level (the subsistence minimum) is needed for viability.

3.5.3 An equilibrium path

The concept of an *equilibrium path*, also called an *intertemporal equilibrium* or *dynamic general equilibrium*, requires more conditions satisfied. The concept refers to a sequence of temporary equilibria such that *expectations* of the agents are *fulfilled* in every period:

DEFINITION 3 An *equilibrium path* is a technically feasible path $\{(k_t, c_{1t}, c_{2t})\}_{t=0}^{\infty}$ such that for $t = 0, 1, 2, \ldots$, the state $(k_t, c_{1t}, c_{2t}, w_t, r_t)$ is a temporary equilibrium with $r_{t+1}^e = r(k_{t+1})$.

To characterize such a path, we forward (3.30) one period and rearrange so as to get

$$K_{t+1} = s_t L_t. \tag{3.31}$$

Since $K_{t+1} \equiv k_{t+1} L_{t+1} = k_{t+1} L_t (1+n)$, this can be written

$$k_{t+1} = \frac{s(w(k_t), r(k_{t+1}))}{1+n}, \tag{3.32}$$

using that $s_t = s(w_t, r_{t+1}^e)$, $w_t = w(k_t)$, and $r_{t+1}^e = r_{t+1} = r(k_{t+1})$ in a sequence of temporary equilibria with fulfilled expectations. Equation (3.32) is a first-order difference equation, known as the *fundamental difference equation* or the *law of motion* of the Diamond model.

PROPOSITION 2 Suppose the No Fast Assumption (A1) applies. Then,
(i) for any $k_0 > 0$ there exists at least one equilibrium path;
(ii) if $k_0 = 0$, an equilibrium path exists if and only if $f(0) > 0$ (i.e., capital not essential);
(iii) in any case, an equilibrium path has $w_t > 0$, $t = 0, 1, 2, \ldots$, and $k_t > 0$, $t = 1, 2, 3, \ldots$;
(iv) an equilibrium path satisfies the first-order difference equation (3.32).

Proof. (i) and (ii): see Appendix D. (iii) For a given t, let $k_t \geq 0$. Then, since an equilibrium path is a sequence of temporary equilibria, we have, from Proposition 1, $w_t = w(k_t) > 0$ and $s_t = s(w(k_t), r_{t+1}^e)$, where $r_{t+1}^e = r(k_{t+1})$. Hence, by Lemma 1, $s(w(k_t), r_{t+1}^e) > 0$, which implies $k_{t+1} > 0$, in view of (3.32). This shows that only for $t = 0$ is $k_t = 0$ possible along an equilibrium path. (iv) This was shown in the text above. □

The formal proofs of point (i) and (ii) of the proposition are quite technical and placed in the appendix. But the graphs in the ensuing figures 3.4-3.7 provide an intuitive verification. The "only if" part of point (ii) reflects the not very surprising fact that *if* capital were an essential production factor, no capital "now" would imply no income "now", hence no saving and investment and thus no capital in the next period and so on. On the other hand, the "if" part of point (ii) says that when capital is not essential, an equilibrium path can set off also from an initial period with no capital. Then point (iii) adds that an equilibrium path will have positive capital in all subsequent periods. Finally, as to point (iv), note that the fundamental difference equation, (3.32), rests on equation (3.31). Recall from the previous subsection that the economic logic behind this key equation is that since capital is the only non-human asset in the economy and the young are born without any inheritance, the aggregate capital stock at the beginning of period $t+1$ *must* be owned by the old generation in that period. It must thereby equal the aggregate saving these people had in the previous period where they were young.

The transition diagram

To be able to further characterize equilibrium paths, we construct a transition diagram in the (k_t, k_{t+1}) plane. The *transition curve* is defined as the set of points (k_t, k_{t+1}) satisfying (3.32). Its form and position depends on the households' preferences and the firms' technology. Fig. 3.4 shows one *possible*, but far from necessary configuration of this curve. A complicating circumstance is that the equation (3.32) has k_{t+1} on both sides. Sometimes we are able to solve the equation explicitly for k_{t+1} as a function of k_t, but sometimes we can do so only implicitly. What is even worse is that there are cases where k_{t+1} is not unique for a given k_t. We will proceed step by step.

First, what can we say about the *slope* of the transition curve? In general, a point on the transition curve has the property that at least in a small neighborhood of this point, the equation (3.32) will define k_{t+1} as an implicit function of k_t.[16] Taking the derivative with respect to k_t on both sides of (3.32), we get

$$\frac{dk_{t+1}}{dk_t} = \frac{1}{1+n}\left(s_w w'\left(k_t\right) + s_r r'\left(k_{t+1}\right)\frac{dk_{t+1}}{dk_t}\right). \tag{3.33}$$

By ordering, the slope of the transition curve within this small neighborhood can be written

$$\frac{dk_{t+1}}{dk_t} = \frac{s_w\left(w\left(k_t\right), r\left(k_{t+1}\right)\right) w'\left(k_t\right)}{D(k_t, k_{t+1})}, \tag{3.34}$$

when the denominator,

$$D(k_t, k_{t+1}) \equiv 1 + n - s_r(w(k_t), r(k_{t+1}))r'(k_{t+1}),$$

differs from nil.

[16] An exception occurs if the denominator in (3.34) below vanishes.

In view of $s_w > 0$ and $w'(k_t) = -k_t\ f''(k_t) > 0$, the numerator in (3.34) is always positive and we have

$$\frac{dk_{t+1}}{dk_t} \gtrless 0 \text{ for } s_r(w(k_t), r(k_{t+1})) \gtrless \frac{1+n}{r'\left(k_{t+1}\right)},$$

respectively, since $r'\left(k_{t+1}\right) = f''(k_{t+1}) < 0$.

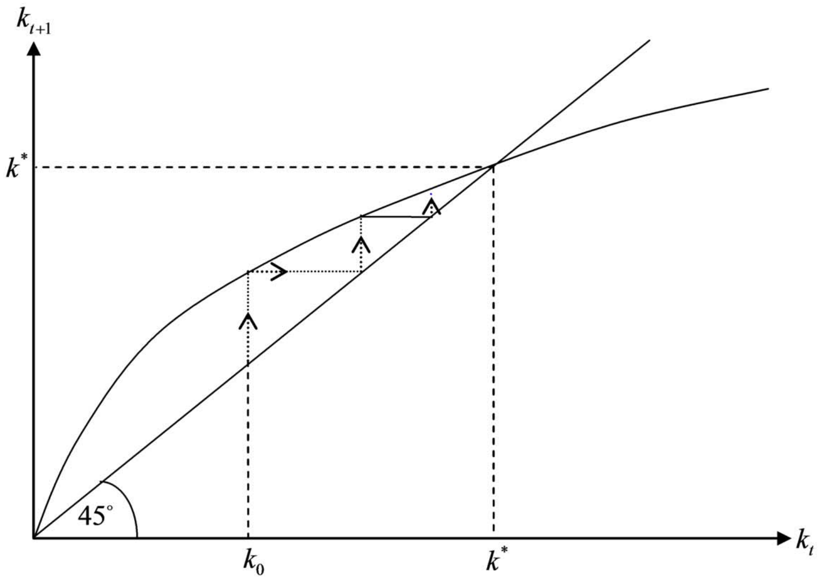


Figure 3.4: Transition curve and the resulting dynamics when $u(c) = \ln c$ and $Y = AK^{\alpha}L^{1-\alpha}$, $0 < \alpha < 1$, cf. Example 2.

It follows that the transition curve is universally upward-sloping if and only if $s_r(w(k_t), r(k_{t+1})) > (1+n)/r'\left(k_{t+1}\right)$ everywhere along the transition curve. The intuition behind this finding becomes visible by rewriting (3.34) in terms of small changes in k_t and k_{t+1}. Since $\Delta k_{t+1}/\Delta k_t \approx dk_{t+1}/dk_t$ for Δk_t "small", (3.34) implies

$$\left[1 + n - s_r\left(\cdot\right) r'\left(k_{t+1}\right)\right] \Delta k_{t+1} \approx s_w\left(\cdot\right)\ w'(k_t)\Delta k_t. \qquad (*)$$

Let $\Delta k_t > 0$. This rise in k_t will always raise wage income and, via the resulting rise in s_t, raise k_{t+1}, everything else equal. Everything else is *not* equal, however, since a rise in k_{t+1} implies a fall in the rate of interest. There are four cases to consider:

Case 1: $s_r\left(\cdot\right) = 0$. Then there is no feedback effect from the fall in the rate of interest. So the tendency to a rise in k_{t+1} is neither offset nor fortified.

Case 2: $s_r\left(\cdot\right) > 0$. Then the tendency to a rise in k_{t+1} will be partly offset through the *dampening* effect on saving resulting from the fall in the interest rate. This negative feedback can not fully offset the tendency to a rise in k_{t+1}. The reason is that the negative feedback on the saving of the young will only be there *if* the interest rate falls in the first place. We cannot in a period have both a *fall* in the interest rate triggering lower saving *and* a *rise* in the interest rate (via a lower k_{t+1}) *because* of the lower saving. So a *sufficient* condition for a universally upward-sloping transition curve is that the saving of the young is a non-decreasing function of the interest rate.

Case 3: $(1+n)/r'(k_{t+1}) < s_r(\cdot) < 0$. Then the tendency to a rise in k_{t+1} will be fortified through the *stimulating* effect on saving resulting from the fall in the interest rate.

Case 4: $s_r(\cdot) < (1+n)/r'(k_{t+1}) < 0$. Then the expression in brackets on the left-hand side of (*) is negative and requires therefore that $\Delta k_{t+1} < 0$ in order to comply with the positive right-hand side. This is a situation where self-fulfilling expectations operate, a case to which we return. We shall explore this case in the next sub-section.

Another feature of the transition curve is the following:

LEMMA 2 *(the transition curve is nowhere flat)* For all $k_t > 0$ such that the denominator, $D(k_t, k_{t+1})$, in (3.34) differs from nil, we have $dk_{t+1}/dk_t \neq 0$.

Proof. Since $s_w > 0$ and $w'(k_t) > 0$ always, the numerator in (3.34) is always positive. $\square$

The implication is that no part of the transition curve can be horizontal.[17]

When the transition curve crosses the 45° degree line for some $k_t > 0$, as in the example in Fig. 3.4, we have a steady state at this k_t. Formally:

DEFINITION 4 An equilibrium path $\{(k_t, c_{1t}, c_{2t})\}_{t=0}^{\infty}$ is in a *steady state* with capital-labor ratio $k^* > 0$ if the fundamental difference equation, (3.32), is satisfied with k_t as well as k_{t+1} replaced by k^*.

This exemplifies the notion of a steady state as a stationary point in a dynamic process. Some economists use the term "dynamic equilibrium" instead of "steady state". In this text the term "equilibrium" refers to situations where the constraints and decided actions of the market participants are mutually compatible. So, an economy can be in "equilibrium" without being in a steady state. We see a steady state as a *special* sequence of temporary equilibria with fulfilled expectations, namely one with the property that the endogenous variable, here k, entering the fundamental difference equation does not change over time.

EXAMPLE 2 (the log utility Cobb-Douglas case) Let $u(c) = \ln c$ and $Y = AK^{\alpha}L^{1-\alpha}$, where $A > 0$ and $0 < \alpha < 1$. Since $u(c) = \ln c$ is the case $\theta = 1$

[17] This would not generally hold if the utility function were not time-separable.

in Example 1, by (3.15) we have $s_r = 0$. Indeed, with logarithmic utility the substitution and income effects on s_t of a rise in the interest rate offset each other; and, as discussed above, in the Diamond model there can be no wealth effect of a rise in r_{t+1}. Further, the equation (3.32) reduces to a transition *function,*

$$k_{t+1} = \frac{(1-\alpha)Ak_t^\alpha}{(1+n)(2+\rho)}. \tag{3.35}$$

The associated transition curve is shown in Fig. 3.4 and there is for $k_0 > 0$ both a unique equilibrium path and a unique steady state with capital-labor ratio

$$k^* = \left(\frac{(1-\alpha)A}{(2+\rho)(1+n)}\right)^{1/(1-\alpha)} > 0.$$

At $k_t = k^*$, the slope of the transition curve is necessarily less than one. The dynamics therefore lead to convergence to the steady state as illustrated in the figure.[18] In the steady state the interest rate is $r^* = f'(k^*) - \delta = \alpha(1+n)(2+\rho)/(1-\alpha) - \delta$. Note that a higher n results in a lower k^*, hence a higher r^*. $\square$

Because the Cobb-Douglas production function implies that capital is essential, (3.35) implies $k_{t+1} = 0$ if $k_t = 0$. The state $k_{t+1} = k_t = 0$ is thus a stationary point of the difference equation (3.35) considered in isolation. This state is not, however, an equilibrium path as defined above (not a steady state of an *economic* system since there is no production). We may call it a *trivial* steady state in contrast to the economically viable steady state $k_{t+1} = k_t = k^* > 0$ which is then called a *non-trivial* steady state.

Theoretically, there may be more than one (non-trivial) steady state. Nonexistence of a steady state is also possible. But before considering these possibilities, the next subsection (which may be skipped in a first reading) addresses an even more defiant feature which is that for a given k_0 there may exist more than one equilibrium path.

The possibility of multiple equilibrium paths*

It turns out that a *backward-bending* transition curve like that in Fig. 3.5 is possible within the model. Not only are there two steady states but for $k_t \in (\underline{k}, \overline{k})$ there are *three temporary equilibria* with self-fulfilling expectations. That is, for a given k_t in this interval, there are three different values of k_{t+1} that are consistent with self-fulfilling expectations. Exercise 3.3 at the end of the chapter documents this possibility by way of a numerical example.

[18] A formal proof can be based on the mean value theorem.

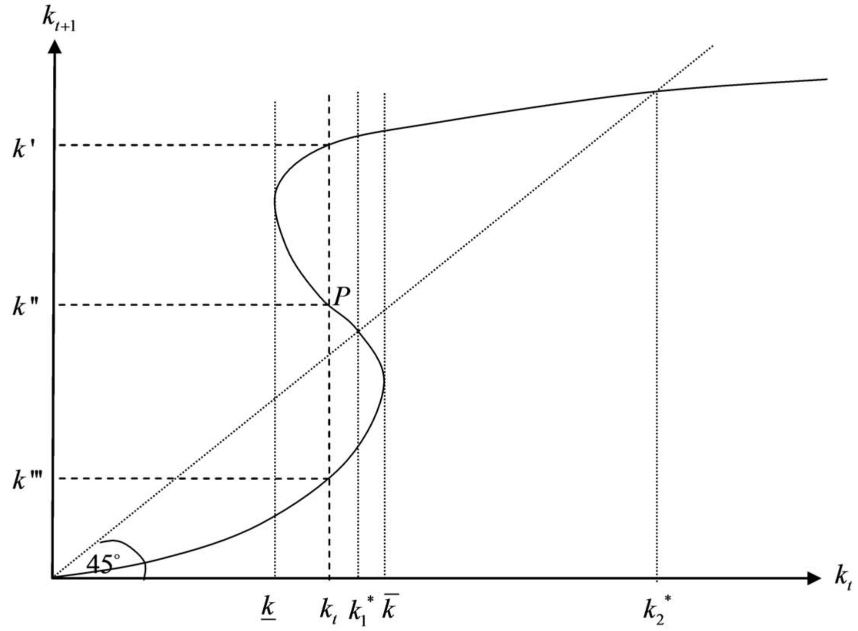


Figure 3.5: A backward-bending transition curve leads to multiple temporary equilibria with self-fulfilling expectations.

The theoretical possibility of multiple equilibria with self-fulfilling expectations requires that there is at least one interval on the horizontal axis where a section of the transition curve has negative slope. What is the intuition behind that in this situation multiple equilibria can arise? Consider the specific configuration in Fig. 3.5 where k', k'', and k''' are the possible values of the capital-labor ratio next period when $k_t \in (\underline{k}, \bar{k})$. In a small neighborhood of the point P associated with the intermediate value, k'', the slope of the transition curve is negative. In the figure such a neighborhood is represented by the rectangle R. Within this rectangle the fundamental difference equation (3.32) does indeed define k_{t+1} as an implicit function of k_t, the graph of which goes through the point P and has negative slope.

Now, as we saw above, the negative slope requires not only that in this neighborhood $s_r(w_t, r(k_{t+1})) < 0$, but that the stricter condition $s_r(w_t, r(k_{t+1})) < (1+n)/f''(k'')$ holds (we take w_t as given since k_t is given and $w_t = w(k_t)$). That the point P with coordinates (k_t, k'') is on the transition curve indicates that, given $w_t = w(k_t)$ and an expected interest rate $r_{t+1}^e = r(k'')$, the induced saving by the young, $s(w_t, r(k'')$, will be such that $k_{t+1} = k''$. Then the expectation is fulfilled. But also the point (k_t, k'), where $k' > k''$, is on transition curve and this reflects that also a lower interest rate, $r(k')$, can be self-fulfilling.

By this is meant that *if* an interest rate at the level $r(k')$ is generally expected, then this expectation induces *more* saving by the young, just enough more to make $k_{t+1} = k' > k''$, thus confirming the expectation of the lower interest rate, $r(k')$. What makes this possible is exactly the negative dependency of s_t on r^e_{t+1}. The fact that also the point (k_t, k'''), satisfying the inequality $k''' < k''$, is on the transition curve has a similar interpretation. It is exactly $s_r < 0$ that makes it possible that *less* saving by the young than at the point P can be induced by an expected *higher* interest rate, $r(k''')$, than at P.

Recognizing the ambiguity arising from the possibility of multiple equilibrium paths, we face an additional ambiguity, known as the "expectational coordination problem". The model presupposes that all the young *agree* in their expectations. Only then will one of the three mentioned temporary equilibria appear. But the model is silent about how the needed coordination of expectations is brought about, and if it is, why this coordination ends up in one rather than another of the three possible equilibria with self-fulfilling expectations. Each single young is isolated in the market and will not know what the others will expect. The market mechanism by itself provides no coordination of expectations.

As it stands, the model consequently cannot determine how the economy will evolve in the present situation with a backward-bending transition curve. It is not uncommon that macroeconomic analysis runs into such difficulties. This kind of difficulties reflect the complexity of an economic system. At this stage we will circumvent the indeterminacy problem by taking an ad-hoc approach. There are at least three ways to try to rule out the possibility of multiple equilibrium paths. One way is to discard the assumption of perfect foresight. Instead, some kind of adaptive expectations may be assumed, for example in the form of *myopic foresight,* also called *static expectations*. This means that the expectation formed by the agents in the current period about the value of a variable next period is that it will stay the same as in the current period. So here the assumption would be that the young have the expectation $r^e_{t+1} = r_t$. Then, given $k_0 > 0$, a *unique* sequence of temporary equilibria $\{(k_t, c_{1t}, c_{2t}, w_t, r_t)\}_{t=0}^{\infty}$ is generated by the model. *Oscillations* in the sense of repetitive movements up and down of k_t are possible. Even *chaotic* trajectories are possible (see Exercise 3.6).

Outside steady state, when agents have static expectations, they will often experience that their expectations are systematically wrong. And the assumption of myopic foresight rules out that learning occurs. This may be too simplistic, although it *can* be argued that human beings to a certain extent have a psychological disposition to myopic foresight.

Another approach to the indeterminacy problem could be motivated by the general observation that sometimes the possibility of multiple equilibria in a model arises because of a "rough" time structure imposed on the model in question. In the present case, each period in the Diamond model corresponds to half of an adult person's lifetime. And in the first period of life there is no capital income, in the second there is no labor income. This coarse notion of time may artificially generate a multiplicity of equilibria or, with myopic foresight, oscillations. An expanded model where people live many periods may "smooth" the responses of the system to the events impinging on it. Indeed, with working life stretching over more than one period, wealth effects of changes in the interest rate arise, thereby reducing the likelihood of a backward-bending transition curve. In Chapter 12 we shall see an example of an overlapping-generations model in continuous time where the indeterminacy problem never arises.

For now, our approach will be to stay with the rough time structure of the Diamond model because of its analytical convenience and then make the best of it by imposing conditions on the utility function, the production function, and/or parameter values so as to rule out multiple equilibria. We stay with the assumption of perfect foresight, but assume that circumstances are such that multiple equilibria with self-fulfilling expectations do not arise. Fortunately, the "circumstances" needed for this in the present model are not defying empirical plausibility.

Conditions for uniqueness of the equilibrium path

Sufficient for the equilibrium path to be unique is that preferences and technology in combination are such that the slope of the transition curve is everywhere positive. Hence we impose the Positive Slope Assumption that

$$s_r(w(k_t), r(k_{t+1})) > \frac{1+n}{f''(k_{t+1})} \tag{A2}$$

for all pairs (k_t, k_{t+1}) consistent with an equilibrium path. This condition is of course always satisfied when $s_r \geq 0$ (reflecting an elasticity of marginal utility of consumption not above one) and *can* be satisfied even if $s_r < 0$ (as long as s_r is "small" in absolute value). Essentially, (A2) is an assumption that the income effect on consumption as young of a rise in the interest rate does not dominate the substitution effect "too much".

Unfortunately, when stated as in (A2), this condition is not as informative as we might wish. a condition like (A2) is not in itself very informative. This is because it is expressed in terms of an *endogenous* variable, k_{t+1}, for given k_t. A model assumption should preferably be stated in terms of what is *given,* also called the "primitives" of the model; in this model the "primitives" comprise the given preferences, demography, technology, and market form. We can state *sufficient* conditions, however, in terms of the "primitives", such that (A2) is ensured. Here we state two such sufficient conditions, both involving a CRRA period utility function with parameter θ as defined in (3.14):

(a) If $0 < \theta \leq 1$, then (A2) holds for all $k_t > 0$ along an equilibrium path.

(b) If the production function is of CES-type with CRS,[19] i.e., $f(k) = A(\alpha k^\gamma + 1 - \alpha)^{1/\gamma}$, $A > 0$, $0 < \alpha < 1$, $-\infty < \gamma < 1$, then (A2) holds along an equilibrium path even for $\theta > 1$, if the elasticity of substitution between capital and labor, $1/(1-\gamma)$, is not "too small", i.e., if

$$\frac{1}{1-\gamma} > \frac{1 - 1/\theta}{1 + (1+\rho)^{-1/\theta}(1 + f'(k) - \delta)^{(1-\theta)/\theta}} \tag{3.36}$$

for all $k > 0$. In turn, sufficient for this is that $(1-\gamma)^{-1} > 1 - \theta^{-1}$.

That (a) is sufficient for (A2) is immediately visible in (3.15). The sufficiency of (b) is proved in Appendix D. The elasticity of substitution between capital and labor is a concept analogue to the elasticity of intertemporal substitution in consumption (Section 3.3). It is a measure of the sensitivity of the chosen $k = K/L$ with respect to the relative factor price. The next chapter goes more into detail with the concept and shows, among other things, that the Cobb-Douglas production function corresponds to $\gamma = 0$. So the Cobb-Douglas production function will satisfy the inequality $(1-\gamma)^{-1} > 1 - \theta^{-1}$ (since $\theta > 0$), hence also the inequality (3.36).

With these or other sufficient conditions in the back of our mind we shall now proceed imposing the Positive Slope Assumption (A2). To summarize:

PROPOSITION 3 *(uniqueness of an equilibrium path)* Suppose the No Fast and Positive Slope assumptions, (A1) and (A2), apply. Then:
(i) if $k_0 > 0$, there exists a unique equilibrium path;
(ii) if $k_0 = 0$, an equilibrium path exists if and only if $f(0) > 0$ (i.e., capital not essential).

When the conditions of Proposition 3 hold, the fundamental difference equation, (3.32), of the model defines k_{t+1} as an implicit *function* of k_t,

$$k_{t+1} = \varphi(k_t), \tag{3.37}$$

for all $k_t > 0$, where $\varphi(k_t)$ is called a *transition function.* The derivative of this implicit function is given by (3.34) with k_{t+1} on the right-hand side replaced by $\varphi(k_t)$, i.e.,

$$\varphi'(k_t) = \frac{s_w\left(w\left(k_t\right), r\left(\varphi(k_t)\right)\right) w'(k_t)}{1 + n - s_r\left(w\left(k_t\right), r\left(\varphi(k_t)\right)\right) r'(\varphi(k_t))} > 0. \tag{3.38}$$

[19] CES stands for Constant Elasticity of Substitution. The CES production function was briefly considered in Section 2.1 and is considered in detail in Chapter 4.

The positivity for all $k_t > 0$ is due to (A2). A specific transition function is shown in Example 2 above.

By simple iteration for $t = 0, 1, 2, \ldots$,.a transition function gives us the evolution of k_t. Then we have in fact determined the evolution of "everything" in the economy: the factor prices $w(k_t)$ and $r(k_t)$, the saving of the young $s_t = s(w(k_t), r(k_{t+1}))$, and the consumption by both the young and the old. The mechanism behind the evolution of the economy is the Walrasian (or Classical) mechanism where prices, here w_t and r_t, always adjust so as to generate market clearing as if there were a Walrasian auctioneer and where expectations always adjust so as to be model consistent.

Existence and stability of a steady state?

Possibly the equilibrium path converges to a steady state. To address this issue, we examine the possible configurations of the transition curve in more detail. In addition to being positively sloped everywhere, the transition curve will always, for $k_t > 0$, be situated strictly below the solid curve, $k_{t+1} = w(k_t)/(1+n)$, shown in Fig. 3.6. In turn, the latter curve is always, for $k_t > 0$, strictly below the stippled curve, $k_{t+1} = f(k_t)/(1+n)$, in the figure. To be precise:

LEMMA 3 *(ceiling)* Suppose the No Fast Assumption (A1) applies. Along an equilibrium path, whenever $k_t > 0$,

$$0 < k_{t+1} < \frac{w(k_t)}{1+n} < \frac{f(k_t)}{1+n}, \qquad t = 0, 1, \ldots. \tag{*}$$

Proof. From (iii) of Proposition 2, an equilibrium path has $w_t = w(k_t) > 0$ and $k_{t+1} > 0$ for $t = 0, 1, 2, \ldots$. Thus,

$$0 < k_{t+1} = \frac{s_t}{1+n} < \frac{w_t}{1+n} = \frac{w(k_t)}{1+n} = \frac{f(k_t) - f'(k_t)k_t}{1+n} < \frac{f(k_t)}{1+n},$$

where the first equality comes from (3.32), the second inequality from Lemma 1 in Section 3.3, and the last inequality from the fact that $f'(k_t)k_t > 0$ when $k_t > 0$. This proves (*). □

We will call the graph $(k_t, w(k_t)/(1+n))$ in Fig. 3.6 a *ceiling*. It acts as a ceiling on k_{t+1} simply because the saving of the young cannot exceed the income of the young, $w(k_t)$. The stippled graph, $(k_t, f(k_t)/(1+n))$, in Fig. 3.6 represents what we name the *roof* ("everything of interest" occurs below it). While the ceiling is the key concept in the proof of Proposition 4 below, the roof is a more straightforward construct since it is directly given by the production function and is always strictly concave. The roof is always above the ceiling and so it appears as a convenient first "enclosure" of the transition curve. Let us therefore start with a characterization of the roof:

LEMMA 4 The roof, $R(k) \equiv f(k)/(1+n)$, has positive slope everywhere, crosses the 45° line for at most one $k > 0$ and can only do that from above. A necessary and sufficient condition for the roof to be above the 45° line for small k is that either $\lim_{k\to 0} f'(k) > 1 + n$ or $f(0) > 0$ (capital not essential).

Proof. Since $f' > 0$, the roof has positive slope. Since $f'' < 0$, it can only cross the 45° line once and only from above. If and only if $\lim_{k\to 0} f'(k) > 1 + n$, then for small k_t, the roof is steeper than the 45° line. Obviously, if $f(0) > 0$, then close to the origin, the roof will be above the 45° line. □

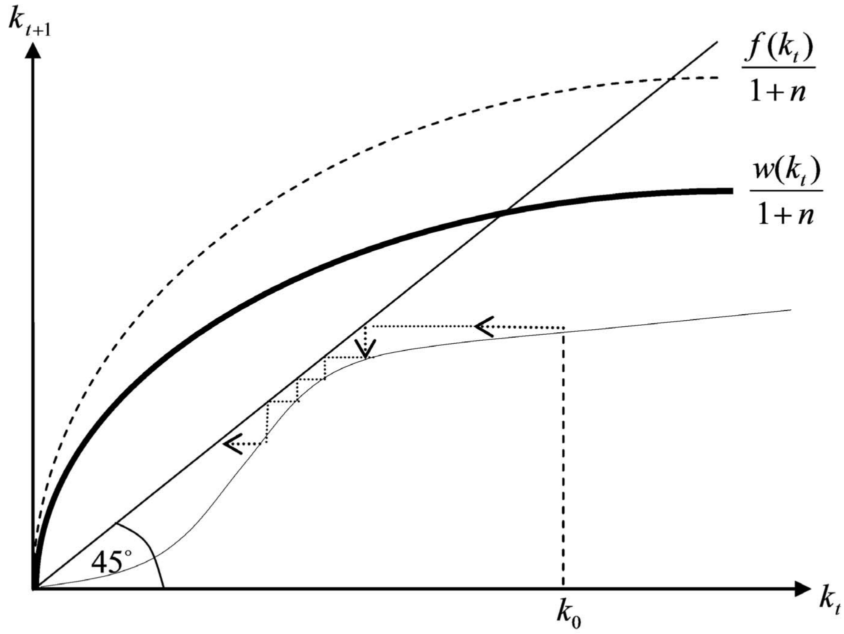


Figure 3.6: A case where both the roof and the ceiling cross the 45° line, but the transition curve does not (no steady state exists).

The ceiling is generally a more complex construct. It can have convex sections and for instance cross the 45° line at more than one point if at all. While the roof can be above the 45° line for *all* $k_t > 0$, the ceiling cannot. Indeed, (ii) of the next lemma implies that if for small k_t the ceiling is above the 45° line, the ceiling will necessarily cross the 45° line at least once for larger k_t.

LEMMA 5 Given $w(k) = f(k) - f'(k)k$ for all $k \geq 0$, where $f(k)$ satisfies $f(0) \geq 0$, $f' > 0$, $f'' < 0$, the following holds:
(i) $\lim_{k\to\infty} w(k)/k = 0$;

(ii) the ceiling, $C(k) \equiv w(k)/(1+n)$, is positive and has positive slope for all $k > 0$; moreover, there exists $\bar{k} > 0$ such that $C(k) < k$ for all $k > \bar{k}$.

Proof. (i) In view of $f(0) \geq 0$ combined with $f'' < 0$, we have $w(k) > 0$ for all $k > 0$. Hence, $\lim_{k\to\infty} w(k)/k \geq 0$ if this limit exists. Consider an arbitrary $k_1 > 0$. We have $f'(k_1) > 0$. For all $k > k_1$, it holds that $0 < f'(k) < f'(k_1)$, in view of $f' > 0$ and $f'' < 0$, respectively. Hence, $\lim_{k\to\infty} f'(k)$ exists and

$$0 \leq \lim_{k\to\infty} f'(k) < f'(k_1). \tag{3.39}$$

We have

$$\lim_{k\to\infty} \frac{w(k)}{k} = \lim_{k\to\infty} \frac{f(k)}{k} - \lim_{k\to\infty} f'(k). \tag{3.40}$$

There are two cases to consider. *Case 1:* $f(k)$ has an upper bound. Then, $\lim_{k\to\infty} f(k)/k = 0$ so that $\lim_{k\to\infty} w(k)/k = -\lim_{k\to\infty} f'(k) = 0$, by (3.40) and (3.39), as $w(k)/k > 0$ for all $k > 0$. *Case 2:* $\lim_{k\to\infty} f(k) = \infty$. Then, by L'Hôpital's rule for "∞/∞", $\lim_{k\to\infty}(f(k)/k) = \lim_{k\to\infty} f'(k)$ so that (3.40) implies $\lim_{k\to\infty} w(k)/k = 0$.

(ii) As $n > -1$ and $w(k) > 0$ for all $k > 0$, $C(k) > 0$ for all $k > 0$. From $w'(k) = -kf''(k) > 0$ follows $C'(k) = -kf''(k)/(1+n) > 0$ for all $k > 0$; that is, the ceiling has positive slope everywhere. For $k > 0$, the inequality $C(k) < k$ is equivalent to $w(k)/k < 1+n$. By (i) follows that for all $\varepsilon > 0$, there exists $k_\varepsilon > 0$ such that $w(k)/k < \varepsilon$ for all $k > k_\varepsilon$. Now, letting $\varepsilon = 1+n$ and $\bar{k} = k_\varepsilon$ proves that there exists $\bar{k} > 0$ such that $w(k)/k < 1+n$ for all $k > \bar{k}$. □

A necessary condition for existence of a (non-trivial) steady state is that the roof is above the 45^0 line for small k_t. But this is not sufficient for also the transition curve to be above the 45^0 line for small k_t. Fig. 3.6 illustrates this. Here the transition curve is in fact everywhere below the 45^0 line. In this case no steady state exists and the dynamics imply convergence towards the "catastrophic" point $(0,0)$. Given the rate of population growth, the saving of the young is not sufficient to avoid famine in the long run. This outcome will occur if the technology implies so low productivity that even when all income of the young were saved, we would have $k_{t+1} < k_t$ for all $k_t > 0$, cf. Exercise 3.2. The Malthusian mechanism will be at work and bring down n (outside the model). This exemplifies that even a trivial steady state (the point (0,0)) may be of interest in so far as it may be the point the economy is heading to (though never reaching it).

To help existence of a steady state we will impose the condition that either capital is not essential or preferences and technology fit together in such a way that the slope of the transition curve is larger than 1 for small k_t. That is, we assume that either

$$\begin{array}{lrcl} \text{(i)} & f(0) & > & 0 \quad \text{or} \\ \text{(ii)} & \lim_{k\to 0}\varphi'(k) & > & 1, \end{array} \tag{A3}$$

where $\varphi'(k)$ is implicitly given in (3.38). Whether condition (i) of (A3) holds in a given situation can be directly checked from the production function. If it does not, we should check condition (ii). But this condition is less amenable because the transition function φ is not one of the "primitives" of the model. There exist cases, though, where we can find an explicit transition function and try out whether (ii) holds (like in Example 2 above). But generally we can not. Then we have to resort to *sufficient* conditions for (ii) of (A3), expressed in terms of the "primitives". For example, if the period utility function belongs to the CRRA class and the production function is Cobb-Douglas at least for small k, then (ii) of (A3) holds (see Appendix E). Anyway, as (i) and (ii) of (A3) can be interpreted as reflecting two different kinds of "early steepness" of the transition curve, we shall call (A3) the Early Steepness Assumption.[20]

Before stating the proposition aimed at, we need a definition of the concept of asymptotic stability.

DEFINITION 5 Consider a first-order autonomous difference equation $x_{t+1} = g(x_t)$, $t = 0, 1, 2, \ldots$. A steady state, x^*, is (locally) *asymptotically stable* if there exists $\varepsilon > 0$ such that $x_0 \in (x^* - \varepsilon, x^* + \varepsilon)$ implies that $x_t \to x^*$ for $t \to \infty$. A steady state $x^* > 0$ is *globally asymptotically stable* if for all feasible $x_0 > 0$, it holds that $x_t \to x^*$ for $t \to \infty$.

Applying this definition on our difference equation $k_{t+1} = \varphi(k_t)$, we have:

PROPOSITION 4 *(existence and stability of a steady state)* Assume that the No Fast Assumption (A1) and the Positive Slope assumption (A2) apply as well as the Early Steepness Assumption (A3). Then there exists at least one steady state $k_1^* > 0$ that is asymptotically stable. If k_t does not converge to k_1^*, k_t converges to another steady state. If there is only one steady state, it is globally asymptotically stable. Oscillations never arise.

Proof. By (A1), Lemma 3 applies. From Proposition 2 we know that if (i) of (A3) holds, then $k_{t+1} = s_t/(1+n) > 0$ even for $k_t = 0$. Alternatively, (ii) of (A3) is enough to ensure that the transition curve lies above the 45° line for small k_t. According to (ii) of Lemma 5, for large k_t the ceiling is below the 45° line. Being below the ceiling, cf. Lemma 3, the transition curve must therefore cross the 45° line at least once. Let k_1^* denote the smallest k_t at which it crosses. Then $k_1^* > 0$

[20] In (i) of (A3), the "steepness" is at $k = 0$ rather a "hop" if we imagine k approaching zero from below.

is a steady state with the property $0 < \varphi'(k_1^*) < 1$. By graphical inspection we see that this steady state is asymptotically stable. If it is the only (non-trivial) steady state, we see it is globally asymptotically stable. Otherwise, if k_t does not converge to k_1^*, k_t converges to one of the other steady states. Indeed, divergence is ruled out since, by Lemma 5, there exists $\bar{k} > 0$ such that $w(k)/(1+n) < k$ for all $k > \bar{k}$ (Fig. 3.7 illustrates). For oscillations to come about there must exist a steady state, k^{**}, with $\varphi'(k^{**}) < 0$, but this is impossible in view of (A2). □

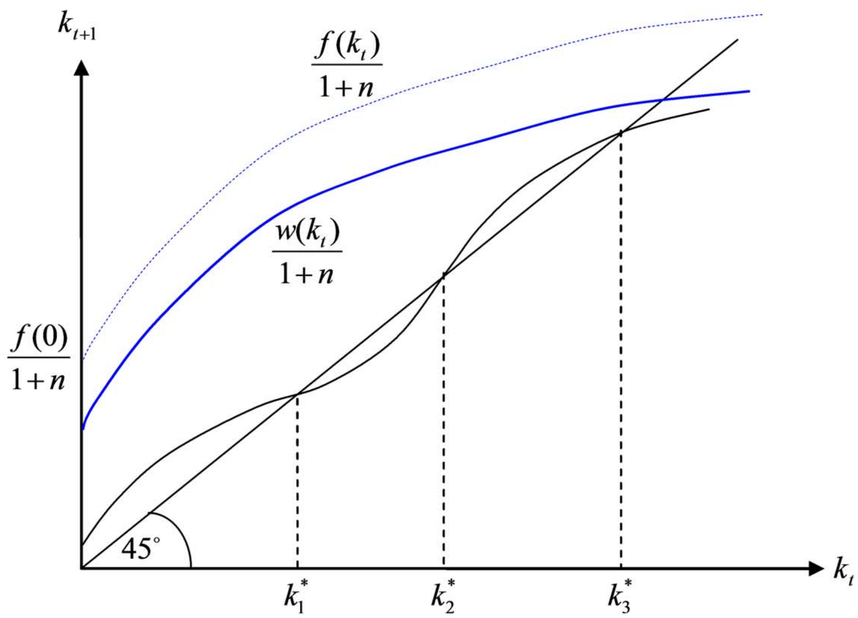


Figure 3.7: A case of multiple steady states (and capital being not essential).

From Proposition 4 we conclude that, given k_0, the assumptions (A1) - (A3) ensure existence and uniqueness of an equilibrium path; moreover, the equilibrium path converges towards *some* steady state. Thus with these assumptions, for any $k_0 > 0$, sooner or later the system settles down at some steady state $k^* > 0$. For the factor prices we therefore have

$$\begin{aligned} r_t &= f'(k_t) - \delta \to f'(k^*) - \delta \equiv r^*, \quad \text{and} \\ w_t &= f(k_t) - k_t f'(k_t) \to f(k^*) - k^* f'(k^*) \equiv w^*, \end{aligned}$$

for $t \to \infty$. But there may be more than one steady state and therefore only *local* stability is guaranteed. This can be shown by examples, where the utility function, the production function, and parameters are specified in accordance with the assumptions (A1) - (A3) (see Exercise 3.5 and ...).

Fig. 3.7 illustrates such a case (with $f(0) > 0$ so that capital is not essential). Moving West-East in the figure, the first steady state, k_1^*, is stable, the second, k_2^*, unstable, and the third, k_3^*, stable. In which of the two stable steady states the economy ends up depends on the initial capital-labor ratio, k_0. The lower steady state, k_1^*, is known as a *poverty trap*. If $0 < k_0 < k_2^*$, the economy is caught in the trap and converges to the low steady state. But with high enough k_0 ($k_0 > k_2^*$), perhaps obtained by foreign aid, the economy avoids the trap and converges to the high steady state. Looking back at Fig. 3.6, we can interpret that figure's scenario as exhibiting an *inescapable* poverty trap.

It turns out that CRRA utility combined with a Cobb-Douglas production function ensures both that (A1) - (A3) hold and that a *unique* (non-trivial) steady state exists. So in this case *global* asymptotic stability of the steady state is ensured.[21] Example 2 and Fig. 3.4 above display a special case of this, the case $\theta = 1$.

This is of course a convenient case for the analyst. A Diamond economy satisfying assumptions (A1) - (A3) *and* featuring a *unique* steady state is called a *well-behaved* Diamond economy.

We end this section with the question: Is it possible that aggregate consumption, along an equilibrium path, for some periods exceeds aggregate income? We shall see that this is indeed the case in this model if K_0 (wealth of the old in the initial period) is large enough. Indeed, from national accounting we have:

$$\begin{aligned} C_{10} + C_{20} &= F(K_0, L_0) - I_0 > F(K_0, L_0) \Leftrightarrow I_0 < 0 \\ &\Leftrightarrow K_1 < (1-\delta)K_0 \Leftrightarrow K_0 - K_1 > \delta K_0. \end{aligned}$$

So aggregate consumption in period 0 being greater than aggregate income is equivalent to a fall in the capital stock from period 0 to period 1 greater than the capital depreciation in period 0. Consider the log utility Cobb-Douglas case in Fig. 3.4 and suppose $\delta < 1$ and $L_t = L_0 = 1$, i.e., $n = 0$. Then $k_t = K_t$ for all t and by (3.35), $K_{t+1} = \frac{(1-\alpha)A}{2+\rho} K_t^{\alpha}$. Thus $K_1 < (1-\delta)K_0$ for

$$K_0 > \left(\frac{(1-\alpha)A}{(2+\rho)(1-\delta)} \right)^{1/(1-\alpha)}.$$

As initial K is arbitrary, this situation is possible. When it occurs, it reflects that the financial wealth of the old is so large that their consumption (recall they consume all their financial wealth as well as the interest on this wealth) exceeds what is left of current aggregate production after subtracting the amount consumed by the young. So aggregate gross investment in the economy will be negative. Of course this is only feasible if capital goods can be "eaten" or at least

[21] See last section of Appendix E.

be immediately (without further resources) converted into consumption goods. As it stands, the model has implicitly assumed this to be the case. And this is in line with the general setup since the output good is assumed homogeneous and can either be consumed or piled up as capital.

We now turn to efficiency problems.

3.6 The golden rule and dynamic inefficiency

An economy described by the Diamond model has the property that even though there is perfect competition and no externalities, the outcome brought about by the market mechanism may not be Pareto optimal.[22] Indeed, the economy may *overaccumulate* forever and thus suffer from a distinctive form of production inefficiency.

A key element in understanding the concept of overaccumulation is the concept of a *golden-rule capital-labor ratio.* Overaccumulation occurs when aggregate saving maintains a capital-labor ratio above the golden-rule value forever. Let us consider these concepts in detail.

In the present section generally the period length is arbitrary except when we relate to the Diamond model and the period length therefore is half of adult lifetime.

The golden-rule capital-labor ratio

The golden rule is a concept that in itself relates to technically feasible paths only. It does not involve the market form.

Consider the economy-wide resource constraint $C_t = Y_t - S_t = F(K_t, L_t) - (K_{t+1} - K_t + \delta K_t)$, where we assume that F is neoclassical with CRS. Accordingly, aggregate consumption per unit of labor can be written

$$c_t \equiv \frac{C_t}{L_t} = \frac{F(K_t, L_t) - (K_{t+1} - K_t + \delta K_t)}{L_t} = f(k_t) + (1 - \delta)k_t - (1 + n)k_{t+1}, \tag{3.41}$$

where k is the capital-labor ratio K/L. Note that C_t will generally be greater than the workers' consumption. One should simply think of C_t as the flow of produced consumption goods in the economy and c_t as this flow divided by aggregate employment, including the labor that in period t produces investment goods. How

[22] Recall that a *Pareto optimal* path is a technically feasible path with the property that no other technically feasible path will make at least one individual better off without making someone else worse off. A technically feasible path which is not Pareto optimal is called *Pareto inferior.*

the consumption goods are distributed to different members of society is not our concern here.

DEFINITION 6 By the *golden-rule capital-labor ratio*, k_{GR}, is meant that value of the capital-labor ratio k, which results in the highest possible sustainable level of consumption per unit of labor.

Sustainability requires replicability forever. We therefore consider a steady state. In a steady state $k_{t+1} = k_t = k$ so that (3.41) simplifies to

$$c = f(k) - (\delta + n)k \equiv c(k). \tag{3.42}$$

Maximizing gives the first-order condition

$$c'(k) = f'(k) - (\delta + n) = 0. \tag{3.43}$$

In view of $c''(k) = f''(k) < 0$, the condition (3.43) is both necessary and sufficient for an interior maximum. Let us assume that $\delta + n > 0$ and that f satisfies the condition

$$\lim_{k\to\infty} f'(k) < \delta + n < \lim_{k\to 0} f'(k).$$

Then (3.43) has a solution in k, and it is unique because $c''(k) < 0$. The solution is called k_{GR} so that

$$f'(k_{GR}) - \delta = n.$$

That is:

PROPOSITION 5 *(the golden rule)* The highest sustainable consumption level per unit of labor in society is obtained when in steady state the net marginal productivity of capital equals the growth rate of the economy.

It follows that if a society aims at the highest sustainable level of consumption and initially has $k_0 < k_{GR}$, society should increase its capital-labor ratio up to the point where the extra output obtainable by a further small increase is exactly offset by the extra gross investment needed to maintain the capital-labor ratio at that level. The intuition is visible from (3.42). The golden-rule capital-labor ratio, k_{GR}, strikes the right balance in the trade-off between high output per unit of labor and a not too high investment requirement. Although a steady state with $k > k_{GR}$ would imply higher output per unit of labor, it would also imply that a large part of that output is set aside for investment (namely the amount $(\delta + n)k$ per unit of labor) to counterbalance capital depreciation and growth in the labor force; without this investment the high capital-labor ratio k^* would not be maintained. With $k > k_{GR}$ this feature would dominate the first effect so that consumption per unit of labor ends up low. Fig. 3.8 illustrates.

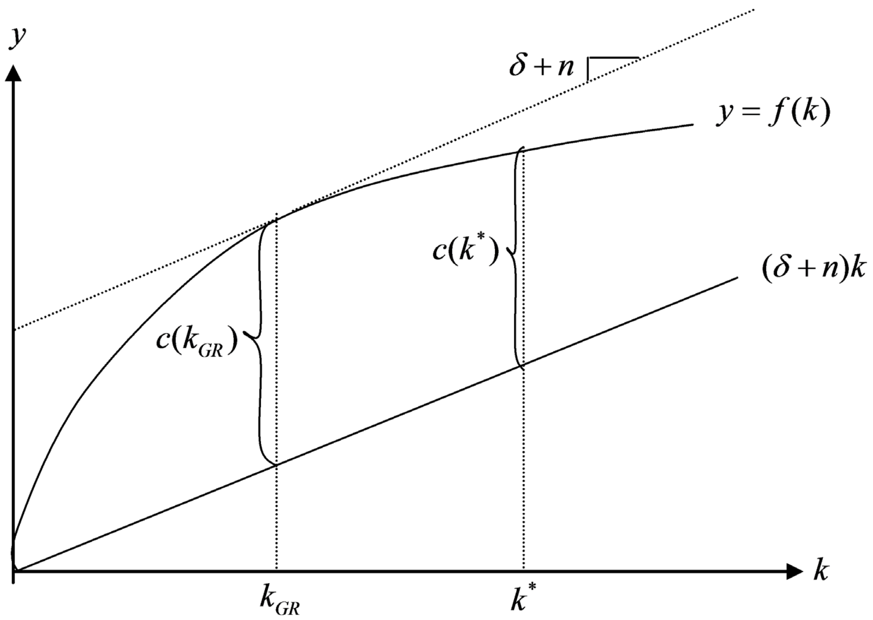


Figure 3.8: A steady state with overaccumulation.

The name golden rule hints at the golden rule from the Bible: "Do unto others as you would have them to do unto you." We imagine that God asks the newly born generation: "What capital-labor ratio would you prefer to be presented with, given that you must hand over the same capital-labor ratio to the next generation?" The appropriate answer is: the golden-rule capital-labor ratio.

The possibility of overaccumulation in a competitive market economy

The equilibrium path in the Diamond model with perfect competition implies an interest rate $r^* = f'(k^*) - \delta$ in a steady state. As an implication,

$$r^* \gtreqless n \Leftrightarrow f'(k^*) - \delta \gtreqless n \Leftrightarrow k^* \lesseqgtr k_{GR}, \text{ respectively,}$$

in view of $f'' < 0$. Hence, a long-run interest rate below the growth rate of the economy indicates that $k^* > k_{GR}$. This amounts to a Pareto-inferior state of affairs. Indeed, everyone can be made better off if by a coordinated reduction of saving and investment, k is reduced. A formal demonstration of this is given in connection with Proposition 6 in the next subsection. Here we give an account in more intuitive terms.

Consider Fig. 3.8. Let k be gradually reduced to the level k_{GR} by refraining from investment in period t_0 and forward until this level is reached. When this happens, let k be maintained at the level k_{GR} forever by providing for the needed investment per young, $(\delta + n)k_{GR}$. Then there would be higher aggregate consumption in period t_0 and every future period. Both the immediate reduction of saving and a resulting lower capital-labor ratio to be maintained contribute to this result. There is thus scope for both young and old to consume more in every future period.

In the Diamond model a simple policy implementing such a Pareto improvement in the case where $k^* > k_{GR}$ (i.e., $r^* < n$) is to incur a lump-sum tax on the young, the revenue of which is immediately transferred lump sum to the old, hence, fully consumed. Suppose this amounts to a transfer of one good from each young to the old. Since there are $1 + n$ young people for each old person, every old receives in this way $1 + n$ goods in the same period. Let this transfer be repeated every future period. By decreasing their saving by one unit, the young can maintain unchanged consumption in their youth, and when becoming old, they receive $1 + n$ goods from the next period's young and so on. In effect, the "return" on the tax payment by the young is $1 + n$ next period. This is more than the $1 + r^*$ that could be obtained via the market through own saving.[23]

A proof that $k^* > k_{GR}$ is indeed theoretically possible in the Diamond model can be based on the log utility-Cobb-Douglas case from Example 2 in Section 3.5.3. As indicated by the formula for r^* in that example, the outcome $r^* < n$, which is equivalent to $k^* > k_{GR}$, can always be obtained by making the parameter $\alpha \in (0, 1)$ in the Cobb-Douglas function small enough. The intuition is that a small α implies a large $1 - \alpha$ and so a large wage income, $wL = (1 - \alpha)K^\alpha L^{-\alpha} \cdot L$ $= (1 - \alpha)Y$. This leads to high saving by the young, since $s_w > 0$. The result is a high k_{t+1} which generates a high real wage also next period and may in this manner be sustained forever.

An intuitive understanding of the fact that the perfectly competitive market mechanism may thus lead to overaccumulation, can be based on the following argument. Assume, first, that $s_r < 0$. In this case, if the young in period t expects the rate of return on their saving to end up small (less than n), the decided saving will be large in order to provide for consumption after retirement. But the aggregate result of this behavior is a high k_{t+1} and therefore a low $f'(k_{t+1})$. In this way the expectation of a low r_{t+1} is confirmed by the actual events. The young persons each do the best they can as atomistic individuals, taking the market conditions as given. Yet the aggregate outcome is an equilibrium with overaccumulation, hence a Pareto-inferior outcome.

[23] In this model with no utility of leisure, a tax on wage income, or a mandatory pay-as-you-go pension contribution (see Chapter 5), would act like a lump-sum tax on the young.

The described tax-transfers policy will affect the equilibrium interest rate negatively. By choosing an appropriate size of the tax this policy, combined with competitive markets, will under certain conditions (see Chapter 5.1) bring the economy to the golden-rule steady state where overaccumulation has ceased and $r^* = n$.

Looking at the issue more closely, we see that $s_r < 0$ is not crucial for this outcome. Suppose $s_r = 0$ (the log utility case) and that in the current period, k_t is, for some historical reason, at least temporarily considerably above k_{GR}. Thus, current wages are high, hence, s_t is relatively high (there is in this case no offsetting effect on s_t from the relatively low expected r_{t+1}). Again, the aggregate result is a high k_{t+1} and thus the expectation is confirmed. Consequently, the situation in the next period is the same and so on. By continuity, even if $s_r > 0$, the argument goes through as long as s_r is not too large.

Dynamic inefficiency and the double infinity

The overaccumulation phenomenon is an example of *dynamic inefficiency.*

DEFINITION 7 A technically feasible path $\{(c_t, k_t)\}_{t=0}^{\infty}$ with the property that there does not exist another technically feasible path with higher c_t in some periods without smaller c_t in other periods is called *dynamically efficient.* A technically feasible path $\{(c_t, k_t)\}_{t=0}^{\infty}$ which is not dynamically efficient is called *dynamically inefficient.*

PROPOSITION 6 A technically feasible path $\{(c_t, k_t)\}_{t=0}^{\infty}$ with the property that for $t \to \infty$, $k_t \to k^* > k_{GR}$, is dynamically inefficient.

Proof. Let $k^* > k_{GR}$. Then there exists an $\varepsilon > 0$ such that $k \in (k^* - 2\varepsilon, k^* + 2\varepsilon)$ implies $f'(k) - \delta < n$ since $f'' < 0$. By concavity of f,

$$f(k) - f(k - \varepsilon) \leq f'(k - \varepsilon)\varepsilon. \tag{3.44}$$

Consider a technically feasible path $\{(c_t, k_t)\}_{t=0}^{\infty}$ with $k_t \to k^*$ for $t \to \infty$ (the reference path). Then there exists a t_0 such that for $t \geq t_0$, $k_t \in (k^* - \varepsilon, k^* + \varepsilon)$, $f'(k_t) - \delta < n$ and $f'(k_t - \varepsilon) - \delta < n$. Consider an alternative feasible path $\left\{(\hat{c}_t, \hat{k}_t)\right\}_{t=0}^{\infty}$, where a) for $t = t_0$ consumption is increased relative to the reference path such that $\hat{k}_{t_0+1} = k_{t_0} - \varepsilon$; and b) for all $t > t_0$, consumption is such that $\hat{k}_{t+1} = k_t - \varepsilon$. We now show that after period t_0, $\hat{c}_t > c_t$. Indeed, for all $t > t_0$, by (3.41),

$$\begin{aligned}
\hat{c}_t &= f(\hat{k}_t) + (1-\delta)\hat{k}_t - (1+n)\hat{k}_{t+1} \\
&= f(k_t - \varepsilon) + (1-\delta)(k_t - \varepsilon) - (1+n)(k_{t+1} - \varepsilon) \\
&\geq f(k_t) - f'(k_t - \varepsilon)\varepsilon + (1-\delta)(k_t - \varepsilon) - (1+n)(k_{t+1} - \varepsilon) \quad \text{(by (3.44))} \\
&> f(k_t) - (\delta + n)\varepsilon + (1-\delta)k_t - (1+n)k_{t+1} + (\delta + n)\varepsilon \\
&= f(k_t) + (1-\delta)k_t - (1+n)k_{t+1} = c_t,
\end{aligned}$$

by (3.41). $\square$

Moreover, it can be shown[24] that:

PROPOSITION 7 A technically feasible path $\{(c_t, k_t)\}_{t=0}^{\infty}$ such that for $t \to \infty$, $k_t \to k^* \leq k_{GR}$, is dynamically efficient.

Accordingly, a steady state with $k^* < k_{GR}$ is never dynamically inefficient. This is because increasing k from this level always has its price in terms of a decrease in *current* consumption; and at the same time decreasing k from this level always has its price in terms of lost *future* consumption. But a steady state with $k^* > k_{GR}$ is always dynamically inefficient. Intuitively, staying forever with $k = k^* > k_{GR}$, implies that society *never* enjoys its great capacity for producing consumption goods.

The fact that $k^* > k_{GR}$ – and therefore dynamic inefficiency – cannot be ruled out might seem to contradict the First Welfare Theorem from the microeconomic theory of general equilibrium. This is the theorem saying that under certain conditions, market equilibria are Pareto optimal. Included in these conditions are that increasing returns to scale are absent, there are no missing markets, markets are competitive, no goods are of public good character, and there are no externalities. An additional (but not always underlined) condition for the First Welfare Theorem is that there is a *finite number of periods or*, if the number of periods is infinite, a *finite number of agents.* In contrast, in the OLG model there is a *double infinity*: an infinite number of periods *and* agents. Hence, the First Welfare Theorem breaks down. Indeed, the case $r^* < n$, i.e., $k^* > k_{GR}$, can arise under *laissez-faire.* Then, as we have seen, everyone can be made better off by a coordinated intervention by some social arrangement (a government for instance) such that k is reduced.

The essence of the matter is that the double infinity opens up for technically feasible reallocations which are definitely beneficial when $r^* < n$ and which a central authority can accomplish but the market can not. That *nobody* need loose by the described kind of redistribution is due to the double infinity: the economy goes on forever and there is no last generation. Nonetheless, some kind of centralized *coordination* is required to accomplish a solution.

There is an analogy in "Gamow's bed problem": There are an infinite number of inns along the road, each with one bed. On a certain rainy night all innkeepers have committed their beds. A late guest comes to the first inn and asks for a bed. "Sorry, full up!" But the minister of welfare hears about it and suggests that from each inn one incumbent guest moves down the road one inn.[25]

Whether the theoretical possibility of overaccumulation should be a matter of practical concern is an empirical question about the relative size of rates of return

[24] Cass (1972).

[25] George Gamow (1904-1968) was a Russian physicist. The problem is also known as *Hilbert's hotel problem*, after the German mathematician David Hilbert (1862-1943).

and economic growth. To answer the question meaningfully, we need an extension of the criterion for overaccumulation so that the presence of technological progress and rising per capita consumption in the long run can be taken into account. This is one of the topics of the next chapter. At any rate, we can already here reveal that there exists no indication that overaccumulation has historically been an actual problem in industrialized market economies.

A final remark before concluding. Proposition 5 about the golden rule can be generalized to the case where instead of one there are n different capital goods in the economy. Essentially the generalization says that assuming CRS-neoclassical production functions with n different capital goods as inputs, one consumption good, no technological change, and perfectly competitive markets, a steady state in which per-unit-of labor consumption is maximized has interest rate equal to the growth rate of the labor force when technological progress is ignored (see, e.g., Mas-Colell, 1989).

3.7 Concluding remarks

There exist several basic long-run models in macroeconomics. Diamonds OLG model is one of them. Its strengths include:

- The *life-cycle* aspect of human behavior is taken into account. Although the economy is infinitely-lived, the individual agents are not. During lifetime one's working capacity, income, and needs change and this is reflected in the individual labor supply and saving behavior. The aggregate implications of the life-cycle behavior of coexisting individual agents at different stages in their life is at the centre of attention.

- The model takes elementary forms of *heterogeneity* in the population into account − there are "old" and "young", there are the currently-alive people and the future generations whose preferences are not reflected in current market transactions. Questions relating to the distribution of income and wealth across generations can be studied. The possibility of coordination failure on a grand scale is laid bare.

Regarding analytical tractability, comparing with the basic representative agent model (the Ramsey model of Chapter 10), the complexity in the OLG model, implied by having in every period two different coexisting generations, is in some respects more than compensated by the fact that the finite time horizon of the households make the *dynamics* of the model *one-dimensional*: we end up with a first-order difference equation in the capital-labor ratio, k_t. In contrast, the dynamics of the basic representative agent model is two-dimensional (owing to the assumed infinite horizon of the households).

3.8 Literature notes

1. The Nobel Laureate Paul A. Samuelson (1915-2009) is one of the pioneers of OLG models. Building on the French economist and Nobel laureate Maurice Allais (1911-2010), a famous article by Samuelson, from 1958, is concerned with a missing market problem. Imagine a two-period OLG economy where, as in the Diamond model, only the young have an income (by Samuelson simplifying considered an exogenous endowment of consumption goods from heaven). Contrary to the Diamond model, however, there is no capital. Also other potential stores of value are absent. Then, in the laissez-faire market economy the old have to starve because they can no longer work and had no possibility of saving - transferring income - as young.

The allocation of resources in the economy is Pareto-inferior. Indeed, if each member of the young generation hands over to the old generation one unit of consumption, and this is next period repeated by the new young generation and so on in the future, everyone will be better off. Since for every old there are $1+n$ young, the implied rate of return would be n, the population growth rate. Such transfers do not arise under laissez-faire. A kind of social contract is required. As Samuelson pointed out, a government could in period 0 issue paper money and transfer these money notes to the members of the old generation who would then use them to buy goods from the young. Provided the young believed the notes to be valuable in the next period, they would accept them in exchange for some of their goods in order to use them in the next period for buying from the new young generation etc.

We have here an example of how a social institution can solve a coordination problem.[26]

2. Diamond (1965) extended Samuelson's contribution by adding capital accumulation. Because of its antecedents Diamonds OLG model is sometimes called the Samuelson-Diamond model or the Allais-Samuelson-Diamond model. In our exposition we have drawn upon clarifications by Galor and Ryder (1989) and de la Croix and Michel (2002). The last mentioned contribution is an extensive exploration of discrete-time OLG models and their applications. An advanced and thorough treatment from a microeconomic general equilibrium perspective is contained in Bewley (2007).

3. The *life-cycle saving hypothesis* was put forward by Franco Modigliani (1918-2003) and associates in the 1950s. See for example Modigliani and Brumberg (1954). Numerous extensions of the framework, relating to the motives (b) - (e) in the list of Section 3.1, see for instance de la Croix and Michel (2002).

[26] To just give a flavor of Samuelson's contribution we have here ignored several aspects, including that Samuelson assumed three periods of life.

4. A review of the empirics of life-cycle behavior and attempts at refining life-cycle models are given in Browning and Crossley (2001).

5. Regarding the dynamic efficiency issue, both the propositions 6 and 7 were shown in a stronger form by the American economist David Cass (1937-2008). Cass established the *general* necessary and sufficient condition for a feasible path $\{(c_t, k_t)\}_{t=0}^{\infty}$ to be dynamically efficient (Cass 1972). Our propositions 6 and 7 are more restrictive in that they are limited to paths that converge. Partly intuitive expositions of the deeper aspects of the theory are given by Shell (1971) and Burmeister (1980).

6. Diamond has also contributed to other fields of economics, including search theory for labor markets. In 2010 Diamond, together with Dale Mortensen and Christopher Pissarides, was awarded the Nobel price in economics.

7. The fact that multiple self-fulfilling equilibrium paths are in several contexts theoretically possible has attracted considerable attention in certain business cycle theories. "Optimism" may result in economic booms and "pessimism" in economic busts, see, e.g., Farmer (2010), which gives an introduction for general readers. We shall have more to say about this literature in Part VI.

From here very incomplete:

The two-period structure of Diamonds OLG model leaves little room for considering, e.g., education and dissaving in the early years of life. This kind of issues is taken up in three-period extensions of the Diamond model, see de la Croix and Michell (2002).

Dynamic inefficiency, see also Burmeister (1980).

Bewley 1977, 1980.

Two-sector OLG: Galor (1992). Galor's book on difference equations.

On the golden rule in a general setup, see Mas-Colell (1989).

3.9 Appendix

A. On CRRA utility

Derivation of the CRRA function Consider a utility function $u(c)$, defined for all $c > 0$ and satisfying $u'(c) > 0$, $u''(c) < 0$. Let the absolute value of the elasticity of marginal utility be denoted $\theta(c)$, that is, $\theta(c) \equiv -cu''(c)/u'(c) > 0$. We claim that if $\theta(c)$ is a positive constant, θ, then, up to a positive linear transformation, $u(c)$ must be of the form

$$u(c) = \begin{cases} \frac{c^{1-\theta}}{1-\theta}, & \text{when } \theta \neq 1, \\ \ln c, & \text{when } \theta = 1, \end{cases} \tag{*}$$

i.e., of CRRA form.

Proof. Suppose $\theta(c) = \theta > 0$. Then, $u''(c)/u'(c) = -\theta/c$. By integration, $\ln u'(c) = -\theta \ln c + A$, where A is an arbitrary constant. Take the antilogarithm function on both sides to get $u'(c) = e^A e^{-\theta \ln c} = e^A c^{-\theta}$. By integration we get

$$u(c) = \begin{cases} e^A \frac{c^{1-\theta}}{1-\theta} + B, & \text{when } \theta \neq 1, \\ e^A \ln c + B, & \text{when } \theta = 1, \end{cases}$$

where B is an arbitrary constant. This proves the claim. Letting $A = B = 0$, we get (*). □

When we want to make the kinship between the members of the "CRRA family" transparent, we maintain $A = 0$ and for $\theta = 1$ also $B = 0$, whereas for $\theta \neq 1$ we set $B = -1/(1-\theta)$. In this way we achieve that all members of the CRRA family will be represented by curves going through the same point as the log function, namely the point $(1, 0)$, cf. Fig. 3.2. For a particular $\theta > 0, \theta \neq 1$, we have $u(c) = (c^{1-\theta} - 1)/(1-\theta)$, which makes up the *CRRA utility function in normalized form*. Given θ, the transformation to normalized form is of no consequence for the economic behavior since adding or subtracting a constant does not affect marginal rates of substitution.

The domain of the CRRA function From an economic point of view it is desirable that the domain of our utility functions include $c = 0$. Starvation is a real-life possibility. Right away, if $\theta \geq 1$, the CRRA function, whether in the form $u(c) = (c^{1-\theta} - 1)/(1-\theta)$ or in the form (*), is defined only for $c > 0$. This is because for $c \to 0$ we get $u(c) \to -\infty$. In this case we simply define $u(0) = -\infty$. This is a natural extension since the CRRA function anyway has the property that $u'(c) \to \infty$, when $c \to 0$ (whether θ is larger or smaller than one). The marginal utility thus becomes very large as c becomes very small, that is, the No Fast Assumption is satisfied. This will ensure that the chosen c is strictly positive whenever there is a positive budget. And the case of a zero budget, leading to zero consumption, is not ruled out. So throughout this book we define the domain of the CRRA function to be $[0, \infty)$.

The range of the CRRA function Considering the CRRA function $u(c) \equiv \left(c^{1-\theta} - 1\right)(1-\theta)^{-1}$ for $c \in [0, \infty)$, we have:

$$\begin{aligned} \text{for } 0 &< \theta < 1, \text{ the range of } u(c) \text{ is } \left[-(1-\theta)^{-1}, \infty\right), \\ \text{for } \theta &= 1, \text{ the range of } u(c) \text{ is } [-\infty, \infty), \\ \text{for } \theta &> 1, \text{ the range of } u(c) \text{ is } [-\infty, -(1-\theta)^{-1}). \end{aligned}$$

Thus, in the latter case $u(c)$ is bounded from above and so allows asymptotic "satiation" to occur.

B. Deriving the elasticity of intertemporal substitution in consumption

Referring to Section 3.3, we here show that the definition of $\sigma(c_1, c_2)$ in (3.17) gives the result (3.18). Let $x \equiv c_2/c_1$ and $\beta \equiv (1+\rho)^{-1}$. Then the first-order condition (3.16) and the equation describing the considered indifference curve constitute a system of two equations

$$\begin{aligned} u'(c_1) &= \beta u'(xc_1)R, \\ u(c_1) + \beta u(xc_1) &= \bar{U}. \end{aligned}$$

For a fixed utility level $U = \bar{U}$ these equations define c_1 and x as implicit functions of R, $c_1 = c(R)$ and $x = x(R)$. We calculate the total derivative with respect to R in both equations and get, after ordering,

$$\begin{aligned} & [u''(c_1) - \beta R u''(xc_1)x]\, c'(R) - \beta R u''(xc_1) c_1 x'(R) \\ &= \beta u'(xc_1), \qquad (3.45) \\ [u'(c_1) + \beta u'(xc_1)x]\, c'(R) &= -\beta u'(xc_1) c_1 x'(R). \qquad (3.46) \end{aligned}$$

Substituting $c'(R)$ from (3.46) into (3.45) and ordering now yields

$$-\left[x\frac{c_1 u''(c_1)}{u'(c_1)} + R\frac{xc_1 u''(xc_1)}{u'(xc_1)}\right]\frac{R}{x}x'(R) = x + R.$$

Since $-cu''(c)/u'(c) \equiv \theta(c)$, this can be written

$$\frac{R}{x}x'(R) = \frac{x+R}{x\theta(c_1) + R\theta(xc_1)}.$$

Finally, in view of $xc_1 = c_2$ and the definition of $\sigma(c_1, c_2)$, this gives (3.18).

C. Walras' law

In the proof of Proposition 1 we referred to Walras' law. Here is how Walras' law works in each period in a model like this. We consider period t, but for simplicity we skip the time index t on the variables. There are three markets, a market for capital services, a market for labor services, and a market for output goods. Suppose a "Walrasian auctioneer" calls out the price vector $(\hat{r}, w, 1)$, where $\hat{r} > 0$ and $w > 0$, and asks all agents, i.e., the young, the old, and the representative firm, to declare their supplies and demands.

The supplies of capital and labor are by assumption inelastic and equal to K units of capital services and L units of labor services. But the demand for capital and labor services depends on the announced $\hat{r}$ and w. Let the potential pure profit of the representative firm be denoted Π. If $\hat{r}$ and w are so that $\Pi < 0$, the firm declares $K^d = 0$ and $L^d = 0$. If on the other hand at the announced $\hat{r}$ and w, $\Pi = 0$ (as when $\hat{r} = r(k) + \delta$ and $w = w(k)$), the desired capital-labor ratio is given as $k^d = f'^{-1}(\hat{r})$ from (3.20), but the firm is indifferent with respect to the absolute level of the factor inputs. In this situation the auctioneer tells the firm to declare $L^d = L$ (recall L is the given labor supply) and $K^d = k^d L^d$ which is certainly acceptable for the firm. Finally, if $\Pi > 0$, the firm is tempted to declare infinite factor demands, but to avoid that, the auctioneer imposes the rule that the maximum allowed demands for capital and labor are $2K$ and $2L$, respectively. Within these constraints the factor demands will be uniquely determined by $\hat{r}$ and w and we have

$$\Pi = \Pi(\hat{r}, w, 1) = F(K^d, L^d) - \hat{r}K^d - wL^d. \tag{3.47}$$

The owners of both the capital stock K and the representative firm must be those who saved in the previous period, namely the currently old. These elderly will together declare the consumption $c_2 L_{-1} = (1 + \hat{r} - \delta)K + \Pi$ and the net investment $-K$ (which amounts to disinvestment). The young will declare the consumption $c_1 L = wL - s(w, r_{+1}^e)L$ and the net investment $sL = s(w, r_{+1}^e)L$. So aggregate declared consumption will be $C = (1 + \hat{r} - \delta)K + \Pi + wL - s(w, r_{+1}^e)L$ and aggregate net investment $I - \delta K = s(w, r_{+1}^e)L - K$. It follows that $C + I = wL + \hat{r}K + \Pi$. The aggregate declared supply of output is $Y^s = F(K^d, L^d)$. The values of excess demands in the three markets now add to

$$\begin{aligned} Z(\hat{r}, w, 1) &\equiv w(L^d - L) + \hat{r}(K^d - K) + C + I - Y^s \\ &= wL^d - wL + \hat{r}K^d - \hat{r}K + wL + \hat{r}K + \Pi - F(K^d, L^d) \\ &= wL^d + \hat{r}K^d + \Pi - F(K^d, L^d) = 0, \end{aligned}$$

by (3.47).

This is a manifestation of Walras' law for each period: *whatever the announced price vector for the period is, the aggregate value of excess demands in the period is zero.* The reason is the following. When each household satisfies its budget constraint and each firm pays out its ex ante profit,[27] then the economy as a whole has to satisfy an aggregate budget constraint for the period considered.

The budget constraints, demands, and supplies operating in this thought experiment (and in Walras' law in general) are the *Walrasian* budget constraints, demands, and supplies. Outside equilibrium these are somewhat artificial constructs. A Walrasian budget constraint is based on the assumption that the desired actions can be realized. This assumption will be wrong unless $\hat{r}$ and w are already at their equilibrium levels. But the assumption that desired actions

[27] By ex ante profit is meant the hypothetical profit calculated on the basis of firms' desired supply evaluated at the announced price vector, $(\hat{r}, w, 1)$.

can be realized is never falsified because the thought experiment does not allow trades to take place outside Walrasian equilibrium. Similarly, the Walrasian consumption demand by the worker is rather hypothetical outside equilibrium. This demand is based on the income the worker *would* get if fully employed at the announced real wage, not on the actual employment (or unemployment) at that real wage.

These ambiguities notwithstanding, the important message of Walras' law goes through, namely that when two of the three markets clear (in the sense of the Walrasian excess demands being nil), so does the third.

D. Proof of (i) and (ii) of Proposition 2

For convenience we repeat the fundamental difference equation characterizing an equilibrium path:

$$k_{t+1} = \frac{s\left(w\left(k_t\right), r\left(k_{t+1}\right)\right)}{1+n},$$

where $w(k) \equiv f(k) - f'(k)k > 0$ for all $k > 0$ and $r(k) \equiv f'(k) - \delta > -1$ for all $k \geq 0$. The key to the proof of Proposition 2 about existence of an equilibrium path is the following lemma.

LEMMA D1 Suppose the No Fast Assumption (A1) applies and let $w > 0$ and $n > -1$ be given. Then the equation

$$\frac{s\left(w, r\left(k\right)\right)}{k} = 1 + n. \tag{3.48}$$

has at least one solution $k > 0$.

Proof. Note that $1 + n > 0$. From Lemma 1 in Section 3.3 follows that for all possible values of $r(k)$, $0 < s(w, r(k)) < w$. Hence, for any $k > 0$,

$$0 < \frac{s\left(w, r\left(k\right)\right)}{k} < \frac{w}{k}.$$

Letting $k \to \infty$ we then have $s\left(w, r\left(k\right)\right)/k \to 0$ since $s\left(w, r\left(k\right)\right)/k$ is squeezed between 0 and 0 (as indicated in the two graphs in Fig. 3.9).

Next we consider $k \to 0$. There are two cases.

Case 1: $\lim_{k\to 0} s\left(w, r\left(k\right)\right) > 0$.[28] Then obviously $\lim_{k\to 0} s\left(w, r\left(k\right)\right)/k = \infty$.

[28] If the limit does not exist, the proof applies to the *limit inferior* of $s\left(w, r\left(k\right)\right)$ for $k \to 0$. The limit inferior for $i \to \infty$ of a sequence $\{x_i\}_{i=0}^{\infty}$ is defined as $\lim_{i\to\infty} \inf\{x_j | \, j = i, i+1, \dots\}$, where inf of a set $S_i = \{x_j | \, j = i, i+1, \dots\}$ is defined as the greatest lower bound for S_i.

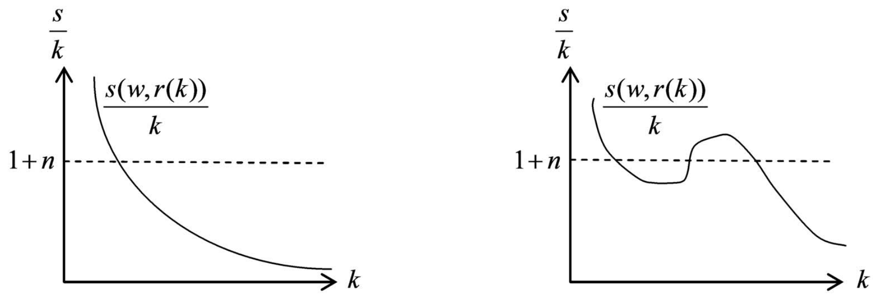


Figure 3.9: Existence of a solution to equation (3.48).

Case 2: $\lim_{k\to 0} s(w, r(k)) = 0$.[29] In this case we have

$$\lim_{k\to 0} r(k) = \infty. \tag{3.49}$$

Indeed, since $f'(k)$ rises monotonically as $k \to 0$, the only alternative would be that $\lim_{k\to 0} r(k)$ exists and is $< \infty$; then, by Lemma 1 in Section 3.3, we would be in case 1 rather than case 2. By the second-period budget constraint, with $r = r(k)$, consumption as old is $c_2 = s(w, r(k))(1 + r(k)) \equiv c(w, k) > 0$ so that

$$\frac{s(w, r(k))}{k} = \frac{c(w, k)}{[1 + r(k)]\, k}.$$

The right-hand side of this equation goes to ∞ for $k \to 0$ since $\lim_{k\to 0} [1 + r(k)]\, k = 0$ by Technical Remark in Section 3.4 and $\lim_{k\to 0} c(w, k) = \infty$; this latter fact follows from the first-order condition (3.8), which can be written

$$0 \leq u'(c(w, k)) = (1 + \rho)\frac{u'(w - s(w, r(k))}{1 + r(k)} \leq (1 + \rho)\frac{u'(w)}{1 + r(k)}.$$

Taking limits on both sides gives

$$\lim_{k\to 0} u'(c(w, k)) = (1 + \rho) \lim_{k\to 0} \frac{u'(w - s(w, r(k)))}{1 + r(k)} = (1 + \rho) \lim_{k\to 0} \frac{u'(w)}{1 + r(k)} = 0,$$

where the second equality comes from the fact that we are in case 2 and the third comes from (3.49). But since $u'(c) > 0$ and $u''(c) < 0$ for all $c > 0$, $\lim_{k\to 0} u'(c(w, k)) = 0$ requires $\lim_{k\to 0} c(w, k) = \infty$, as was to be shown.

In both Case 1 and Case 2 we thus have that $k \to 0$ implies $s(w, r(k))/k \to \infty$. Since $s(w, r(k))/k$ is a continuous function of k, there must be at least one $k > 0$ such that (3.48) holds (as illustrated by the two graphs in Fig. 3.14). □

[29] If the limit does not exist, the proof applies to the *limit inferior* of $s(w, r(k))$ for $k \to 0$.

Now, to prove (i) of Proposition 2, consider an arbitrary $k_t > 0$. We have $w(k_t) > 0$. In (3.48), let $w = w(k_t)$. By Lemma C1, (3.48) has a solution $k > 0$. Set $k_{t+1} = k$. Starting with $t = 0$, from a given $k_0 > 0$ we thus find a $k_1 > 0$ and letting $t = 1$, from the now given k_1 we find a k_2 and so on. The resulting infinite sequence $\{k_t\}_{t=0}^{\infty}$ is an equilibrium path. In this way we have proved existence of an equilibrium path if $k_0 > 0$. Thereby (i) of Proposition 2 is proved.

But what if $k_0 = 0$? Then, if $f(0) = 0$, no temporary equilibrium is possible in period 0, in view of (ii) of Proposition 1; hence there can be no equilibrium path. Suppose $f(0) > 0$. Then $w(k_0) = w(0) = f(0) > 0$, as explained in Technical Remark in Section 3.4. Let w in equation (3.48) be equal to $f(0)$. By Lemma C1 this equation has a solution $k > 0$. Set $k_1 = k$. Letting period 1 be the new initial period, we are back in the case with initial capital positive. This proves (ii) of Proposition 2.

E. Sufficient conditions for certain properties of the transition curve

Positive slope everywhere For convenience we repeat here the condition (3.36):

$$\frac{1}{1-\gamma} > \frac{1-\sigma}{1+(1+\rho)^{-\sigma}(1+f'(k)-\delta)^{\sigma-1}}, \tag{*}$$

where we have substituted $\sigma \equiv 1/\theta$. In Section 3.5.3 we claimed that in the CRRA-CES case this condition is sufficient for the transition curve to be positively sloped everywhere. We here prove the claim.

Consider an arbitrary $k_t > 0$ and let $w \equiv w(k_t) > 0$. Knowing that $w'(k_t) > 0$ for all $k_t > 0$, we can regard k_{t+1} as directly linked to w. With k representing k_{t+1}, k must satisfy the equation $k = s(w, r(k))/(1 + n)$. A sufficient condition for this equation to implicitly define k as an increasing function of w is also a sufficient condition for the transition curve to be positively sloped for all $k_t > 0$.

When $u(c)$ belongs to the CRRA class, by (3.15) with $\sigma \equiv 1/\theta$, we have $s(w, r(k)) = \left[1 + (1 + \rho)^{\sigma}(1 + r(k))^{1-\sigma}\right]^{-1} w$. The equation $k = s(w, r(k))/(1+n)$ then implies

$$\frac{w}{1+n} = k\left[1+(1+\rho)^{\sigma}R(k)^{1-\sigma}\right] \equiv h(k), \tag{3.50}$$

where $R(k) \equiv 1 + r(k) \equiv 1 + f'(k) - \delta > 0$ for all $k > 0$. It remains to provide a sufficient condition for obtaining $h'(k) > 0$ for all $k > 0$. We have

$$h'(k) = 1 + (1+\rho)^{\sigma}R(k)^{1-\sigma}\left[1-(1-\sigma)\eta(k)\right], \tag{3.51}$$

since $\eta(k) \equiv -kR'(k)/R(k) > 0$, the sign being due to $R'(k) = f''(k) < 0$. So $h'(k) > 0$ if and only if $1-(1-\sigma)\eta(k) > -(1+\rho)^{-\sigma}R(k)^{\sigma-1}$, a condition equivalent to

$$\frac{1}{\eta(k)} > \frac{1-\sigma}{1+(1+\rho)^{-\sigma}R(k)^{\sigma-1}}. \tag{3.52}$$

To make this condition more concrete, consider the CES production function

$$f(k) = A(\alpha k^\gamma + 1 - \alpha), \qquad A > 0, 0 < \alpha < 1, \gamma < 1. \tag{3.53}$$

Then $f'(k) = \alpha A^\gamma (f(k)/k)^{1-\gamma}$ and defining $\pi(k) \equiv f'(k)k/f(k)$ we find

$$\eta(k) = (1-\gamma)\frac{(1-\pi(k))f'(k)}{1-\delta+f'(k)} \leq (1-\gamma)(1-\pi(k)) < 1-\gamma, \tag{3.54}$$

where the first inequality is due to $0 \leq \delta \leq 1$ and the second to $0 < \pi(k) < 1$, which is an implication of strict concavity of f combined with $f(0) \geq 0$. Thus, $\eta(k)^{-1} > (1-\gamma)^{-1}$ so that if (*) holds for all $k > 0$, then so does (3.52), i.e., $h'(k) > 0$ for all $k > 0$. We have hereby shown that (*) is sufficient for the transition curve to be positively sloped everywhere.

Transition curve steep for k small Here we specialize further and consider the CRRA-Cobb-Douglas case: $u(c) = (c^{1-\theta} - 1)/(1-\theta)$, $\theta > 0$, and $f(k) = Ak^\alpha$, $A > 0$, $0 < \alpha < 1$. In the prelude to Proposition 4 in Section 3.5 it was claimed that if this combined utility and technology condition holds at least for small k, then (ii) of (A3) is satisfied. We now show this.

Letting $\gamma \to 0$ in (3.53) gives the Cobb-Douglas function $f(k) = Ak^\alpha$ (this is proved in the appendix to Chapter 4). With $\gamma = 0$, clearly $(1-\gamma)^{-1} = 1 > 1-\sigma$, where $\sigma \equiv \theta^{-1} > 0$. This inequality implies that (*) above holds and so the transition curve is positively sloped everywhere. As an implication there is a transition function, φ, such that $k_{t+1} = \varphi(k_t)$, $\varphi'(k_t) > 0$. Moreover, since $f(0) = 0$, we have, by Lemma 5, $\lim_{k_t \to 0} \varphi(k_t) = 0$.

Given the imposed CRRA utility, the fundamental difference equation of the model is

$$k_{t+1} = \frac{w(k_t)}{(1+n)\left[1 + (1+\rho)^\sigma R(k_{t+1})^{1-\sigma}\right]} \tag{3.55}$$

or, equivalently,

$$h(k_{t+1}) = \frac{w(k_t)}{1+n},$$

where $h(k_{t+t})$ is defined as in (3.50). By implicit differentiation we find $h'(k_{t+1})\varphi'(k_t) = w'(k_t)/(1+n)$, i.e.,

$$\varphi'(k_t) = \frac{w'(k_t)}{(1+n)h'(k_{t+1})} > 0.$$

If $k^* > 0$ is a steady-state value of k_t, (3.55) implies

$$1 + (1+\rho)^\sigma R(k^*)^{1-\sigma} = \frac{w(k^*)}{(1+n)k^*}, \tag{3.56}$$

and the slope of the transition curve at the steady state will be

$$\varphi'(k^*) = \frac{w'(k^*)}{(1+n)h'(k^*)} > 0. \tag{3.57}$$

If we can show that such a $k^* > 0$ exists, is unique, and implies $\varphi'(k^*) < 1$, then the transition curve crosses the 45° line from above, and so (ii) of (A3) follows in view of $\lim_{k_t \to 0} = 0$.

Defining $x(k) \equiv f(k)/k = Ak^{\alpha-1}$, where $x'(k) = (\alpha-1)Ak^{\alpha-2} < 0$, and using that $f(k) = Ak^\alpha$, we have $R(k) = 1 + \alpha x(k) - \delta$ and $w(k)/k = (1-\alpha)x(k)$. Hence, (3.56) can be written

$$1 + (1+\rho)^\sigma (1 + \alpha x^* - \delta)^{1-\sigma} = \frac{1-\alpha}{1+n} x^*, \tag{3.58}$$

where $x^* = x(k^*)$. It is easy to show graphically that this equation has a unique solution $x^* > 0$ whether $\sigma < 1$, $\sigma = 1$, or $\sigma > 1$. Then $k^* = (x^*/A)^{1/(\alpha-1)} > 0$ is also unique.

By (3.51) and (3.58),

$$\begin{aligned} h'(k^*) &= 1 + (\frac{1-\alpha}{1+n}x^* - 1)\left[1 - (1-\sigma)\eta(k^*)\right] > 1 + (\frac{1-\alpha}{1+n}x^* - 1)(1 - \eta(k^*)) \\ &\geq 1 + (\frac{1-\alpha}{1+n}x^* - 1)\alpha, \end{aligned}$$

where the first inequality is due to $\sigma > 0$ and the second to the fact that $\eta(k) \leq 1 - \alpha$ in view of (3.54) with $\gamma = 0$ and $\pi(k) = \alpha$. Substituting this together with $w'(k^*) = (1-\alpha)\alpha x^*$ into (3.57) gives

$$0 < \varphi'(k^*) < \frac{\alpha x^*}{1 + n + \alpha x^*} < 1, \tag{3.59}$$

as was to be shown.

The CRRA-Cobb-Douglas case is well-behaved For the case of CRRA utility and Cobb-Douglas technology with CRS, existence and uniqueness of a steady state has just been proved. Asymptotic stability follows from (3.59). So the CRRA-Cobb-Douglas case is well-behaved.

3.10 Exercises

3.1 The dynamic accounting relation for a closed economy is

$$K_{t+1} = K_t + S^N \tag{*}$$

where K_t is the aggregate capital stock and S_t^N is aggregate net saving. In the Diamond model, let S_{1t} be aggregate net saving of the young in period t and S_{2t} aggregate net saving of the old in the same period. On the basis of (*) give a direct proof that the link between two successive periods takes the form $k_{t+1} = s_t/(1+n)$, where s_t is the saving of each young, n is the population growth rate, and k_{t+1} is the capital/labor ratio at the beginning of period $t+1$. *Hint:* by definition, the increase in financial wealth is the same as net saving (ignoring gifts).

3.2 Suppose the production function in Diamond's OLG model is $Y = A(\alpha K^\gamma + (1-\alpha)L^\gamma)^{1/\gamma}$, $A > 0, 0 < \alpha < 1, \gamma < 0$, and $A\alpha^{1/\gamma} < 1+n$. a) Given $k \equiv K/L$, find the equilibrium real wage, $w(k)$. b) Show that $w(k) < (1+n)k$ for all $k > 0$. *Hint:* consider the roof. c) Comment on the implication for the long-run evolution of the economy. *Hint:* consider the ceiling.

3.3 *(multiple temporary equilibria with self-fulfilling expectations)* Fig. 3.10 shows the transition curve for a Diamond OLG model with $u(c) = c^{1-\theta}/(1-\theta)$, $\theta = 8$, $\rho = 0.4$, $n = 0.2$, $\delta = 0.6$, $f(k) = A(bk^p + 1 - b)^{1/p}$, $A = 7$, $b = 0.33$, $p = -0.4$.

a) Let $t = 0$. For a given k_0 slightly below 1, how many temporary equilibria with self-fulfilling expectations are there?

b) Suppose the young in period 0 expect the real interest rate on their saving to be relatively low. Describe by words the resulting equilibrium path in this case. Comment (what is the economic intuition behind the path?).

c) In the first sentence under b), replace "low" by "high". How is the answer to b) affected? What kind of difficulty arises?

3.4 *(plotting the transition curve by MATLAB)* This exercise requires computation on a computer. You may use *MATLAB OLG program.*[30]

a) Enter the model specification from Exercise 3.3 and plot the transition curve.

b) Plot examples for two other values of the substitution parameter: $p = -1.0$ and $p = 0.5$. Comment.

c) Find the approximate largest lower bound for p such that higher values of p eliminates multiple equilibria.

[30] Made by Marc P. B. Klemp and available at the address: http://www.econ.ku.dk/okocg/Computation/.

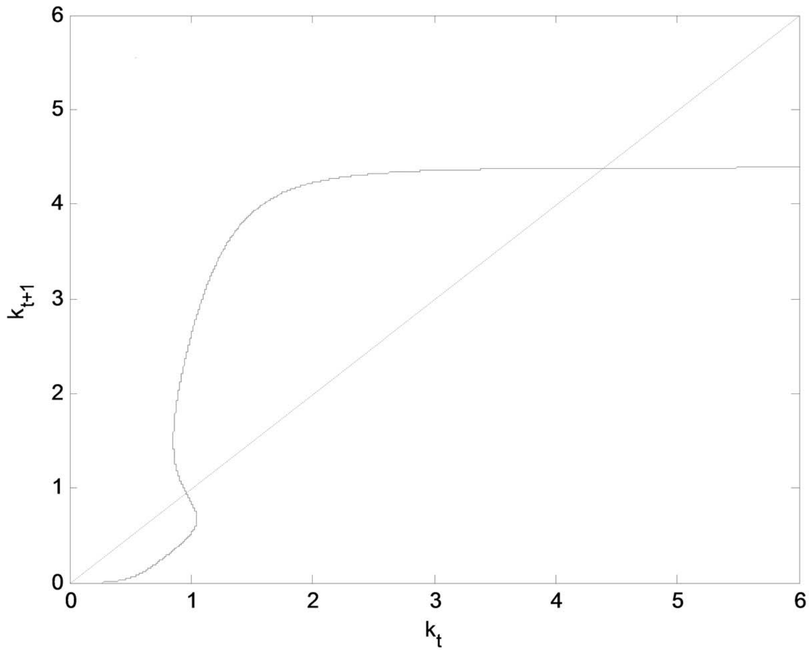


Figure 3.10: Transition curve for Diamond's OLG model in the case described in Exercise 3.3.

d) In continuation of c), what is the corresponding elasticity of factor substitution, ψ? *Hint:* as shown in §4.4, the formula is $\psi = 1/(1-p)$.

e) The empirical evidence for industrialized countries suggests that $0.4 < \psi < 1.0$. Is your ψ from d) empirically realistic? Comment.

3.5 *(one stable and one unstable steady state)* Consider the following Diamond model: $u(c) = \ln c$, $\rho = 2.3$, $n = 2.097$, $\delta = 1.0$, $f(k) = A(bk^p + 1 - b)^{1/p}$, $A = 20$, $b = 0.5$, $p = -1.0$.

a) Plot the transition curve of the model. *Hint:* you may use either a program like *MATLAB OLG Program* (available on the course website) or first a little algebra and then Excel (or similar simple software).

b) Comment on the result you get. Will there exist a poverty trap? Why or why not?

c) At the stable steady state calculate numerically the output-capital ratio, the aggregate saving-income ratio, the real interest rate, and the capital income share of gross national income.

d) Briefly discuss how your results in c) comply with your knowledge of corresponding empirical magnitudes in industrialized Western countries?

e) There is one feature which this model, as a long-run model, ought to incorporate, but does not. Extend the model, taking this feature into account, and write down the fundamental difference equation for the extended model in algebraic form.

f) Plot the new transition curve. *Hint:* given the model specification, this should be straightforward if you use Excel (or similar); and if you use MATLAB OLG Program, note that by a simple "trick" you can transform your new model into the "old" form.

g) The current version of the MATLAB OLG Program is not adapted to this question. So at least here you need another approach, for instance based on a little algebra and then Excel (or similar simple software). Given $k_0 = 10$, calculate numerically the time path of k_t and plot the *time profile* of k_t, i.e., the graph (t, k_t) in the tk-plane. Next, do the same for $k_0 = 1$. Comment.

3.6 *(dynamics under myopic foresight)*
(incomplete) Show the possibility of a chaotic trajectory.

3.7 Given the period utility function is CRRA, derive the saving function of the young in Diamond's OLG model. *Hint:* substitute the period budget constraints into the Euler equation.

3.8 *Short questions* a) A steady-state capital-labor ratio can be in the "dynamically efficient" region or in the "dynamically inefficient" region. How are the two mentioned regions defined? b) Give a simple characterization of the two regions. c) The First Welfare Theorem states that, given certain conditions, any competitive equilibrium ($\equiv$ Walrasian equilibrium) is Pareto optimal. Give a list of circumstances that each tend to obstruct Pareto optimality of a competitive equilibrium.

3.9 Consider a Diamond OLG model for a closed economy. Let the utility discount rate be denoted ρ and let the period utility function be specified as $u(c) = \ln c$.

a) Derive the saving function of the young. Comment.

b) Let the aggregate production function be a neoclassical production function with CRS and ignore technological progress. Let L_t denote the number of young in period t. Derive the fundamental difference equation of the model.

From now, assume that the production function is $Y = \alpha L + \beta KL/(K+L)$, where $\alpha > 0$ and $\beta > 0$ (as in Problem 2.4).

c) Draw a transition diagram illustrating the dynamics of the economy. Make sure that you draw the diagram so as to exhibit consistency with the production function.

d) Given the above information, can we be sure that there exists a unique and globally asymptotically stable steady state? Why or why not?

e) Suppose the economy is in a steady state up to and including period $t_0 > 0$. Then, at the shift from period t_0 to period $t_0 + 1$, a negative technology shock occurs such that the technology level in period $t_0 + 1$ is below that of period t_0. Illustrate by a transition diagram the evolution of the economy from period t_0 onward. Comment.

f) Let $k \equiv K/L$. In the $(t, \ln k)$ plane, draw a graph of $\ln k_t$ such that the qualitative features of the time path of $\ln k$ before and after the shock, including the long run, are exhibited.

g) How, if at all, is the real interest rate in the long run affected by the shock?

h) How, if at all, is the real wage in the long run affected by the shock?

i) How, if at all, is the labor income share of national income in the long run affected by the shock?

j) Explain by words the economic intuition behind your results in h) and i).

3.10

Chapter 4

A growing economy

In the previous chapter we ignored technological progress. An incontestable fact of real life in industrialized countries is, however, the presence of a persistent rise in GDP per capita – on average between 1.5 and 2.5 percent per year since 1870 in many developed economies. In regard to UK, USA, and Japan, see Fig. 4.1; and in regard to Denmark, see Fig. 4.2. In spite of the somewhat dubious quality of the data from before the Second World War, this observation should be taken into account in a model which, like the Diamond model, aims at dealing with long-run issues. For example, in relation to the question of dynamic inefficiency, cf. Chapter 3, the cut-off value of the steady-state interest rate is the steady-state GDP growth rate of the economy and this growth rate increases one-to-one with the rate of technological progress. We shall therefore now introduce technological progress.

On the basis of a summary of "stylized facts" about growth, Section 4.1 motivates the assumption that technological progress at the aggregate level takes the Harrod-neutral form. In Section 4.2 we extend the Diamond OLG model by incorporating this form of technological progress. Section 4.3 extends the concept of the golden rule to allow for the existence of technological progress. In Section 4.4 we address what is known as the marginal productivity theory of the functional income distribution and apply an expedient analytical tool, the elasticity of factor substitution. The next section defines the concept of elasticity of factor substitution at the general level. Section 4.6 then goes into detail with the special case of a constant elasticity of factor substitution (the CES production function). Finally, Section 4.7 concludes with some general considerations regarding the concept of economic growth.

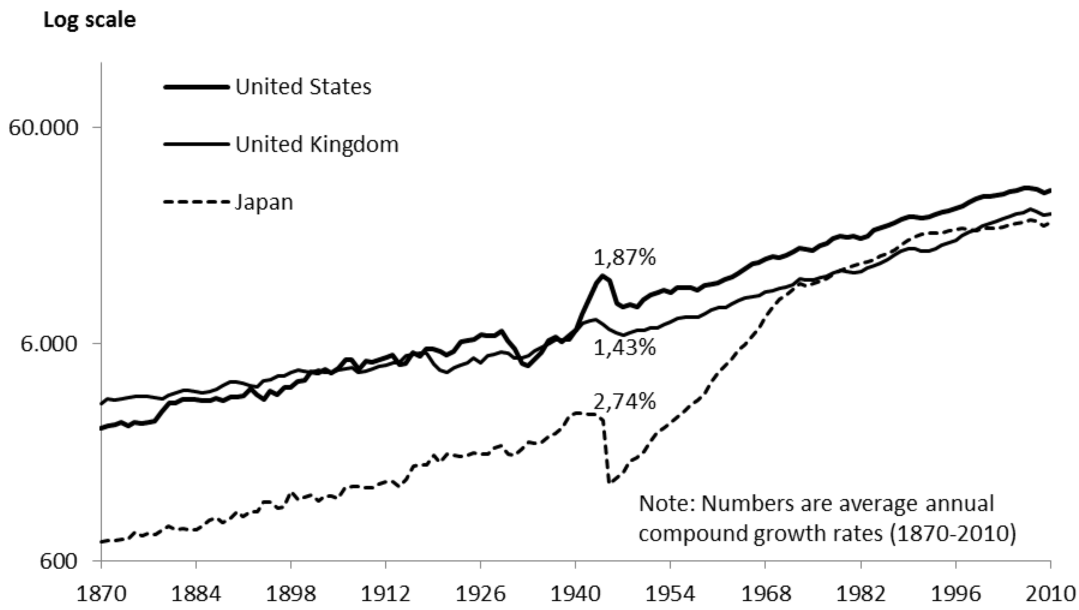


Figure 4.1: GDP per capita in USA, UK, and Japan 1870-2010. Source: Bolt and van Zanden (2013).

4.1 Harrod-neutrality and Kaldor's stylized facts

To allow for technological change, we may write aggregate production this way

$$Y_t = \tilde{F}(K_t, L_t, t), \tag{4.1}$$

where Y_t, K_t, and L_t stand for output, capital input, and labor input, respectively. Changes in technology are here represented by the dependency of the production function $\tilde{F}$ on time, t. For fixed t, the production function may still be for instance neoclassical with respect to the role of the factor inputs, the first two arguments. Often we assume that $\tilde{F}$ depends in a smooth way on time such that the partial derivative, $\partial\tilde{F}_t/\partial t$, exists and is a continuous function of (K_t, L_t, t). When $\partial\tilde{F}_t/\partial t > 0$, technological change amounts to technological *progress:* for K_t and L_t held constant, output increases with t.

A particular form of the time-dependency of the production function has attracted the attention of macroeconomists. This is known as *Harrod-neutral technological progress* and is present when we can rewrite $\tilde{F}$ such that

$$Y_t = F(K_t, T_tL_t), \tag{4.2}$$

where the "level of technology" is represented by a coefficient, T_t, on the labor input, and this coefficient is rising over time. An alternative name for this

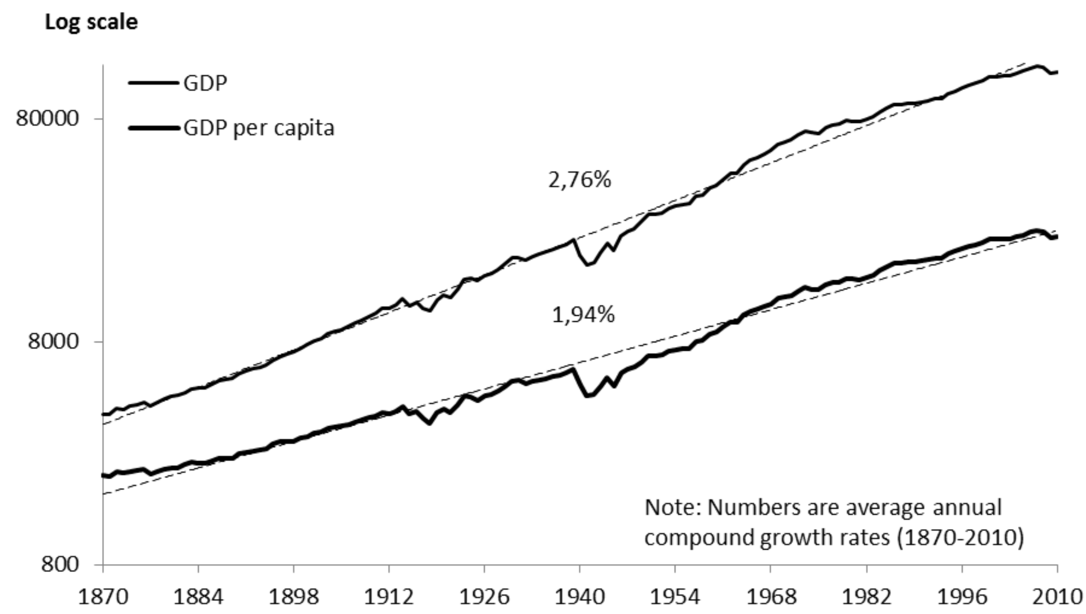


Figure 4.2: GDP and GDP per capita. Denmark 1870-2006. Sources: Bolt and van Zanden (2013); Maddison (2010); The Conference Board Total Economy Database (2013).

in the literature is *labor-augmenting* technological progress. The name "labor-augmenting" may sound as if *more* labor is required to reach a given output level for given capital. In fact, the opposite is the case, namely that T_t has risen so that *less* labor input is required. The idea is that the technological change $-$ a certain percentage increase in T $-$ affects the output level *as if* the labor input had been increased exactly by this percentage, and nothing else had happened.

The interpretation of Harrod neutrality is not that something miraculous happens to the labor input. The content of (4.2) is just that technological innovations are assumed to predominantly be such that not only do labor and capital *in combination* become more productive, but this happens to *manifest itself* in the form (4.2), that is, *as if* an improvement in the *quality* of the labor input had occurred.[1]

[1] As is usual in simple macroeconomic models, in both (4.1) and (4.2) it is simplifying assumed that technological progress is *disembodied.* This means that new technical and organizational knowledge increases the combined productivity of capital and workers independently of when the first were constructed and the latter educated, cf. Chapter 2.2.

Kaldor's stylized facts

The reason that macroeconomists often assume that technological change at the aggregate level takes the Harrod-neutral form, as in (4.2), and not for example the form $Y_t = F(X_t K_t, T_t L_t)$ (where both X and T are changing over time, at least one of them growing), is the following. You want the long-run properties of the model to comply with Kaldor's list of "stylized facts" (Kaldor 1961) concerning the long-run evolution of certain "Great Ratios" of industrialized economies. Abstracting from short-run fluctuations, Kaldor's "stylized facts" are:

1. K/L and Y/L are growing over time and have roughly constant growth rates;

2. the output-capital ratio, Y/K, the income share of labor, wL/Y, and the economy-wide "rate of return", $(Y - wL - \delta K)/K$,[2] are roughly constant over time;

3. the growth rate of Y/L can vary substantially across countries for quite long time.

Ignoring the conceptual difference between the path of Y/L and that of Y *per capita* (a difference not so important in this context), the figures 4.1 and 4.2 illustrate Kaldor's "fact 1" about the long-run property of the Y/L path for the more developed countries. Japan had an extraordinarily high growth rate of GDP per capita for a couple of decades after World War II, usually explained by fast technology transfer from the most developed countries (the catching-up process which can only last until the technology gap is eliminated). Fig. 4.3 gives rough support for a part of Kaldor's "fact 2", namely the claim about long-run constancy of the labor income share of national income. "Fact 3" about large diversity across countries regarding the growth rate of Y/L over long time intervals is well documented empirically.[3]

It is fair to add, however, that the claimed regularities 1 and 2 do not fit all developed countries equally well. While Solow's famous growth model (Solow, 1956) can be seen as the first successful attempt at building a model consistent with Kaldor's "stylized facts", Solow himself once remarked about them: "There is no doubt that they are stylized, though it is possible to question whether they are facts" (Solow, 1970). Recently, many empiricists (see Literature notes) have questioned the methods which standard national income accounting applies

[2] In this formula w is the real wage and δ is the capital depreciation rate. Land is ignored. For countries where land is a quantitatively important production factor, the denominator should be replaced by $K + p_J J$, where J is land and p_J is the real price of land, J.

[3] For a summary, see Pritchett (1997).

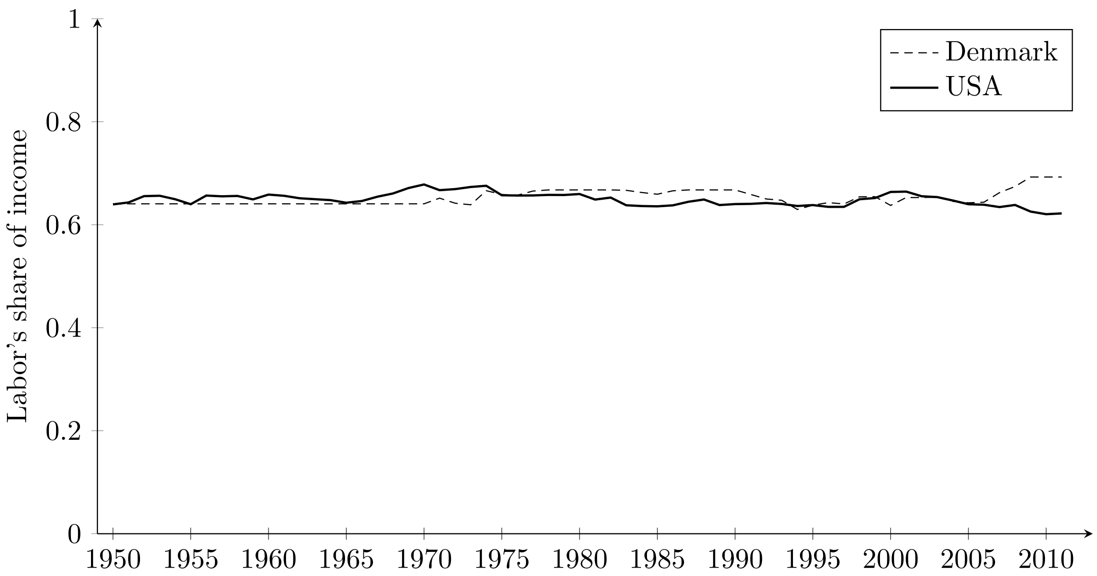


Figure 4.3: Labor's share of GDP in USA (1950-2011) and Denmark (1970-2011). Source: Feenstra, Inklaar and Timmer (2013), www.ggdc.net/pwt.

to separate the income of entrepreneurs, sole proprietors, and unincorporated businesses into labor and capital income. It is claimed that these methods obscure a tendency of the labor income share to fall in recent decades.

Notwithstanding these ambiguities, it is definitely a fact that many long-run models are constructed so a to comply with Kaldor's stylized facts. Let us briefly take a look at the Solow model (in discrete time) and check its consistency with Kaldor's "stylized facts". The point of departure of the Solow model and many other growth models is the *dynamic resource constraint for a closed economy*:

$$K_{t+1} - K_t = I_t - \delta K_t = S_t - \delta K_t \equiv Y_t - C_t - \delta K_t, \qquad K_0 > 0 \text{ given}, \tag{4.3}$$

where I_t is gross investment, which in a closed economy equals gross saving, $S_t \equiv Y_t - C_t$; δ is a constant capital depreciation rate, $0 \leq \delta \leq 1$.

The Solow model and Kaldor's stylized facts

As is well-known, the Solow model postulates a constant aggregate saving-income *ratio*, $\hat{s}$, so that $S_t = \hat{s} Y_t$, $0 < \hat{s} < 1$.[4] Further, the model assumes that the aggregate production function is neoclassical and features Harrod-neutral technological progress. So, let F in (4.2) be Solow's production function. To this Solow adds assumptions of CRS and exogenous geometric growth in both the technology

[4] Note that $\hat{s}$ is a *ratio* while the s in the Diamond model stands for the *saving* per young.

level T and the labour force L, i.e., $T_t = T_0(1+g)^t$, $g \geq 0$, and $L_t = L_0(1+n)^t$, $n > -1$. In view of CRS, we have $Y = F(K, AL) = TLF(\tilde{k}, 1) \equiv TLf(\tilde{k})$, where $\tilde{k} \equiv K/(TL)$ is the *effective capital-labor ratio* while $f' > 0$ and $f'' < 0$.

Substituting $S_t = \hat{s}Y_t$ into $K_{t+1} - K_t = S_t - \delta K_t$, dividing through by $T_t(1+g)L_t(1+n)$ and rearranging gives the "law of motion" of the Solow economy:

$$\tilde{k}_{t+1} = \frac{\hat{s}f(\tilde{k}_t) + (1-\delta)\tilde{k}_t}{(1+g)(1+n)} \equiv \varphi(\tilde{k}_t). \tag{4.4}$$

Defining $G \equiv (1+g)(1+n)$, we have $\varphi'(\tilde{k}) = (\hat{s}f'(\tilde{k}) + 1 - \delta)/G > 0$ and $\varphi''(\tilde{k}) = \hat{s}f''(\tilde{k})/G < 0$. If $G > 1 - \delta$ and f satisfies the Inada conditions $\lim_{\tilde{k}\to 0} f'(\tilde{k}) = \infty$ and $\lim_{\tilde{k}\to\infty} f'(\tilde{k}) = 0$, there is a unique and globally asymptotically stable steady state $\tilde{k}^* > 0$. The transition diagram looks entirely as in Fig. 3.4 of the previous chapter (ignoring the tildes).[5] The convergence of $\tilde{k}$ to $\tilde{k}^*$ implies that in the long run we have $K/L = \tilde{k}^*T$ and $Y/L = f(\tilde{k}^*)T$. Both K/L and Y/L are consequently growing at the same constant rate as T, the rate g. And constancy of $\tilde{k}$ implies that $Y/K = f(\tilde{k})/\tilde{k}$ is constant and so is the labor income share, $wL/Y = (f(\tilde{k}) - \tilde{k}f'(\tilde{k}))/f(\tilde{k})$, and hence also the net rate of return, $(1 - wL/Y)Y/K - \delta$.

It follows that the Solow model complies with the stylized facts 1 and 2 above. Many different models aim at doing that. What these models must then have *in common* is a capability of generating *balanced growth*.

Balanced growth

With K_t, Y_t, and C_t denoting aggregate capital, output, and consumption as above, we define a balanced growth path the following way:

DEFINITION 1 A *balanced growth path*, BGP, is a path $\{(K_t, Y_t, C_t)\}_{t=0}^{\infty}$ along which the variables K_t, Y_t, and C_t are positive and grow at constant rates (not necessarily positive).

At least for a closed economy there is a general equivalence relationship between balanced growth and constancy of certain key ratios like Y/K and C/Y. This relationship is an implication of accounting based on the above aggregate dynamic resource constraint (4.3).

For an arbitrary variable $x_t \in \mathbb{R}_{++}$, we define $\Delta x_t \equiv x_t - x_{t-1}$. Whenever $x_{t-1} > 0$, the *growth rate* of x from $t-1$ to t, denoted $g_x(t)$, is defined by $g_x(t)$

[5] What makes the Solow model so easily tractable compared to the Diamond OLG model is the constant saving-income ratio which makes the transition function essentially dependent only on the production function in intensive form. Owing to dimishing marginal productivity of capital, this is a strictly concave function. Anyway, the Solow model emerges as a special case of the Diamond model, see Exercise IV.??.

$\equiv \Delta x_t/x_{t-1}$. When there is no risk of confusion, we suppress the explicit dating and write $g_x \equiv \Delta x/x$.

PROPOSITION 1 *(the balanced growth equivalence theorem).* Let $P \equiv \{(K_t, Y_t,\ C_t)\}_{t=0}^{\infty}$ be a path along which K_t, Y_t, C_t, and $S_t \equiv Y_t - C_t$ are positive for all $t = 0, 1, 2, \ldots$. Then, given the dynamic resource constraint for a closed economy, (4.3), the following holds:
(i) If P is a BGP, then $g_Y = g_K = g_C$ and the ratios Y/K and C/Y are constant.
(ii) If Y/K and C/Y are constant, then P is a BGP with $g_Y = g_K = g_C$, i.e., not only is balanced growth present but the constant growth rates of Y, K, and C are the same.

Proof Consider a path $\{(K_t, Y_t, C_t)\}_{t=0}^{\infty}$ along which K, Y, C, and $S_t \equiv Y - C_t$ are positive for all $t = 0, 1, 2, \ldots$.

(i) Suppose the path is a balanced growth path. Then, by definition, g_Y, g_K, and g_C are constant. Hence, by (4.3), $S/K = g_K + \delta$ must be constant, implying[6]

$$g_S = g_K. \tag{*}$$

By (4.3), $Y \equiv C + S$, and so

$$\begin{aligned} g_Y &= \frac{\Delta Y}{Y} = \frac{\Delta C}{Y} + \frac{\Delta S}{Y} = \frac{C}{Y} g_C + \frac{S}{Y} g_S = \frac{C}{Y} g_C + \frac{S}{Y} g_K \qquad \text{(by (*))} \\ &= \frac{C}{Y} g_C + \frac{Y - C}{Y} g_K = \frac{C}{Y}(g_C - g_K) + g_K. \qquad (**) \end{aligned}$$

Let us provisionally assume that $g_C \neq g_K$. Then (**) gives

$$\frac{C}{Y} = \frac{g_Y - g_K}{g_C - g_K}, \tag{***}$$

a constant since g_Y, g_K, and g_C are constant. Constancy of C/Y requires $g_C = g_Y$, hence, by (***), $C/Y = 1$, i.e., $C = Y$. In view of $Y \equiv C + S$, however, this implication contradicts the given condition that $S > 0$. Hence, our provisional assumption and its implication (***) are falsified. Instead we have $g_C = g_K$. By (**), this implies $g_Y = g_K = g_C$, but now without the condition $C/Y = 1$ being implied. It follows that Y/K and C/Y are constant.

(ii) Suppose Y/K and C/Y are positive constants. Applying that the ratio between two variables is constant if and only if the variables have the same (not necessarily constant or positive) growth rate, we can conclude that $g_Y = g_K = g_C$.

[6] The ratio between two positive variables is constant if and only if the variables have the same growth rate (not necessarily constant or positive). For this and similar simple growth-arithmetic rules, see Appendix A.

By constancy of C/Y follows that $S/Y \equiv 1-C/Y$ is constant. So $g_S = g_Y = g_K$, which in turn implies that S/K is constant. By (4.3),

$$\frac{S}{K} = \frac{\Delta K + \delta K}{K} = g_K + \delta,$$

so that also g_K is constant. This, together with constancy of Y/K and C/Y, implies that also g_Y and g_C are constant. $\square$

Remark. It is part (i) of the proposition which requires the assumption $S > 0$ for all $t \geq 0$. If $S = 0$, we would have $g_K = -\delta$ and $C \equiv Y - S = Y$, hence $g_C = g_Y$ for all $t \geq 0$. Then there would be balanced growth if the common value of C and Y had a constant growth rate. This growth rate, however, could easily differ from that of K. Suppose $Y = AK^\alpha L^{1-\alpha}$, $0 < \alpha < 1$, $g_A = \gamma$ and $g_L = n$, where γ and n are constants. By the product and power function rule (see Appendix A), we would then have $1 + g_C = 1 + g_Y = (1+\gamma)(1-\delta)^\alpha(1+n)^{1-\alpha}$, which could easily be larger than 1 and thereby different from $1 + g_K = 1 - \delta \leq 1$ so that (i) no longer holds. *Example:* If $\delta = n = 0 < \gamma$, then $1 + g_Y = 1 + \gamma > 1 = 1 + g_K$.

It is part (ii) of the proposition which requires the assumption of a closed economy. In an open economy we do not necessarily have $I = S$, hence constancy of S/K no longer implies constancy of $g_K = I/K - \delta$. $\square$

For many long-run closed-economy models, including the Diamond OLG model, it holds that if and only if the dynamic system implied by the model is in a steady state, will the economy feature balanced growth, cf. Proposition 4 below. There *exist* cases, however, where this equivalence between steady state and balanced growth does not hold (some open economy models and some models with *embodied* technological change). Hence, we shall maintain a distinction between the two concepts.

Note that Proposition 1 pertains to *any* model for which (4.3) is valid. No assumption about market form and economic agents' behavior are involved. And except for the assumed constancy of the capital depreciation rate δ, no assumption about the technology is involved, not even that constant returns to scale is present.

Proposition 1 suggests that if one accepts Kaldor's stylized facts as a rough description of more than a century's growth experience and therefore wants the model to be consistent with them, one should construct the model so that it can generate balanced growth.

Balanced growth requires Harrod-neutrality

Our next proposition states that for a model to be capable of generating balanced growth, technological progress *must* take the Harrod-neutral form (i.e., be labor-augmenting). Also this proposition holds in a fairly general setting, but not as general as that of Proposition 1. Constant returns to scale and a constant growth rate in the labor force, two aspects about which Proposition 1 is silent, will now have a role to play.[7]

Consider an aggregate production function

$$Y_t = \tilde{F}(K_t, BL_t, t), \qquad B > 0,\ \tilde{F}_2' \geq 0, \tilde{F}_3' > 0, \tag{4.5}$$

where B is a constant that depends on measurement units, and the function $\tilde{F}$ is homogeneous of degree one with respect to the first two arguments (CRS) and is non-decreasing in its second argument and increasing in the third, time. The latter property represents technological progress: as time proceeds, unchanged inputs of capital and labor result in more and more output. Note that $\tilde{F}$ need not be neoclassical.

Let the labor force change at a constant rate:

$$L_t = L_0(1+n)^t, \qquad n > -1, \tag{4.6}$$

where $L_0 > 0$. The Japanese economist Hirofumi Uzawa (1928-) is famous for several contributions, not least his balanced growth theorem (Uzawa 1961).

PROPOSITION 2 *(Uzawa's balanced growth theorem).* Let $P \equiv \{(K_t, Y_t, C_t)\}_{t=0}^{\infty}$ be a path along which K_t, Y_t, C_t, and $S_t \equiv Y_t - C_t$ are positive for all $t = 0, 1, 2, \ldots$, and satisfy the dynamic resource constraint for a closed economy, (4.3), given the production function (4.5) and the labor force (4.6). Then:
(i) A *necessary* condition for the path P to be a BGP is that along P it holds that

$$Y_t = \tilde{F}(K_t, T_t L_t, 0), \tag{4.7}$$

where $T_t = T_0(1+g)^t$ with $T_0 = B$ and $1 + g \equiv (1+g_Y)/(1+n) > 1$, g_Y being the constant growth rate of output along the BGP.
(ii) Assume $(1+g)(1+n) > 1 - \delta$. Then, for any $g \geq 0$ such that there is a $q > (1+g)(1+n) - (1-\delta)$ with the property that the production function $\tilde{F}$ in (4.5) allows an output-capital ratio equal to q at $t = 0$ (i.e., $\tilde{F}(1, \tilde{k}^{-1}, 0) = q$ for some real number $\tilde{k} > 0$), a *sufficient* condition for $\tilde{F}$ to be compatible with a BGP with output-capital ratio equal to q is that $\tilde{F}$ can be written as in (4.7) with $T_t = B(1+g)^t$.

Proof (i) Suppose the given path $\{(K_t, Y_t, C_t)\}_{t=0}^{\infty}$ is a BGP. By definition, g_K and g_Y are then constant so that $K_t = K_0(1+g_K)^t$ and $Y_t = Y_0(1+g_Y)^t$. With $t = 0$ in (4.5) we then have

$$Y_t(1+g_Y)^{-t} = Y_0 = \tilde{F}(K_0, BL_0, 0) = \tilde{F}(K_t(1+g_K)^{-t}, BL_t(1+n)^{-t}, 0). \tag{4.8}$$

[7] On the other hand we do *not* imply that CRS is *always* necessary for a balanced growth path (see Exercise 4.??).

In view of the assumption that $S_t \equiv Y_t - C_t > 0$, we know from (i) of Proposition 1, that Y/K is constant so that $g_Y = g_K$. By CRS, (4.8) then implies

$$\tilde{F}(K_t, B(1+g_Y)^t(1+n)^{-t}L_t, 0) = (1+g_Y)^t Y_0 = Y_t.$$

As $1 + g \equiv (1+g_Y)/(1+n)$, this implies

$$Y_t = \tilde{F}(K_t, B(1+g)^t L_t, 0) = \tilde{F}(K_t, BL_t, t),$$

where the last equality comes from combining the first equality with (4.5). Now, the first equality shows that (4.7) holds for $T_t = B(1+g)^t = T_0(1+g)^t$. By $\tilde{F}_3'$ $(= \partial\tilde{F}/\partial t) > 0$ follows that for K_t and L_t fixed over time, Y_t is rising over time. For this to be consistent with the first equality, we must have $g > 0$.

(ii) See Appendix B. □

The form (4.7) indicates that along a BGP, technological progress must be Harrod-neutral, and we can interpret the variable T as the "technology level". By defining a new CRS production function F by $F(K_t, T_tL_t) \equiv \tilde{F}(K_t, T_tL_t, 0)$, we see that (i) of the proposition implies that at least along the BGP, we can rewrite the original production function this way:

$$Y_t = \tilde{F}(K_t, BL_t, t) = \tilde{F}(K_t, T_tL_t, 0) \equiv F(K_t, T_tL_t). \tag{4.9}$$

where F has CRS, and $T_t = T_0(1+g)^t$, with $T_0 = B$ and $1+g \equiv (1+g_Y)/(1+n)$.

What is the intuition behind the Uzawa result that for balanced growth to be possible, technological progress must at the aggregate level have the purely labor-augmenting form? We may first note that there is an asymmetry between capital and labor. Capital is an accumulated amount of non-consumed output and has thus at least a "tendency" to inherit the trend in output. In contrast, labor is a non-produced production factor. The labor force grows in an exogenous way and does *not* inherit the trend in output. Indeed, the ratio L_t/Y_t is free to adjust as t proceeds.

More specifically, consider the point of departure, the original production function (4.5). Because of CRS, it must satisfy

$$1 = \tilde{F}\left(\frac{K_t}{Y_t}, \frac{BL_t}{Y_t}, t\right). \tag{4.10}$$

We know from Proposition 1 that along a BGP, K_t/Y_t is constant. The assumption $\tilde{F}_3'$ $(= \partial\tilde{F}/\partial t) > 0$ implies that technological progress is present. Along a BGP, this progress must manifest itself in the form of a compensating change in L_t/Y_t in (4.10) as t proceeds, because otherwise the right-hand side of (4.10) would increase, which would contradict the constancy of the left-hand side. As we have in (4.5) assumed $\partial \tilde{F}/\partial L \geq 0$, the needed change in L_t/Y_t is a *fall*. The fall in L_t/Y_t must exactly offset the effect on $\tilde{F}$ of the rising t, when there is a fixed capital-output ratio and the left-hand side of (4.10) remains unchanged. It follows that along the considered BGP, L_t/Y_t is a decreasing function of t. The inverse, Y_t/L_t, is thus an *increasing* function of t. If we denote this function T_t, we end up with (4.9).

The generality of Uzawa's theorem is noteworthy. Like Proposition 1, Uzawa's theorem is about technically feasible paths, while economic institutions, market forms, and agents' behavior are not involved. The theorem presupposes CRS, but does not need that the technology has neoclassical properties not to speak of satisfying the Inada conditions. And the theorem holds for exogenous as well as endogenous technological progress.

A simple implication of the theorem is the following. Let y_t denote "labor productivity" in the sense of Y_t/L_t, k_t denote the capital-labor ratio, K_t/L_t, and c_t the consumption-labor ratio, C_t/L_t. We have:

COROLLARY Consider an economy with labor force satisfying (4.6) and a CRS-production function $Y_t = F(K_t, T_t L_t)$, $F_2' > 0$. Along a BGP with positive gross saving and the technology level T growing at a constant rate $g \geq 0$, output grows at the rate $(1+g)(1+n)-1$ ($\approx g+n$ for g and n "small") while labor productivity, y, capital-labor ratio, k, and consumption-labor ratio, c, all grow at the rate g.

Proof That $g_Y = (1+g)(1+n) - 1$ follows from (i) of Proposition 2. As to g_y we have

$$y_t \equiv \frac{Y_t}{L_t} = \frac{Y_0(1+g_Y)^t}{L_0(1+n)^t} = y_0(1+g)^t,$$

since $1+g = (1+g_Y)/(1+n)$. This shows that y grows at the rate g. Moreover, $y/k = Y/K$, which is constant along a BGP, by (i) of Proposition 1. Hence k grows at the same rate as y. Finally, also $c/y \equiv C/Y$ is constant along a BGP, implying that also c grows at the same rate as y. □

Factor income shares

There is one facet of Kaldor's stylized facts which we have not yet related to Harrod-neutral technological progress, namely the claimed long-run "approximate" constancy of both the income share of labor and the rate of return on capital. It turns out that, if we assume (a) neoclassical technology, (b) profit maximizing firms, and (c) perfect competition in the output and factor markets, then these constancies are inherent in the combination of constant returns to scale and balanced growth.

To see this, let the aggregate production function be

$$Y_t = F(K_t, T_t L_t), \tag{4.11}$$

where F is neoclassical and has CRS. In view of perfect competition, the representative firm chooses inputs such that

$$\frac{\partial Y_t}{\partial K_t} = F_1(K_t, T_t L_t) = r_t + \delta, \qquad \text{and,} \tag{4.12}$$

$$\frac{\partial Y_t}{\partial L_t} = F_2(K_t, T_t L_t) T_t = w_t, \tag{4.13}$$

where the right-hand sides indicate the factor prices, r_t being the interest rate, δ the depreciation rate, and w_t the real wage.

In equilibrium the labor income share will be

$$\frac{w_t L_t}{Y_t} = \frac{\frac{\partial Y_t}{\partial L_t} L_t}{Y_t} = \frac{F_2(K_t, T_t L_t) T_t L_t}{Y_t}. \tag{4.14}$$

Since land as a production factor is ignored, gross capital income equals non-labor income, $Y_t - w_t L_t$. Denoting the gross capital income share by α_t, we thus have

$$\begin{aligned} \alpha_t &= \frac{Y_t - w_t L_t}{Y_t} = \frac{F(K_t, T_t L_t) - F_2(K_t, T_t L_t) T_t L_t}{Y_t} \\ &= \frac{F_1(K_t, T_t L_t) K_t}{Y_t} = \frac{\frac{\partial Y_t}{\partial K_t} K_t}{Y_t} = (r_t + \delta) \frac{K_t}{Y_t}, \end{aligned} \tag{4.15}$$

where we have used (4.13), Euler's theorem,[8] and then (4.12). Finally, when the capital good is nothing but a non-consumed output good, it has price equal to 1, and so the economy-wide rate of return on capital can be written

$$\frac{Y_t - w_t L_t - \delta K_t}{1 \cdot K_t} = \frac{Y_t - w_t L_t}{Y_t} \cdot \frac{Y_t}{K_t} - \delta = \alpha_t \cdot \frac{Y_t}{K_t} - \delta = r_t, \tag{4.16}$$

where the last equality comes from (4.15).

PROPOSITION 3 *(factor income shares under perfect competition)* Let the dynamic resource constraint for a closed economy be given as in (4.3). Assume F is neoclassical with CRS, and that the economy is competitive. Let the path $P = \{(K_t, Y_t, C_t)\}_{t=0}^{\infty}$ be BGP with positive gross saving. Then, along the path P:
(i) The gross capital income share equals some constant $\alpha \in (0, 1)$, and the labor income then equals $1 - \alpha$.
(ii) The rate of return on capital is $\alpha q - \delta$, where q is the constant output-capital ratio along the BGP.

[8] Indeed, from Euler's theorem follows that $F_1 K + F_2 TL = F(K, TL)$, when F is homogeneous of degree one.

Proof In view of CRS, $Y_t = F(K_t, T_tL_t) = T_tL_tF(\tilde{k}_t, 1) \equiv T_tL_tf(\tilde{k}_t)$, where $\tilde{k}_t \equiv K_t/(T_tL_t)$, and $f' > 0$, $f'' < 0$. From Proposition 1 follows that along the given path P, which is a BGP, Y_t/K_t is some constant, say q, equal to $f(\tilde{k}_t)/\tilde{k}_t$. Hence, $\tilde{k}_t$ is constant, say equal to $\tilde{k}^*$. Consequently, along P, $\partial Y_t/\partial K_t = f'(\tilde{k}^*) = r_t + \delta$. From this follows that r_t is a constant, r. (i) From (4.15) now follows that $\alpha_t = f'(\tilde{k}^*)/q \equiv \alpha$. Moreover, $0 < \alpha < 1$, since $0 < \alpha$ is implied by $f' > 0$, and $\alpha < 1$ is implied by the fact that $f'(\tilde{k}^*) < f(\tilde{k}^*)/\tilde{k}^* = Y/K = q$, where "$<$" is due to $f'' < 0$ and $f(0) \geq 0$ (draw the graph of $f(\tilde{k})$). By the first equality in (4.15), the labor income share can be written $w_tL_t/Y_t = 1 - \alpha_t = 1 - \alpha$. (ii) Consequently, by (4.16), the rate of return on capital equals r_t $(= r) = \alpha q - \delta$. □

What this proposition amounts to is that a BGP in this economy exhibits both the first and the second "Kaldor fact" (point 1 and 2, respectively, in the list at the beginning of the chapter).

Although the proposition implies constancy of the factor income shares under balanced growth, it does not *determine* them. The proposition expresses the factor income shares in terms of the unknown constants α and q. These constants will generally depend on the effective capital-labor ratio in steady state, $\tilde{k}^*$, which will generally be an unknown as long as we have not formulated a theory of saving. This takes us back to Diamond's OLG model which provides such a theory.

4.2 The Diamond OLG model with Harrod-neutral technological progress

Recall from the previous chapter that in the Diamond OLG model people live in two periods, as young and as old. Only the young work and each young supplies one unit of labor inelastically. The period utility function, $u(c)$, satisfies the No Fast Assumption. The saving function of the young is $s_t = s(w_t, r_{t+1})$. We now include Harrod-neutral technological progress in the Diamond model.

Let (4.11) be the aggregate production function in the economy and assume, as before, that F is neoclassical with CRS. The technology level T_t grows at a constant exogenous rate:

$$T_t = T_0(1+g)^t, \qquad g \geq 0. \tag{4.17}$$

The initial level of technology, T_0, is historically given. The employment level L_t equals the number of young and thus grows at the constant exogenous rate $n > -1$.

Suppressing for a while the explicit dating of the variables, in view of CRS with respect to K and TL, we have

$$\tilde{y} \equiv \frac{Y}{TL} = F(\frac{K}{TL}, 1) = F(\tilde{k}, 1) \equiv f(\tilde{k}), \qquad f' > 0, f'' < 0,$$

where TL is *labor input in efficiency units;* $\tilde{k} \equiv K/(TL)$ is known as the *effective* or *technology-corrected capital-labor ratio* - also sometimes just called the "capital intensity". There is perfect competition in all markets. In each period the representative firm maximizes profit, $\Pi = F(K, TL) - \hat{r}K - wL$. Given the constant capital depreciation rate $\delta \in [0, 1]$, this leads to the first-order conditions

$$\frac{\partial Y}{\partial K} = \frac{\partial \left[TLf(\tilde{k})\right]}{\partial K} = f'(\tilde{k}) = r + \delta, \tag{4.18}$$

and

$$\frac{\partial Y}{\partial L} = \frac{\partial \left[TLf(\tilde{k})\right]}{\partial L} = \left[f(\tilde{k}) - f'(\tilde{k})\tilde{k}\right] T = w. \tag{4.19}$$

In view of $f'' < 0$, a $\tilde{k}$ satisfying (4.18) is unique. We let its value in period t be denoted $\tilde{k}_t^d$. Assuming equilibrium in the factor markets, this desired effective capital-labor ratio equals the effective capital-labor ratio from the supply side, $\tilde{k}_t \equiv K_t/(T_t L_t) \equiv k_t/T_t$, which is predetermined in every period. The equilibrium interest rate and real wage in period t are thus determined by

$$r_t = f'(\tilde{k}_t) - \delta \equiv r(\tilde{k}_t), \quad \text{where } r'(\tilde{k}_t) = f''(\tilde{k}_t) < 0, \tag{4.20}$$

$$w_t = \left[f(\tilde{k}_t) - f'(\tilde{k}_t)\tilde{k}\right] T_t \equiv \tilde{w}(\tilde{k}_t)T_t, \quad \text{where } \tilde{w}'(\tilde{k}_t) = -\tilde{k}_t f''(\tilde{k}_t) > 0. \tag{4.21}$$

Here, $\tilde{w}(\tilde{k}_t) = w_t/T_t$ is known as the *technology-corrected real wage.*

The equilibrium path

The aggregate capital stock at the beginning of period $t + 1$ must still be owned by the old generation in that period and thus equal the aggregate saving these people had as young in the previous period. Hence, as before, $K_{t+1} = s_t L_t = s(w_t, r_{t+1})L_t$. In view of $K_{t+1} \equiv \tilde{k}_{t+1}T_{t+1}L_{t+1} = \tilde{k}_{t+1}T_t(1+g)L_t(1+n)$, together with (4.20) and (4.21), we get

$$\tilde{k}_{t+1} = \frac{s(\tilde{w}(\tilde{k}_t)T_t, r(\tilde{k}_{t+1}))}{T_t(1 + g)(1 + n)}. \tag{4.22}$$

This is the general version of the law of motion of the Diamond OLG model with Harrod-neutral technological progress.

For the model to comply with Kaldor's "stylized facts", the model should be capable of generating balanced growth. Essentially, this capability is equivalent to being able to generate a steady state. In the presence of technological progress this latter capability requires a restriction on the lifetime utility function, U. Indeed, we see from (4.22) that the model is consistent with existence of a steady state only if the time-dependent technology level, T_t, in the numerator and denominator cancels out. This requires that the saving function is homogeneous of degree one in its first argument such that $s(\tilde{w}(\tilde{k}_t)T_t, r(\tilde{k}_{t+1})) = s(\tilde{w}(\tilde{k}_t), r(\tilde{k}_{t+1}))T_t$. In turn, this is so if and only if the lifetime utility function of the young is *homothetic.* So, in addition to the No Fast Assumption from Chapter 3, we impose the Homotheticity Assumption:

$$\text{the lifetime utility function } U \text{ is homothetic.} \tag{A4}$$

This property entails that if the value of the "endowment", here the human wealth w_t, is multiplied by an arbitrary constant $\lambda > 0$, then the chosen c_{1t} and c_{2t+1} will also be multiplied by this factor (see Appendix C). It then follows from the period budget constraints, $c_{1t} + s_t = w_t$ and $c_{2t+1} = (1 + r_{t+1})s_t$, that s_t is multiplied by λ as well. Letting $\lambda = 1/(\tilde{w}(\tilde{k}_t)T_t)$, (A4) thus allows us to write

$$s_t = s(1, r(\tilde{k}_{t+1}))\tilde{w}(\tilde{k}_t)T_t \equiv \hat{s}(r(\tilde{k}_{t+1}))\tilde{w}(\tilde{k}_t)T_t, \tag{4.23}$$

where $\hat{s}(r(\tilde{k}_{t+1}))$ is the saving-wealth *ratio* of the young. The distinctive feature is that the homothetic lifetime utility function U allows a decomposition of the young's saving into two factors, where one is the saving-wealth ratio, which depends only on the interest rate, and the other is the human wealth. By (4.22), the law of motion of the economy reduces to

$$\tilde{k}_{t+1} = \frac{\hat{s}(r(\tilde{k}_{t+1}))}{(1+g)(1+n)}\tilde{w}(\tilde{k}_t). \tag{4.24}$$

The equilibrium path of the economy can be analyzed in a similar way as in the case of no technological progress. In the assumptions (A2) and (A3) from Chapter 3 we replace k by $\tilde{k}$ and $1 + n$ by $(1 + g)(1 + n)$. As a generalization of Proposition 4 from Chapter 3, these generalized versions of (A2) and (A3), together with the No Fast Assumption (A1) and the Homotheticity Assumption (A4), guarantee that k_t over time converges to some steady state value $\tilde{k}^* > 0$.

Let an economy that can be described by the Diamond model be called a *Diamond economy.* Our conclusion is then that a Diamond economy will sooner or later settle down in a steady state. The convergence of $\tilde{k}$ implies convergence of many key variables, for instance the interest rate and the technology-corrected real wage. In view of (4.20) and (4.21), respectively, we get, for $t \to \infty$,

$$\begin{aligned} r_t &= f'(\tilde{k}_t) - \delta \to f'(\tilde{k}^*) - \delta \equiv r^*, \quad \text{and} \\ \frac{w_t}{T_t} &= f(\tilde{k}_t) - \tilde{k}_t f'(\tilde{k}_t) \to f(k^*) - k^* f'(k^*) \equiv \tilde{w}(\tilde{k}^*). \end{aligned}$$

Moreover, for instance the labor income share converges to a constant:

$$\frac{w_t L_t}{Y_t} = \frac{w_t/T_t}{Y_t/(T_t L_t)} = \frac{f(\tilde{k}_t) - \tilde{k}_t f'(\tilde{k}_t)}{f(\tilde{k}_t)} \to 1 - \frac{k^* f'(k^*)}{f(k^*)} \equiv 1 - \alpha^* \text{ for } t \to \infty.$$

The prediction from the model is thus that a Diamond economy will in the long run behave in accordance with Kaldor's stylized facts. The background for this is that convergence to a steady state is, in this and many other models, equivalent to "convergence" to a BGP. This equivalence follows from:

PROPOSITION 4 Consider a Diamond economy with Harrod-neutral technological progress at the constant rate $g \geq 0$ and positive gross saving for all t.
(i) If the economy features balanced growth, then it is in a steady state.
(ii) If the economy is in a steady state, then it features balanced growth.

Proof (i) Suppose the economy features balanced growth. Then, by Proposition 1, Y/K is constant. As $Y/K = \tilde{y}/\tilde{k} = f(\tilde{k})/\tilde{k}$, also $\tilde{k}$ is constant. Thereby the economy is in a steady state. (ii) Suppose the economy is in a steady state, i.e., for some $\tilde{k}^* > 0$, (4.24) holds for $\tilde{k}_t = \tilde{k}_{t+1} = \tilde{k}^*$. The constancy of $\tilde{k} \equiv K/(TL)$ and $\tilde{y} \equiv Y/(TL) = f(\tilde{k})$ implies that both g_K and g_Y equal the constant g_{TL} $= (1+g)(1+n) - 1 > 0$. As $S \equiv Y - C$, constancy of g_K implies constancy of $S/K = (\Delta K + \delta K)/K = g_K + \delta$, so that also S grows at the rate g_K and thereby at the same rate as output. Hence S/Y, and thereby also $C/Y \equiv 1 - S/Y$, is constant. Hence, also C grows at the constant rate g_Y. All criteria for a BGP are thus satisfied. □

Let us portray the dynamics by a transition diagram. Fig. 4.4 shows a "well-behaved" case in the sense that there is only one steady state and it is globally asymptotically stable. In the figure the initial effective capital-labor ratio, $\tilde{k}_0$, is assumed to be relatively large. This need not be interpreted as if the economy is highly developed and has a high initial capital-labor ratio, K_0/L_0. Indeed, the reason that $\tilde{k}_0 \equiv K_0/(T_0 L_0)$ is large relative to its steady-steady value may be that the economy is "backward" in the sense of having a relatively low initial level of technology. Growing at a given rate g, the technology will in this situation grow faster than the capital-labor ratio, K/L, so that the effective capital-labor ratio declines over time. The process continues until the steady state is essentially reached with a real interest rate $r^* = f'(\tilde{k}^*) - \delta$. This is to remind ourselves that

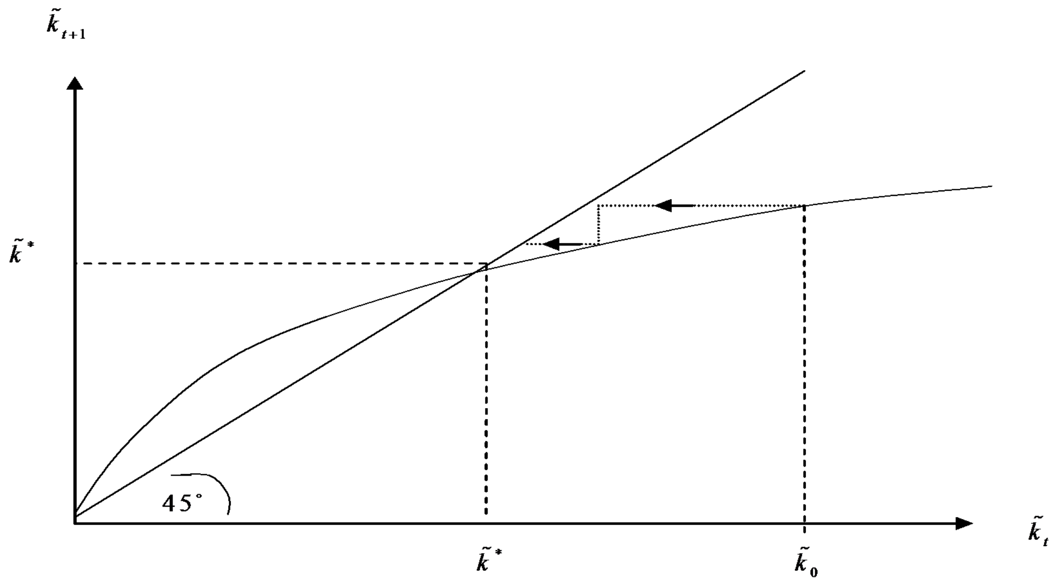


Figure 4.4: Transition curve for a "well-behaved" Diamond OLG model with Harrod-neutral technical progress.

from an empirical point of view, the adjustment towards a steady state can be from above as well as from below.

The output growth rate in steady state, $(1+g)(1+n)-1$, is sometimes called the "natural rate of growth". Since $(1+g)(1+n)-1 = g+n+gn \approx g+n$ for g and n "small", the natural rate of growth approximately equals the sum of the rate of technological progress and the growth rate of the labor force.

Warning: When measured on an *annual* basis, the growth rates of technology and labor force, $\bar{g}$ and $\bar{n}$, do indeed tend to be "small", say $\bar{g} = 0.02$ and $\bar{n} = 0.005$, so that $\bar{g}+\bar{n}+\bar{g}\bar{n} = 0.0251 \approx 0.0250 = \bar{g}+\bar{n}$. But in the context of models like Diamond's, the period length is, say, 30 years. Then the corresponding g and n will satisfy the equations $1+g = (1+\bar{g})^{30} = 1.02^{30} = 1.8114$ and $1+n = (1+\bar{n})^{30}$ $= 1.005^{30} = 1.1614$, respectively. We get $g+n = 0.973$, which is about 10 percent smaller than the true output growth rate over 30 years, which is $g+n+gn = 1.104$.

We end our account of Diamond's OLG model with some remarks on a popular special case of a homothetic utility function.

Example: CRRA utility

An example of a homothetic lifetime utility function is obtained by letting the period utility function take the CRRA form introduced in the previous chapter.

Then

$$U(c_1, c_2) = \frac{c_1^{1-\theta} - 1}{1-\theta} + (1+\rho)^{-1}\frac{c_2^{1-\theta} - 1}{1-\theta}, \qquad \theta > 0. \tag{4.25}$$

Recall that the CRRA utility function with parameter θ has the property that the (absolute) elasticity of marginal utility of consumption equals the constant $\theta > 0$ for all $c > 0$. Up to a positive linear transformation it is, in fact, the only period utility function having this property. A proof that the utility function (4.25) is indeed homothetic is given in Appendix C.

One of the reasons that the CRRA function is popular in macroeconomics is that in *representative* agent models, the period utility function *must* have this form to obtain consistency with balanced growth and Kaldor's stylized facts (this is shown in Chapter 7). In contrast, a model with heterogeneous agents, like the Diamond model, does not need CRRA utility to comply with the Kaldor facts. CRRA utility is just a convenient special case leading to homothetic lifetime utility. And *this* is what is needed for a BGP to exist and thereby for compatibility with Kaldor's stylized facts.

Given the CRRA assumption in (4.25), the saving-wealth ratio of the young becomes

$$\hat{s}(r) = \frac{1}{1 + (1+\rho)\left(\frac{1+r}{1+\rho}\right)^{(\theta-1)/\theta}}. \tag{4.26}$$

It follows that $\hat{s}'(r) \gtreqless 0$ for $\theta \lesseqgtr 1$.

When $\theta = 1$ (the case $u(c) = \ln c$), $\hat{s}(r) = 1/(2+\rho) \equiv \hat{s}$, a constant, and the law of motion (4.24) thus simplifies to

$$\tilde{k}_{t+1} = \frac{1}{(1+g)(1+n)(2+\rho)}\tilde{w}(\tilde{k}_t).$$

We see that in the $\theta = 1$ case, whatever the production function, $\tilde{k}_{t+1}$ enters only at the left-hand side of the fundamental difference equation, which thereby reduces to a simple transition function. Since $\tilde{w}'(\tilde{k}) > 0$, the transition curve is positively sloped everywhere. If the production function is Cobb-Douglas, $Y_t = K_t^\alpha (T_t L_t)^{1-\alpha}$, then $\tilde{w}(\tilde{k}_t) = (1-\alpha)\tilde{k}_t^\alpha$. Combining this with $\theta = 1$ yields a "well-behaved" Diamond model (thus having a unique and globally asymptotically stable steady state), cf. Fig. 4.4 above. In fact, as noted in Chapter 3, in combination with Cobb-Douglas technology, CRRA utility results in "well-behavedness" whatever the value of $\theta > 0$.

4.3 The golden rule under Harrod-neutral technological progress

Given that there is technological progress, consumption per unit of labor is likely to grow over time. Therefore the definition of the golden-rule capital-labor ratio from Chapter 3 has to be generalized to cover the case of growing consumption per unit of labor. To allow existence of steady states and BGPs, we maintain the assumption that technological progress is Harrod-neutral, that is, we maintain the production function (4.11) where the technology level, T, grows at a constant rate $g > 0$. We also maintain the assumption that the labor force, L_t, is fully employed and grows at a constant rate, $n \geq 0$.

Since we need not have a Diamond economy in mind, we can consider an arbitrary period length. It could be one year for instance. Consumption per unit of labor is

$$
\begin{aligned}
c_t &\equiv \frac{C_t}{L_t} = \frac{F(K_t, T_t L_t) - S_t}{L_t} = \frac{f(\tilde{k}_t) T_t L_t - (K_{t+1} - K_t + \delta K_t)}{L_t} \\
&= f(\tilde{k}_t) T_t - (1+g) T_t (1+n) \tilde{k}_{t+1} + (1-\delta) T_t \tilde{k}_t \\
&= \left[f(\tilde{k}_t) + (1-\delta)\tilde{k}_t - (1+g)(1+n)\tilde{k}_{t+1} \right] T_t.
\end{aligned}
$$

DEFINITION 2 The golden-rule capital intensity, $\tilde{k}_{GR}$, is that level of $\tilde{k} \equiv K/(TL)$ which gives the highest sustainable path for consumption per unit of labor in the economy.

To comply with the sustainability requirement, we consider a steady state. So $\tilde{k}_{t+1} = \tilde{k}_t = \tilde{k}$ and therefore

$$
c_t = \left[f(\tilde{k}) + (1-\delta)\tilde{k} - (1+g)(1+n)\tilde{k} \right] T_t \equiv \tilde{c}(\tilde{k}) T_t, \tag{4.27}
$$

where $\tilde{c}(\tilde{k})$ is the "technology-corrected" level of consumption per unit of labor in the steady state. We see that in steady state, consumption per unit of labor will grow at the same rate as the technology. Thus,

$$
\ln c_t = \ln \tilde{c}(\tilde{k}) + \ln T_0 + t \ln(1+g).
$$

Fig. 4.5 illustrates.

Since the evolution of the technology level T_t in (4.27) is exogenous, the highest possible path of c_t is found by maximizing $\tilde{c}(\tilde{k})$. This gives the first-order condition

$$
\tilde{c}'(\tilde{k}) = f'(\tilde{k}) + (1-\delta) - (1+g)(1+n) = 0. \tag{4.28}
$$

When $n \geq 0$, we have $(1+g)(1+n) - (1-\delta) > 0$ in view of $g > 0$. Then, by continuity, the equation (4.28) necessarily has a unique solution in $\tilde{k} > 0$, if the production function satisfies the condition

$$\lim_{\tilde{k}\to 0} f'(\tilde{k}) > (1+g)(1+n) - (1-\delta) > \lim_{\tilde{k}\to\infty} f'(\tilde{k}),$$

which we assume. This is a milder condition than the Inada conditions. Considering the second-order condition $\tilde{c}''(\tilde{k}) = f''(\tilde{k}) < 0$, the $\tilde{k}$ satisfying (4.28) does indeed maximize $\tilde{c}(\tilde{k})$. By definition, this $\tilde{k}$ is the golden-rule capital intensity, $\tilde{k}_{GR}$. Thus

$$f'(\tilde{k}_{GR}) - \delta = (1+g)(1+n) - 1 \approx g + n, \tag{4.29}$$

where the right-hand side is the "natural rate of growth". This says that the golden-rule capital intensity is that level of the capital intensity at which the net marginal productivity of capital equals the output growth rate in steady state.

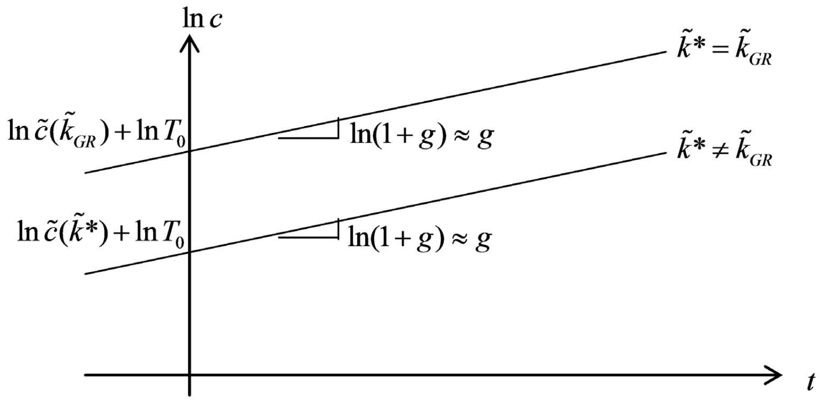


Figure 4.5: The highest sustainable path of consumption is where $\tilde{k}^* = \tilde{k}_{GR}$.

Has dynamic inefficiency been a problem in practice? As in the Diamond model without technological progress, it is theoretically possible that the economy ends up in a steady state with $\tilde{k}^* > \tilde{k}_{GR}$.[9] If this happens, the economy is dynamically inefficient and $r^* < (1+g)(1+n) - 1 \approx g + n$. To check whether dynamic inefficiency is a realistic outcome in an industrialized economy or not, we should compare the observed average GDP growth rate over a long stretch of time to the average real interest rate or rate of return in the economy. For the period after the Second World War the average GDP growth rate ($\approx g + n$) in Western countries is typically about 3 percent per year. But what interest rate

[9] The proof is analogue to that in Chapter 3 for the case $g = 0$.

should one choose? In simple macro models, like the Diamond model, there is no uncertainty and no need for money to carry out trades. In such models all assets earn the same rate of return, r, in equilibrium. In the real world there is a spectrum of interest rates, reflecting the different risk and liquidity properties of the different assets. The expected real rate of return on a short-term government bond is typically less than 3 percent per year (a relatively safe and liquid asset). This is much lower than the expected real rate of return on corporate stock, say 10 percent per year. Our model cannot tell which rate of return we should choose, but the conclusion hinges on that choice.

Abel et al. (1989) study the problem on the basis of a model with *uncertainty.* They show that a sufficient condition for dynamic efficiency is that gross investment, I, does not exceed the gross capital income in the long run, that is $I \leq Y - \partial Y/\partial L \approx Y - wL$. They find that for the U.S. and six other major OECD nations this seems to hold. Indeed, for the period 1929-85 the U.S. has, on average, $I/Y = 0.15$ and $(Y - wL)/Y = 0.29$. A quite similar difference is found for other industrialized countries, suggesting that they do not suffer from dynamic inefficiency. At least in these countries, therefore, the potential coordination failure laid bare by OLG models does not seem to have been operative in practice.

4.4 The functional income distribution

By the *functional income distribution* is meant the distribution of national income on the different basic income categories: income to providers of labor, capital, and land (including other natural resources). Theory of the functional income distribution is thus theory about the determination and evolution of factor income shares.

The simplest theory about the functional income distribution is the *neoclassical theory* of the functional income distribution. It relies on competitive markets and an aggregate production function,

$$Y = F(K, L, J),$$

where K and L have the usual meaning, but the new symbol J measures the input of land. The production function is assumed to be neoclassical with CRS. Until further notice, we ignore technological change. When the representative firm maximizes profit and the factor markets clear, the equilibrium factor prices must satisfy:

$$\hat{r} = F_K(K, L, J), \quad w = F_L(K, L, J), \quad z = F_J(K, L, J),$$

where z denotes land rent (the charge for the use of land per unit of land), and, as usual, $\hat{r}$ is the rental rate per unit of capital, and w is the real wage per unit of labor. If in the given period, the supply of three production factors is predetermined, the three equations *determine* the three factor prices, and the factor income shares are determined as

$$\frac{\hat{r}K}{Y} = \frac{F_K(K,L,J)K}{F(K,L,J)}, \quad \frac{wL}{Y} = \frac{F_L(K,L,J)L}{F(K,L,J)}, \quad \frac{zJ}{Y} = \frac{F_J(K,L,J)J}{F(K,L,J)}.$$

The theory is also called the *marginal productivity theory* of the functional income distribution.

In advanced economies the role of land is relatively minor.[10] In fact many theoretical models completely ignore land. Below we follow that tradition, while considering the question: How is the direction of movement of the labor and capital income shares, respectively, determined during the adjustment process from arbitrary initial conditions toward steady state?

(currently here a gap in the manuscript)

How the labor income share depends on the capital-labor ratio

Ignoring, to begin with, technological progress, we write aggregate output as $Y = F(K,L)$, where F is neoclassical with CRS. From Euler's theorem follows that $F(K,L) = F_1K + F_2L = f'(k)K + (f(k) - kf'(k))L$, where $k \equiv K/L$ and f is the production function in intensive form. In equilibrium under perfect competition we have

$$Y = \hat{r}K + wL,$$

where $\hat{r} = r + \delta = f'(k) \equiv \hat{r}(k)$ and $w = f(k) - kf'(k) \equiv w(k)$.

The *labor income share* is

$$\frac{wL}{Y} = \frac{f(k) - kf'(k)}{f(k)} \equiv \frac{w(k)}{f(k)} \equiv SL(k) = \frac{wL}{\hat{r}K + wL} = \frac{\frac{w/\hat{r}}{k}}{1 + \frac{w/\hat{r}}{k}}, \tag{4.30}$$

where the function $SL(\cdot)$ is the income share of labor function, $w/\hat{r}$ is the *factor price ratio*, and $(w/\hat{r})/k = w/(\hat{r}k)$ is the *factor income ratio*. As $\hat{r}'(k) = f''(k) < 0$ and $w'(k) = -kf''(k) > 0$, the relative factor price $w/\hat{r}$ is an increasing function of k.

[10] In 1750 land rent made up 20 percent of national income in England, in 1850 8 percent, and in 2010 less than 0.1 percent (Jones and Vollrath, 2013). The approximative numbers often used for the labor income share and capital income share in advanced economies are 2/3 and 1/3, respectively.

Suppose that capital tends to grow faster than labor so that k rises over time. Unless the production function is Cobb-Douglas, this will under perfect competition affect the labor income share. But apriori it is not obvious in what direction. By (4.30) we see that the labor income share moves in the same direction as the factor *income* ratio, $(w/\hat{r})/k$. The latter goes up (down) depending on whether the percentage rise in the factor price ratio $w/\hat{r}$ is greater (smaller) than the percentage rise in k. So, if we let $\mathrm{E}\ell_x g(x)$ denote the elasticity of a function $g(x)$ w.r.t. x, that is, $xg'(x)/g(x)$), we can only say that

$$SL'(k) \gtreqless 0 \text{ for } \mathrm{E}\ell_k \frac{w}{\hat{r}} \gtreqless 1, \tag{4.31}$$

respectively. In words: if the production function is such that the ratio of the marginal productivities of the two production factors is strongly (weakly) sensitive to the capital-labor ratio, then the labor income share rises (falls) along with a rise in K/L.

Usually, however, the inverse elasticity is considered, namely $\mathrm{E}\ell_{w/\hat{r}} k$ $(= 1/\mathrm{E}\ell_k \frac{w}{\hat{r}})$. This elasticity indicates how sensitive the cost minimizing capital-labor ratio, k, is to a given factor price ratio $w/\hat{r}$. Under perfect competition $\mathrm{E}\ell_{w/\hat{r}} k$ coincides with what is known as the *elasticity of factor substitution* (for a general definition, see below). The latter is often denoted σ. In the CRS case, σ will be a function of only k so that we can write $\mathrm{E}\ell_{w/\hat{r}} k = \sigma(k)$. By (4.31), we therefore have

$$SL'(k) \gtreqless 0 \text{ for } \sigma(k) \lesseqgtr 1,$$

respectively.

The size of the elasticity of factor substitution is a property of the production function, hence of the technology. In special cases the elasticity of factor substitution is a constant, i.e., independent of k. For instance, if F is Cobb-Douglas, i.e., $Y = K^{\alpha} L^{1-\alpha}, 0 < \alpha < 1$, we have $\sigma(k) \equiv 1$, as we will see in Section 4.6. In this case variation in k does not change the labor income share under perfect competition. Empirically there is not agreement about the "normal" size of the elasticity of factor substitution for industrialized economies, but the bulk of studies seems to conclude with $\sigma(k) < 1$ (see below).

Adding Harrod-neutral technical progress We now add Harrod-neutral technical progress. We write aggregate output as $Y = F(K, TL)$, where F is neoclassical with CRS, and $T = T_t = T_0(1+g)^t$, $g \geq 0$. Then the labor income share is

$$\frac{wL}{Y} = \frac{w/T}{Y/(TL)} \equiv \frac{\tilde{w}}{\tilde{y}}.$$

The above formulas still hold if we replace k by $\tilde{k} \equiv K/(TL)$ and w by $\tilde{w} \equiv w/T$. We get

$$SL'(\tilde{k}) \gtreqless 0 \text{ for } \sigma(\tilde{k}) \lesseqgtr 1,$$

respectively. We see that if $\sigma(\tilde{k}) < 1$ in the relevant range for $\tilde{k}$, then market forces tend to *increase* the income share of the factor that is becoming relatively more scarce. This factor is efficiency-adjusted labor, TL, if $\tilde{k}$ is increasing, which $\tilde{k}$ will be during the transitional dynamics in a well-behaved Diamond model if $\tilde{k}_0 < \tilde{k}^*$. And if instead $\sigma(\tilde{k}) > 1$ in the relevant range for $\tilde{k}$, then market forces tend to *decrease* the income share of the factor that is becoming relatively more scarce. This factor is K, if $\tilde{k}$ is decreasing, which $\tilde{k}$ will be during the transitional dynamics in a well-behaved Diamond model if $\tilde{k}_0 > \tilde{k}^*$, cf. Fig. 4 above. Note that, given the production function in intensive form, f, the elasticity of substitution between capital and labor does not depend on whether $g = 0$ or $g > 0$, but only on the function f itself and the level of $K/(TL)$. This follows from Section 4.6.

While k empirically is clearly growing, $\tilde{k} \equiv k/T$ is not necessarily so because also T is increasing. Indeed, according to Kaldor's "stylized facts", apart from short- and medium-term fluctuations, $\tilde{k}$ − and therefore also $\hat{r}$ and the labor income share − tend to be more or less constant over time. This can happen whatever the sign of $\sigma(\tilde{k}^*) - 1$, where $\tilde{k}^*$ is the long-run value of the effective capital-labor ratio $\tilde{k}$.

As alluded to earlier, there are empiricists who reject Kaldor's "facts" as a general tendency. For instance Piketty (2014) essentially claims that in the very long run the effective capital-labor ratio $\tilde{k}$ has an upward trend, temporarily braked by two world wars and the Great Depression in the 1930s. If so, the sign of $\sigma(\tilde{k}) - 1$ becomes decisive for in what direction wL/Y will move. Piketty interprets the econometric literature as favoring $\sigma(\tilde{k}) > 1$, which means there should be downward pressure on wL/Y. This particular source behind a falling wL/Y can be questioned, however. Indeed, $\sigma(\tilde{k}) > 1$ contradicts the more general empirical view.

Immigration*

The phenomenon of migration provides another example that illustrates how the size of $\sigma(\tilde{k})$ matters. Consider a competitive economy with perfect competition, a given aggregate capital stock K, and a given technology level T (entering the production function in the labor-augmenting way as above). Suppose that due to immigration an upward shift in aggregate labor supply, L, occurs. Full employment is maintained by the needed downward real wage adjustment. Given the present model, in what direction will aggregate labor income $wL = \tilde{w}(\tilde{k})TL$

then change? The effect of the larger L is to some extent offset by a lower w brought about by the lower effective capital-labor ratio. Indeed, in view of $d\tilde{w}/d\tilde{k} = -\tilde{k}f''(\tilde{k}) > 0$, we have $\tilde{k} \downarrow$ implies $w \downarrow$ for fixed T. So we cannot apriori sign the change in wL. The following relationship can be shown (Exercise ??), however:

$$\frac{\partial(wL)}{\partial L} = (1 - \frac{\alpha(\tilde{k})}{\sigma(\tilde{k})})w \gtreqless 0 \text{ for } \sigma(\tilde{k}) \gtreqless \alpha(\tilde{k}), \tag{4.32}$$

respectively, where $a(\tilde{k}) \equiv \tilde{k}f'(\tilde{k})/f(\tilde{k})$ is the output elasticity w.r.t. capital which under perfect competition equals the gross capital income share. It follows that the larger L will not be fully offset by the lower w as long as the elasticity of factor substitution, $\sigma(\tilde{k})$, exceeds the gross capital income share, $\alpha(\tilde{k})$. This condition seems confirmed by most of the empirical evidence, see next section.

The next section describes the concept of elasticity of factor substitution at a more general level. The subsequent section introduces the special case known as the CES production function.

4.5 The elasticity of factor substitution

We shall here discuss the concept of elasticity of factor substitution at a more general level. Fig. 4.6 depicts an isoquant, $F(K, L) = \bar{Y}$, for a given neoclassical production function, $F(K, L)$, which need not have CRS. Let MRS denote the marginal rate of substitution of K for L, i.e.,

$$MRS = -\frac{dK}{dL}_{|Y=\bar{Y}} = F_L(K, L)/F_K(K, L).$$

MRS thus measures how much extra of K is needed to compensate for a reduction in L by one unit?[11] At a given point (K, L) on the isoquant curve, MRS is given by the absolute value of the slope of the tangent to the isoquant at that point. This tangent coincides with that isocost line which, given the factor prices, has minimal intercept with the vertical axis while at the same time touching the isoquant. In view of $F(\cdot)$ being neoclassical, the isoquants are by definition strictly convex to the origin. Consequently, MRS is rising along the curve when L decreases and thereby K increases. Conversely, we can let MRS be the independent variable and consider the corresponding point on the indifference curve, and thereby the ratio K/L, as a function of MRS. If we let MRS rise along the given isoquant, the corresponding value of the ratio K/L will also rise.

[11] When there is no risk of confusion as to what is up and what is down, we use MRS as a shorthand for the more precise notation MRS_{KL}.

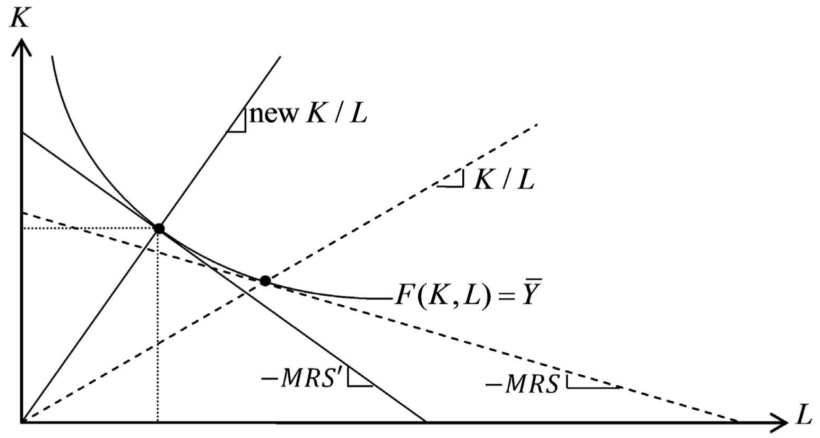


Figure 4.6: Substitution of capital for labor as the marginal rate of substitution increases from MRS to MRS'.

The *elasticity of substitution* between capital and labor, denoted $\hat{\sigma}(K, L)$, measures how sensitive K/L is vis-a-vis a rise in MRS. More precisely, $\hat{\sigma}(K, L)$ is defined as the elasticity of the ratio K/L with respect to MRS when moving along a given isoquant, evaluated at the point (K, L). Thus,

$$\hat{\sigma}(K,L) \equiv E\ell_{MRS}K/L = \frac{MRS}{K/L}\frac{d(K/L)}{dMRS}_{|Y=\bar{Y}} \approx \frac{\frac{\Delta(K/L)}{K/L}}{\frac{\Delta MRS}{MRS}}_{|Y=\bar{Y}}. \tag{4.33}$$

Although the elasticity of factor substitution is a characteristic of the technology as such and is here defined without reference to markets and factor prices, it helps the intuition to refer to factor prices. At a cost-minimizing point, MRS equals the factor price ratio $w/\hat{r}$. Thus, the *elasticity of factor substitution* will under cost minimization coincide with *the percentage increase in the ratio of the cost-minimizing factor ratio induced by a one percentage increase in the inverse factor price ratio, holding the output level unchanged.*[12] The elasticity of factor substitution is thus a positive number and reflects how sensitive the capital-labor ratio K/L is under cost minimization to a one percentage increase in the factor price ratio $w/\hat{r}$ for a given output level. The less curvature the isoquant has, the greater is the elasticity of factor substitution. In an analogue way, in consumer theory one considers the elasticity of substitution between two consumption goods or between consumption today and consumption tomorrow, cf. Chapter 3. In that context the role of the given isoquant is taken over by an indifference curve. That

[12] This characterization is equivalent to interpreting the elasticity of substitution as the percentage *decrease* in the factor ratio (when moving along a given isoquant) induced by a one-percentage *increase* in the *corresponding* factor price ratio.

is also the case when we consider the intertemporal elasticity of substitution in labor supply, cf. the next chapter.

Calculating the elasticity of substitution between K and L at the point (K, L), we get

$$\hat{\sigma}(K, L) = -\frac{F_K F_L (F_K K + F_L L)}{KL\left[(F_L)^2 F_{KK} - 2F_K F_L F_{KL} + (F_K)^2 F_{LL}\right]}, \tag{4.34}$$

where all the derivatives are evaluated at the point (K, L). When $F(K, L)$ has CRS, the formula (4.34) simplifies to

$$\hat{\sigma}(K, L) = \frac{F_K(K, L) F_L(K, L)}{F_{KL}(K, L) F(K, L)} = -\frac{f'(k)\left(f(k) - f'(k)k\right)}{f''(k) k f(k)} \equiv \sigma(k), \tag{4.35}$$

where $k \equiv K/L$.[13] We see that under CRS, the elasticity of substitution depends only on the capital-labor ratio k, not on the output level.

There is an alternative way of interpreting the substitution elasticity formula (4.33). This is based on the fact that any elasticity of a function $y = \varphi(x)$ can be written as a "double-log derivative": $\mathrm{E}\ell_x y \equiv (x/y)dy/dx = d\ln y/d\ln x$.[14] So, we can rewrite (4.33) as $\hat{\sigma}(K, L) = d\ln(K/L)/d\ln MRS$, which is a simple derivative when the data for K/L and MRS are given in logs.

We will now consider the case where the elasticity of substitution is independent also of the capital-labor ratio.

4.6 The CES production function

It can be shown[15] that if a neoclassical production function with CRS has a constant elasticity of factor substitution different from one, it must be of the form

$$Y = A\left[\alpha K^\beta + (1-\alpha)L^\beta\right]^{\frac{1}{\beta}}, \tag{4.36}$$

where A, α, and β are parameters satisfying $A > 0$, $0 < \alpha < 1$, and $\beta < 1$, $\beta \neq 0$. This function has been used intensively in empirical studies and is called a *CES production function* (CES for Constant Elasticity of Substitution). For a given choice of measurement units, the parameter A reflects efficiency (or what

[13] The formulas (4.34) and (4.35) are derived in Appendix D.

[14] To see this, let $X \equiv \ln x$ and $Y \equiv \ln y$. Then, by the chain rule,

$$\frac{d\ln y}{d\ln x} = \frac{dY}{dX} = \frac{dY}{dy}\frac{dy}{dx}\frac{dx}{dX} = \frac{1}{y}\frac{dy}{dx}e^X = \frac{x}{y}\frac{dy}{dx} = \mathrm{E}\ell_x y,$$

where the third equal sign comes from the fact that $x = e^{\ln x}$ so that $X \equiv \ln x \Rightarrow x = e^X \Rightarrow dx/dX = e^X$.

[15] See, e.g., Arrow et al. (1961).

is known as *total factor productivity*) and is thus called the *efficiency parameter.* The parameters α and β are called the *distribution parameter* and the *substitution parameter,* respectively. The restriction $\beta < 1$ ensures that the isoquants are strictly convex to the origin. Note that if $\beta < 0$, the right-hand side of (4.36) is not defined when either K or L (or both) equal 0. We can circumvent this problem by extending the domain of the CES function and assign the function value 0 to these points when $\beta < 0$. Continuity is maintained in the extended domain (see Appendix E).

By taking partial derivatives in (4.36) and substituting back we get

$$\frac{\partial Y}{\partial K} = \alpha A^{\beta}\left(\frac{Y}{K}\right)^{1-\beta} \quad \text{and} \quad \frac{\partial Y}{\partial L} = (1-\alpha)A^{\beta}\left(\frac{Y}{L}\right)^{1-\beta}, \tag{4.37}$$

where $Y/K = A\left[\alpha + (1-\alpha)k^{-\beta}\right]^{\frac{1}{\beta}}$ and $Y/L = A\left[\alpha k^{\beta} + 1 - \alpha\right]^{\frac{1}{\beta}}$.[16] The marginal rate of substitution of K for L therefore is

$$MRS = \frac{\partial Y/\partial L}{\partial Y/\partial K} = \frac{1-\alpha}{\alpha}k^{1-\beta} > 0.$$

Consequently,

$$\frac{dMRS}{dk} = \frac{1-\alpha}{\alpha}(1-\beta)k^{-\beta},$$

where the inverse of the right-hand side is the value of $dk/dMRS$. Substituting these expressions into (4.33) gives

$$\hat{\sigma}(K,L)\ (=\sigma(k)) = \frac{1}{1-\beta} \equiv \sigma, \tag{4.38}$$

confirming the constancy of the elasticity of substitution. Since $\beta < 1$, $\sigma > 0$ always. A higher substitution parameter, β, results in a higher elasticity of factor substitution, σ. And $\sigma \lesseqgtr 1$ for $\beta \lesseqgtr 0$, respectively.

Since $\beta = 0$ is not allowed in (4.36), at first sight we cannot get $\sigma = 1$ from this formula. Yet, $\sigma = 1$ can be introduced as the *limiting* case of (4.36) when $\beta \rightarrow 0$, which turns out to be the Cobb-Douglas function. Indeed, one can show[17] that, for fixed K and L,

$$A\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]^{\frac{1}{\beta}} \rightarrow AK^{\alpha}L^{1-\alpha}, \text{ for } \beta \rightarrow 0 \text{ (so that } \sigma \rightarrow 1).$$

By a similar procedure as above we find that a Cobb-Douglas function always has elasticity of substitution equal to 1; this is exactly the value taken by σ in

[16] The calculations are slightly simplified if we start from the transformation $Y^{\beta} = A^{\beta}\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]$.

[17] Proofs of this and the further claims below are in Appendix E.

(4.38) when $\beta = 0$. In addition, the Cobb-Douglas function is the *only* production function that has unit elasticity of substitution whatever the capital-labor ratio.

Another interesting limiting case of the CES function appears when, for fixed K and L, we let $\beta \to -\infty$ so that $\sigma \to 0$. We get

$$A\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]^{\frac{1}{\beta}} \to A \min(K, L), \text{ for } \beta \to -\infty \text{ (so that } \sigma \to 0). \quad (4.39)$$

So in this case the CES function approaches a Leontief production function, the isoquants of which form a right angle, cf. Fig. 4.7. In the limit there is *no* possibility of substitution between capital and labor. In accordance with this the elasticity of substitution calculated from (4.38) approaches zero when β goes to $-\infty$.

Finally, let us consider the "opposite" transition. For fixed K and L we let the substitution parameter rise towards 1 and get

$$A\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]^{\frac{1}{\beta}} \to A\left[\alpha K + (1-\alpha)L\right], \text{ for } \beta \to 1 \text{ (so that } \sigma \to \infty).$$

Here the elasticity of substitution calculated from (4.38) tends to ∞ and the isoquants tend to straight lines with slope $-(1-\alpha)/\alpha$. In the limit, the production function thus becomes linear and capital and labor become *perfect substitutes*.

Fig. 4.7 depicts isoquants for alternative CES production functions and their limiting cases. In the Cobb-Douglas case, $\sigma = 1$, the horizontal and vertical asymptotes of the isoquant coincide with the coordinate axes. When $\sigma < 1$, the horizontal and vertical asymptotes of the isoquant belong to the interior of the positive quadrant. This implies that both capital and labor are essential inputs. When $\sigma > 1$, the isoquant terminates in points *on* the coordinate axes. Then neither capital, nor labor are essential inputs. Empirically there is not complete agreement about the "normal" size of the elasticity of factor substitution for industrialized economies. The elasticity also differs across the production sectors. A thorough econometric study (Antràs, 2004) of U.S. data indicate the aggregate elasticity of substitution to be in the interval $(0.5, 1.0)$. The survey by Chirinko (2008) concludes with the interval $(0.4, 0.6)$. Starting from micro data, a recent study by Oberfield and Raval (2014) finds that the elasticity of factor substitution for the US manufacturing sector as a whole has been stable since 1970 at about 0.7.

The CES production function in intensive form

Dividing through by L on both sides of (4.36), we obtain the CES production function in intensive form,

$$y \equiv \frac{Y}{L} = A(\alpha k^{\beta} + 1 - \alpha)^{\frac{1}{\beta}}, \quad (4.40)$$

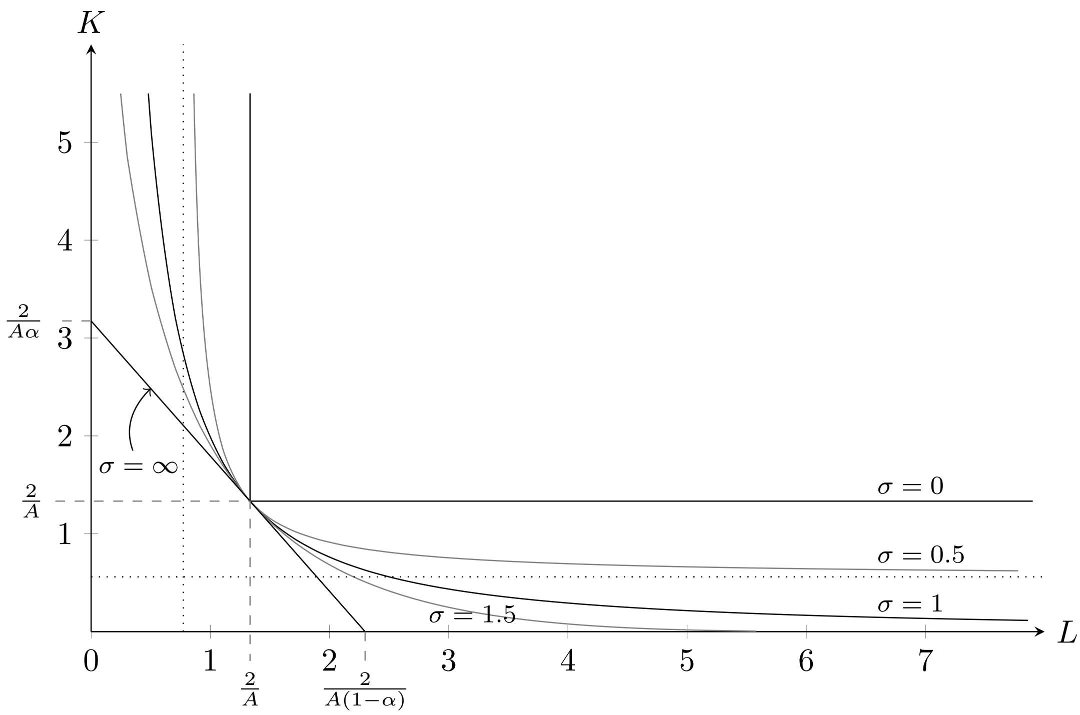


Figure 4.7: Isoquants for the CES function for alternative values of σ ($A = 1.5$, $\bar{Y} = 2$, and $\alpha = 0.42$).

where $k \equiv K/L$. The marginal productivity of capital can be written

$$MPK = \frac{dy}{dk} = \alpha A \left[\alpha + (1-\alpha)k^{-\beta}\right]^{\frac{1-\beta}{\beta}} = \alpha A^{\beta}\left(\frac{y}{k}\right)^{1-\beta},$$

which of course equals $\partial Y/\partial K$ in (4.37). We see that the CES function violates either the lower or the upper Inada condition for MPK, depending on the sign of β. Indeed, when $\beta < 0$ (i.e., $\sigma < 1$), then for $k \to 0$ both y/k and dy/dk approach an upper bound equal to $A\alpha^{1/\beta} < \infty$, thus violating the *lower* Inada condition for MPK (see the left-hand panel of Fig. 4.8). It is also noteworthy that in this case, for $k \to \infty$, y approaches an upper bound equal to $A(1-\alpha)^{1/\beta} < \infty$. These features reflect the low degree of substitutability when $\beta < 0$.

When instead $\beta > 0$, there is a high degree of substitutability ($\sigma > 1$). Then, for $k \to \infty$ both y/k and $dy/dk \to A\alpha^{1/\beta} > 0$, thus violating the *upper* Inada condition for MPK (see right-hand panel of Fig. 4.8). It is also noteworthy that for $k \to 0$, y approaches a positive lower bound equal to $A(1-\alpha)^{1/\beta} > 0$. Thus, when $\sigma > 1$, capital is not essential. At the same time $dy/dk \to \infty$ for $k \to 0$ (so the lower Inada condition for the marginal productivity of capital holds). Details are in Appendix E.

The marginal productivity of labor is

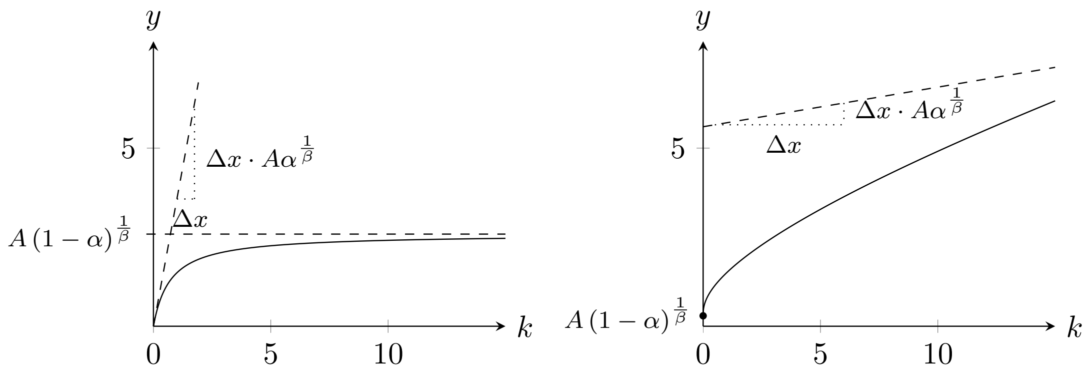
$$MPL = \frac{\partial Y}{\partial L} = (1-\alpha)A^{\beta}y^{1-\beta} = (1-\alpha)A(\alpha k^{\beta}+1-\alpha)^{(1-\beta)/\beta} \equiv w(k),$$

from (4.37). Under perfect competition, the equilibrium labor income share is thus

$$\frac{wL}{Y} = \frac{(1-\alpha)(\alpha k^{\beta}+1-\alpha)^{1/\beta-1}}{(\alpha k^{\beta}+1-\alpha)^{\frac{1}{\beta}}} = \frac{1-\alpha}{\alpha k^{\beta}+1-\alpha}.$$

Since (4.36) is symmetric in K and L, we get a series of symmetric results by considering output per unit of capital as $x \equiv Y/K = A\left[\alpha + (1-\alpha)(L/K)^{\beta}\right]^{1/\beta}$. In total, therefore, when there is low substitutability ($\sigma < 1$), for fixed input of either of the production factors, there is an upper bound for how much an unlimited input of the other production factor can increase output. And when there is high substitutability ($\sigma > 1$), there is no such bound and an unlimited input of either production factor take output to infinity.

The Cobb-Douglas case, i.e., the limiting case for $\beta \to 0$, constitutes in several respects an intermediate case in that *all* four Inada conditions are satisfied and we have $y \to 0$ for $k \to 0$, and $y \to \infty$ for $k \to \infty$.



a) The case of $\sigma < 1$. a) The case of $\sigma > 1$.

Figure 4.8: The CES production function in intensive form, $\sigma = 1/(1-\beta)$, $\beta < 1$.

Returning to the general CES function, in case of Harrod-neutral technological progress, (4.36) and (4.40) are replaced by

$$Y = A\left[\alpha K^{\beta} + (1-\alpha)(TL)^{\beta}\right]^{\frac{1}{\beta}}$$

and

$$\tilde{y} \equiv \frac{Y}{TL} = A(\alpha \tilde{k}^{\beta} + 1 - \alpha)^{\frac{1}{\beta}},$$

respectively, where T is the technology level, and $\tilde{k} \equiv K/(TL)$.

Generalizations*

The CES production functions considered above have CRS. By adding an elasticity of scale parameter, γ, we get the generalized form (the case without technological progress):

$$Y = A\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]^{\frac{\gamma}{\beta}}, \qquad \gamma > 0. \tag{4.41}$$

In this form the CES function is homogeneous of degree γ. For $0 < \gamma < 1$, there are DRS, for $\gamma = 1$ CRS, and for $\gamma > 1$ IRS. If $\gamma \neq 1$, it may be convenient to consider $Q \equiv Y^{1/\gamma} = A^{1/\gamma}\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]^{1/\beta}$ and $q \equiv Q/L = A^{1/\gamma}(\alpha k^{\beta} + 1 - \alpha)^{1/\beta}$.

The elasticity of substitution between K and L is $\sigma = 1/(1-\beta)$ whatever the value of γ. So including the limiting cases as well as non-constant returns to scale in the "family" of production functions with constant elasticity of substitution, we have the simple classification displayed in Table 4.1.

Table 4.1 The family of production functions with constant elasticity of substitution.

$\sigma = 0$	$0 < \sigma < 1$	$\sigma = 1$	$\sigma > 1$
Leontief	CES	Cobb-Douglas	CES

Note that only for $\gamma \leq 1$ is (4.41) a *neoclassical* production function. This is because, when $\gamma > 1$, the conditions $F_{KK} < 0$ and $F_{NN} < 0$ do not hold everywhere.

We may generalize further by assuming there are n inputs, in the amounts $X_1, X_2, ..., X_n$. Then the CES production function takes the form

$$Y = A\left[\alpha_1 {X_1}^{\beta} + \alpha_2 {X_2}^{\beta} + ...\alpha_n {X_n}^{\beta}\right]^{\frac{\gamma}{\beta}}, \ \alpha_i > 0 \text{ for all } i, \sum_i \alpha_i = 1, \gamma > 0. \tag{4.42}$$

In analogy with (4.33), for an n-factor production function the *partial elasticity of substitution* between factor i and factor j is defined as

$$\sigma_{ij} = \frac{MRS_{ij}}{X_i/X_j} \frac{d(X_i/X_j)}{dMRS_{ij}}_{|Y=\bar{Y}},$$

where it is understood that not only the output level but also all X_k, $k \neq i, j$, are kept constant. Note that $\sigma_{ji} = \sigma_{ij}$. In the CES case considered in (4.42), all the partial elasticities of substitution take the same value, $1/(1-\beta)$.

4.7 Concluding remarks

(incomplete)

When speaking of "sustained growth" in variables like K, Y, and C, we do not mean growth in a narrow physical sense. Given limited natural resources (matter and energy), sustained exponential growth in a physical sense is not possible. But sustained exponential growth in terms of economic value is not ruled out. We should for instance understand K broadly as "produced means of production" of *rising quality* and *falling material intensity* (think of the rising efficiency of microprocessors). Similarly, C must be seen as a composite of consumer goods and services with declining material intensity over time. This accords with the empirical fact that as income rises, the share of consumption expenditures devoted to agricultural and industrial products declines and the share devoted to services, hobbies, and amusement increases. Although "economic development" is a more appropriate term (suggesting qualitative and structural change), we will in this book retain standard terminology and speak of "economic growth".

A further remark about terminology. In the branch of economics called economic growth theory, the term "economic growth" is used primarily for growth of *productivity* and *income per capita* rather than just growth of GDP.

4.8 Literature notes

1. We introduced the assumption that at the macroeconomic level the "direction" of technological progress tends to be Harrod-neutral. Otherwise the model will not be consistent with Kaldor's stylized facts. The Harrod-neutrality of the "direction" of technological progress is in the present model just an exogenous feature. This raises the question whether there are *mechanisms* tending to generate Harrod-neutrality. Fortunately new growth theory provides clues as to the sources of the speed as well as the direction of technological change. A facet of this theory is that the direction of technological change is linked to the same economic forces as the speed, namely profit incentives. See Acemoglu (2003) and Jones (2006).

2. Recent literature discussing Kaldor's "stylized facts" includes Rognlie (2015), Gollin (2002), Elsby et al. (2013), and Karabarbounis and Neiman (2014). The latter three references conclude with serious scepticism. Attfield and Temple (2010) and others, however, find support for the Kaldor "facts" considering the US and UK based on time-series econometrics. This means an observed evolution roughly obeying balanced growth in terms of *aggregate* variables. *Structural change* is not ruled out by this. *A* changing sectorial composition of the economy is under certain conditions compatible with balanced growth (in a generalized

sense) at the aggregate level, cf. the "Kuznets facts" (see Kongsamut et al., 2001, and Acemoglu, 2009).

3. In Section 4.2 we claimed that from an empirical point of view, the adjustment towards a steady state can be from above as well as from below. Indeed, Cho and Graham (1996), based on the Penn World Table, find that "on average, countries with a lower income per adult are above their steady-state positions, while countries with a higher income are below their steady-state positions".

4. As to the assessment of whether dynamic inefficiency is – or at least has been – part of reality, in addition to Abel et al. (1989) other useful sources include Ball et al. (1998), Blanchard and Weil (2001), and Barbie, Hagedorn, and Kaul (2004). A survey is given in Weil (2008).

5. In the Diamond OLG model as well as in many other models, a *steady state* and a *balanced growth path* imply each other. Indeed, they are two sides of the same process. There *exist* cases, however, where this equivalence does not hold (some open economy models and some models with *embodied* technological change, see Groth et al., 2010). Therefore, it is recommendable always to maintain a terminological distinction between the two concepts.

6. On the declining material intensity of consumer goods and services as technology develops, see Fagnart and Germain (2011).

From here incomplete:

The term "Great Ratios" of the economy was coined by Klein and Kosubud (1961).

La Grandville (1989): normalization of the CES function. La Grandville (2009) contains a lot about analytical aspects linked to the CES production function and the concept of elasticity of factor substitution.

Piketty (2014), Zucman ( ).

According to Summers (2014), Piketty's interpretation of data relevant for estimation of the elasticity of factor substitution relies on conflating gross and net returns to capital. Krusell and Smith (2015) and Ronglie (2015).

Demange and Laroque (1999, 2000) extend Diamond's OLG model to uncertain environments.

For expositions in depth of OLG modeling and dynamics in discrete time, see Azariadis (1993), de la Croix and Michel (2002), and Bewley (2007).

Dynamic inefficiency, see also Burmeister (1980).

Uzawa's theorem: Uzawa (1961), Schlicht (2006).

The way the intuition behind the Uzawa theorem was presented in Section 4.1 draws upon Jones and Scrimgeour (2008).

For more general and flexible production functions applied in econometric work, see, e.g., Nadiri (1982).

Other aspects of life cycle behavior: education. OLG where people live three periods. Also Eggertsson and Mehrotra (2015).

4.9 Appendix

A. Growth and interest arithmetic in discrete time

Let $t = 0, \pm 1, \pm 2, \ldots$, and consider the variables z_t, x_t, and y_t, assumed positive for all t. Define $\Delta z_t = z_t - z_{t-1}$ and Δx_t and Δy_t similarly. These Δ's need not be positive. The *growth rate* of x_t from period $t-1$ to period t is defined as the relative rate of increase in x, i.e., $\Delta x_t / x_{t-1} \equiv x_t / x_{t-1}$. And the *growth factor* for x_t from period $t-1$ to period t is defined as $1 + x_t/x_{t-1}$.

As we are here interested not in the time evolution of growth rates, we simplify notation by suppressing the t's. So we write the growth rate of x as $g_x \equiv \Delta x / x_{-1}$ and similarly for y and z.

PRODUCT RULE If $z = xy$, then $1 + g_z = (1 + g_x)(1 + g_y)$ and $g_z \approx g_x + g_y$, when g_x and g_y are "small".

Proof. By definition, $z = xy$, which implies $z_{-1} + \Delta z = (x_{-1} + \Delta x)(y_{-1} + \Delta y)$. Dividing by $z_{-1} = x_{-1}y_{-1}$ gives $1 + \Delta z/z_{-1} = (1 + \Delta x/x_{-1})(1 + \Delta y/y_{-1})$ as claimed. By carrying out the multiplication on the right-hand side of this equation, we get $1 + \Delta z/z_{-1} = 1 + \Delta x/x_{-1} + \Delta y/y_{-1} + (\Delta x/x_{-1})(\Delta y/y_{-1}) \approx 1 + \Delta x/x_{-1} + \Delta y/y_1$ when $\Delta x/x_{-1}$ and $\Delta y/y_{-1}$ are "small". Subtracting 1 on both sides gives the stated approximation. $\square$

So the product of two positive variables will grow at a rate approximately equal to the sum of the growth rates of the two variables.

QUOTIENT RULE If $z = x/y$, then $1 + g_z = (1 + g_x)/(1 + g_y)$ and $g_z \approx g_x - g_y$, when g_x and g_y are "small".

Proof. By interchanging z and x in Product Rule and rearranging, we get $1 + \Delta z/z_{-1} = \frac{1+\Delta x/x_{-1}}{1+\Delta y/y_{-1}}$, as stated in the first part of the claim. Subtracting 1 on both sides gives $\Delta z/z_{-1} = \frac{\Delta x/x_{-1} - \Delta y/y_{-1}}{1+\Delta y/y_{-1}} \approx \Delta x/x_{-1} - \Delta y/y_{-1}$, when $\Delta x/x_{-1}$ and $\Delta y/y_{-1}$ are "small". This proves the stated approximation. $\square$

So the ratio between two positive variables will grow at a rate approximately equal to the excess of the growth rate of the numerator over that of the denominator. An implication of the first part of Claim 2 is: the ratio between two positive variables is constant if and only if the variables have the same growth rate (not necessarily constant or positive).

POWER FUNCTION RULE If $z = x^\alpha$, then $1 + g_z = (1 + g_x)^\alpha$.

Proof. $1 + g_z \equiv z/z_{-1} = (x/x_{-1})^\alpha \equiv (1 + g_x)^\alpha$. $\square$

Given a time series $x_0, x_1, ..., x_n$, by the *average growth rate* per period, or more precisely, the *average compound growth rate,* is meant a g which satisfies $x_n = x_0(1+g)^n$. The solution for g is $g = (x_n/x_0)^{1/n} - 1$.

Finally, the following approximation may be useful (for intuition) if used with caution:

THE GROWTH FACTOR With n denoting a positive integer above 1 but "not too large", the growth factor $(1+g)^n$ can be approximated by $1+ng$ when g is "small". For $g \neq 0$, the approximation error is larger the larger is n.

Proof. (i) We prove the claim by induction. Suppose the claim holds for a fixed $n \geq 2$, i.e., $(1+g)^n \approx 1+ng$ for g "small". Then $(1+g)^{n+1} = (1+g)^n(1+g)$ $\approx (1+ng)(1+g) = 1+ng+g+ng^2 \approx 1+(n+1)g$ since g "small" implies g^2 "very small" and therefore ng^2 "small" if n is not "too" large. So the claim holds also for $n+1$. Since $(1+g)^2 = 1+2g+g^2 \approx 1+2g$, for g "small", the claim does indeed hold for $n = 2$. □

THE EFFECTIVE ANNUAL RATE OF INTEREST Suppose interest on a loan is charged n times a year at the rate r/n per year. Then the *effective annual interest rate* is $(1+r/n)^n - 1$.

B. Proof of the sufficiency part of Uzawa's theorem

For convenience we restate the full theorem here:

PROPOSITION 2. Let $P \equiv \{(K_t, Y_t, C_t)\}_{t=0}^{\infty}$ be a path along which Y_t, K_t, C_t, and $S_t \equiv Y_t - C_t$ are positive for all $t = 0, 1, 2, \ldots$, and satisfy the dynamic resource constraint for a closed economy, (4.3), given the production function (4.5) and the labor force (4.6). Then:
(i) A *necessary* condition for the path P to be a BGP is that along P it holds that

$$Y_t = \tilde{F}(K_t, T_t L_t, 0), \tag{*}$$

where $T_t = T_0(1+g)^t$ with $T_0 = B$ and $1+g \equiv (1+g_Y)/(1+n) > 1$, g_Y being the constant growth rate of output along the BGP.
(ii) Assume $(1+g)(1+n) > 1-\delta$. Then, for any $g \geq 0$ such that there is a $q > (1+g)(1+n) - (1-\delta)$ with the property that the production function $\tilde{F}$ in (4.5) allows an output-capital ratio equal to q at $t = 0$ (i.e., $\tilde{F}(1, \tilde{k}^{-1}, 0) = q$ for some real number $\tilde{k} > 0$), a *sufficient* condition for $\tilde{F}$ to be compatible with a BGP with output-capital ratio equal to q is that $\tilde{F}$ can be written as in (4.7) with $T_t = B(1+g)^t$.

Proof (i) See Section 4.1. (ii) Suppose (*) holds with $T_t = B(1+g)^t$. Let $g \geq 0$ be given such that there is a $q > (1+g)(1+n) - (1-\delta) > 0$ with the property that

$$\tilde{F}(1, \tilde{k}^{-1}, 0) = q \tag{**}$$

for some constant $\tilde{k} > 0$. Our strategy is to prove the claim by constructing a path $P = (Y_t, K_t, C_t)_{t=0}^{\infty}$ which satisfies it. We let P be such that the saving-income ratio is a constant $\hat{s} \equiv [(1+g)(1+n) - (1-\delta)]/q \in (0,1)$, i.e., $Y_t - C_t \equiv S_t = \hat{s}Y_t$ for all $t = 0, 1, 2, \ldots$. Inserting this, together with $Y_t = f(\tilde{k}_t)T_tL_t$, where $f(\tilde{k}_t) \equiv \tilde{F}(\tilde{k}_t, 1, 0)$ and $\tilde{k}_t \equiv K_t/(T_tL_t)$, into (4.3), rearranging gives the Solow equation (4.4), which we may rewrite as

$$\tilde{k}_{t+1} - \tilde{k}_t = \frac{\hat{s}f(\tilde{k}_t) - [(1+g)(1+n) - (1-\delta)]\,\tilde{k}_t}{(1+g)(1+n)}.$$

We see that $\tilde{k}_t$ is constant if and only if $\tilde{k}_t$ satisfies the equation $f(\tilde{k}_t)/\tilde{k}_t = [(1+g)(1+n) - (1-\delta)]/\hat{s} \equiv q$. By (**) and the definition of f, the required value of $\tilde{k}_t$ is $\tilde{k}$, which is thus the steady state for the constructed Solow model. Letting K_0 satisfy $K_0 = \tilde{k}BL_0$, where $B = T_0$, we thus have $\tilde{k}_0 = K_0/(T_0L_0) = \tilde{k}$. So that the initial value of $\tilde{k}_t$ equals the steady-state value. It follows that $\tilde{k}_t = \tilde{k}$ for all $t = 0, 1, 2, \ldots$, and so $Y_t/K_t = f(\tilde{k}_t)/\tilde{k}_t = f(\tilde{k})/\tilde{k} = q$ for all $t \geq 0$. In addition, $C_t = (1-\hat{s})Y_t$, so that C_t/Y_t is constant along the path P. As both Y/K and C/Y are thus constant along the path P, by (ii) of Proposition 1 follows that P is a BGP. $\square$

It is noteworthy that the proof of the sufficiency part of the theorem is *constructive*. It provides a method for constructing a BGP with a given technology growth rate and a given output-capital ratio.

C. Homothetic utility functions

Generalities A set C in R^n is called a *cone* if $x \in C$ and $\lambda > 0$ implies $\lambda x \in C$. A function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ is *homothetic* in the cone C if for all $\mathbf{x}, \mathbf{y} \in C$ and all $\lambda > 0$, $f(\mathbf{x}) = f(\mathbf{y})$ implies $f(\lambda\mathbf{x}) = f(\lambda\mathbf{y})$.

Consider the continuous utility function $U(x_1, x_2)$, defined in R_+^2. Suppose U is *homothetic* and that the consumption bundles (x_1, x_2) and (y_1, y_2) belong to the same indifference curve, i.e., $U(x_1, x_2) = U(y_1, y_2)$. Then for any $\lambda > 0$ we have $U(\lambda x_1, \lambda x_2) = U(\lambda y_1, \lambda y_2)$, meaning that also the bundles $(\lambda x_1, \lambda x_2)$ and $(\lambda y_1, \lambda y_2)$ belong to the same indifference curve.

CLAIM C1. Let $U(x_1, x_2)$ be a continuous two-good utility function, increasing in each of its arguments (as is the life time utility function of the Diamond model). Then: U is homothetic if and only if U can be written $U(x_1, x_2) \equiv F(f(x_1, x_2))$ where the function f is homogeneous of degree one and F is an increasing function.

Proof. The "if" part is easily shown. Indeed, if $U(x_1, x_2) = U(y_1, y_2)$, then $F(f(x_1, x_2)) = F(f(y_1, y_2))$. Since F is increasing, this implies $f(x_1, x_2) = f(y_1, y_2)$. Because f is homogeneous of degree one, if $\lambda > 0$, then $f(\lambda x_1, \lambda x_2) = \lambda f(x_1, x_2)$ and $f(\lambda y_1, \lambda y_2) = \lambda f(y_1, y_2)$ so that $U(\lambda x_1, \lambda x_2) = F(f(\lambda x_1, \lambda x_2)) = F(f(\lambda y_1, \lambda y_2)) = U(\lambda y_1, \lambda y_2)$, which shows that U is homothetic. As to the "only if" part, see Sydsæter et al. (2002). □

Using differentiability of our homothetic time utility function $U(x_1, x_2) \equiv F(f(x_1, x_2))$, we find the marginal rate of substitution of good 2 for good 1 to be

$$MRS = \frac{dx_2}{dx_1}_{|U=\bar{U}} = \frac{\partial U/\partial x_1}{\partial U/\partial x_2} = \frac{F' f_1(x_1, x_2)}{F' f_2(x_1, x_2)} = \frac{f_1(1, \frac{x_2}{x_1})}{f_2(1, \frac{x_2}{x_1})}. \tag{4.43}$$

The last equality is due to Euler's theorem saying that when f is homogeneous of degree 1, then the first-order partial derivatives of f are homogeneous of degree 0. Now, (4.43) implies that for a given MRS, in optimum reflecting a given relative price of the two goods, the same consumption ratio, x_2/x_1, will be chosen whatever the budget. For a given relative price, a rising budget (rising wealth) will change the position of the budget line, but not its slope. So MRS will not change, which implies that the chosen pair, (x_1, x_2), will move outward along a given ray in R_+^2. Indeed, as the intercepts with the axes rise proportionately with the budget (the wealth), so will x_1 and x_2.

Proof that the utility function in (4.25) is homothetic In Section 4.2 we claimed that (4.25) is a homothetic utility function. Based on Claim C1, this can be proved in the following way. There are two cases to consider. *Case 1:* $\theta > 0$, $\theta \neq 1$. We rewrite (4.25) as

$$U(c_1, c_2) = \frac{1}{1-\theta}\left[(c_1^{1-\theta} + \beta c_2^{1-\theta})^{1/(1-\theta)}\right]^{1-\theta} - \frac{1+\beta}{1-\theta},$$

where $\beta \equiv (1+\rho)^{-1}$. The function $x = g(c_1, c_2) \equiv (c_1^{1-\theta} + \beta c_2^{1-\theta})^{1/(1-\theta)}$ is homogeneous of degree one and the function $G(x) \equiv (1/(1-\theta))x^{1-\theta} - (1+\beta)/(1-\theta)$ is an increasing function, given $\theta > 0$, $\theta \neq 1$. *Case 2:* $\theta = 1$. Here we start from $U(c_1, c_2) = \ln c_1 + \beta \ln c_2$. This can be written

$$U(c_1, c_2) = (1+\beta)\ln\left[(c_1 c_2^{\beta})^{1/(1+\beta)}\right],$$

where $x = g(c_1, c_2) = (c_1 c_2^{\beta})^{1/(1+\beta)}$ is homogeneous of degree one and $G(x) \equiv (1+\beta)\ln x$ is an increasing function. □

D. General formulas for the elasticity of factor substitution

We here prove (4.34) and (4.35). Given the neoclassical production function $F(K, L)$, the slope of the isoquant $F(K, L) = \bar{Y}$ at the point $(\bar{K}, \bar{L})$ is

$$\frac{dK}{dL}_{|Y=\bar{Y}} = -MRS = -\frac{F_L(\bar{K}, \bar{L})}{F_K(\bar{K}, \bar{L})}. \tag{4.44}$$

We consider this slope as a function of the value of $k \equiv K/L$ as we move along the isoquant. The derivative of this function is

$$\begin{aligned} -\frac{dMRS}{dk}_{|Y=\bar{Y}} &= -\frac{dMRS}{dL}_{|Y=\bar{Y}} \frac{dL}{dk}_{|Y=\bar{Y}} \\ &= -\frac{(F_L)^2 F_{KK} - 2F_K F_L F_{KL} + (F_K)^2 F_{LL}}{F_K^3} \frac{dL}{dk}_{|Y=\bar{Y}} \end{aligned} \tag{4.45}$$

by (2.53) of Chapter 2. In view of $L \equiv K/k$ we have

$$\frac{dL}{dk}_{|Y=\bar{Y}} = \frac{k\frac{dK}{dk}_{|Y=\bar{Y}} - K}{k^2} = \frac{k\frac{dK}{dL}_{|Y=\bar{Y}} \frac{dL}{dk}_{|Y=\bar{Y}} - K}{k^2} = \frac{-kMRS\frac{dL}{dk}_{|Y=\bar{Y}} - K}{k^2}.$$

From this we find

$$\frac{dL}{dk}_{|Y=\bar{Y}} = -\frac{K}{(k + MRS)k},$$

to be substituted into (4.45). Finally, we substitute the inverse of (4.45) together with (4.44) into the definition of the elasticity of factor substitution:

$$\begin{aligned} \sigma(K, L) &\equiv \frac{MRS}{k} \frac{dk}{dMRS}_{|Y=\bar{Y}} \\ &= -\frac{F_L/F_K}{k} \frac{(k + F_L/F_K)k}{K} \frac{F_K^3}{[(F_L)^2 F_{KK} - 2F_K F_L F_{KL} + (F_K)^2 F_{LL}]} \\ &= -\frac{F_K F_L (F_K K + F_L L)}{KL\left[(F_L)^2 F_{KK} - 2F_K F_L F_{KL} + (F_K)^2 F_{LL}\right]}, \end{aligned}$$

which is the same as (4.34).

Under CRS, this reduces to

$$\begin{aligned} \sigma(K, L) &= -\frac{F_K F_L F(K, L)}{KL\left[(F_L)^2 F_{KK} - 2F_K F_L F_{KL} + (F_K)^2 F_{LL}\right]} \quad \text{(from (2.54) with } h = 1) \\ &= -\frac{F_K F_L F(K, L)}{KLF_{KL}\left[-(F_L)^2 L/K - 2F_K F_L - (F_K)^2 K/L\right]} \quad \text{(from (2.60))} \\ &= \frac{F_K F_L F(K, L)}{F_{KL}(F_L L + F_K K)^2} = \frac{F_K F_L}{F_{KL} F(K, L)}, \quad \text{(using (2.54) with } h = 1) \end{aligned}$$

which proves the first part of (4.35). The second part is an implication of rewriting the formula in terms of the production in intensive form.

E. Properties of the CES production function

The generalized CES production function is

$$Y = A\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]^{\frac{\gamma}{\beta}}, \tag{4.46}$$

where A, α, and β are parameters satisfying $A > 0$, $0 < \alpha < 1$, and $\beta < 1$, $\beta \neq 0, \gamma > 0$. If $\gamma < 1$, there is DRS, if $\gamma = 1$, CRS, and if $\gamma > 1$, IRS. The elasticity of substitution is always $\sigma = 1/(1-\beta)$. Throughout below, k means K/L.

The limiting functional forms We claimed in the text that, for fixed $K > 0$ and $L > 0$, (4.46) implies:

$$\begin{aligned} \lim_{\beta\to 0} Y &= A(K^{\alpha}L^{1-\alpha})^{\gamma} = AL^{\gamma}k^{\alpha\gamma}, && (*) \\ \lim_{\beta\to -\infty} Y &= A\min(K^{\gamma}, L^{\gamma}) = AL^{\gamma}\min(k^{\gamma}, 1). && (**) \end{aligned}$$

Proof. Let $q \equiv Y/(AL^{\gamma})$. Then $q = (\alpha k^{\beta} + 1 - \alpha)^{\gamma/\beta}$ so that

$$\ln q = \frac{\gamma \ln(\alpha k^{\beta} + 1 - \alpha)}{\beta} \equiv \frac{m(\beta)}{\beta}, \tag{4.47}$$

where

$$m'(\beta) = \frac{\gamma\alpha k^{\beta}\ln k}{\alpha k^{\beta} + 1 - \alpha} = \frac{\gamma\alpha \ln k}{\alpha + (1-\alpha)k^{-\beta}}. \tag{4.48}$$

Hence, by L'Hôpital's rule for "0/0",

$$\lim_{\beta\to 0} \ln q = \lim_{\beta\to 0} \frac{m'(\beta)}{1} = \gamma\alpha\ln k = \ln k^{\gamma\alpha},$$

so that $\lim_{\beta\to 0} q = k^{\gamma\alpha}$. In view of $Y = AL^{\gamma}q$, $\lim_{\beta\to 0} Y = AL^{\gamma}\lim_{\beta\to 0} q = AL^{\gamma}k^{\gamma\alpha}$. This proves (*).

As to (**), note that

$$\lim_{\beta\to -\infty} k^{\beta} = \lim_{\beta\to -\infty} \frac{1}{k^{-\beta}} \to \begin{cases} 0 \text{ for } k > 1, \\ 1 \text{ for } k = 1, \\ \infty \text{ for } k < 1. \end{cases}$$

Hence, by (4.47),

$$\lim_{\beta\to -\infty} \ln q = \begin{cases} 0 \text{ for } k \geq 1, \\ \lim_{\beta\to -\infty} \frac{m'(\beta)}{1} = \gamma \ln k = \ln k^{\gamma} \text{ for } k < 1, \end{cases}$$

where the result for $k < 1$ is based on L'Hôpital's rule for "∞/-∞". Consequently,

$$\lim_{\beta \to -\infty} q = \begin{cases} 1 \text{ for } k \geq 1, \\ k^{\gamma} \text{ for } k < 1. \end{cases}$$

In view of $Y = AL^{\gamma}q$, $\lim_{\beta \to -\infty} Y = AL^{\gamma} \lim_{\beta \to -\infty} q = AL^{\gamma} \min(k^{\gamma}, 1)$. This proves (**). $\square$

Properties of the isoquants of the CES function The absolute value of the slope of an isoquant for (4.46) in the (L, K) plane is

$$MRS_{KL} = \frac{\partial Y/\partial L}{\partial Y/\partial K} = \frac{1-\alpha}{\alpha} k^{1-\beta} \to \begin{cases} 0 \text{ for } k \to 0, \\ \infty \text{ for } k \to \infty. \end{cases} \qquad (*)$$

This holds whether $\beta < 0$ or $0 < \beta < 1$.

Concerning the asymptotes and terminal points, if any, of the isoquant $Y = \bar{Y}$ we have from (4.46) $\bar{Y}^{\beta/\gamma} = A\left[\alpha K^{\beta} + (1-\alpha)L^{\beta}\right]$. Hence,

$$\begin{aligned} K &= \left(\frac{\bar{Y}^{\frac{\beta}{\gamma}}}{A\alpha} - \frac{1-\alpha}{\alpha} L^{\beta} \right)^{\frac{1}{\beta}}, \\ L &= \left(\frac{\bar{Y}^{\frac{\beta}{\gamma}}}{A(1-\alpha)} - \frac{\alpha}{1-\alpha} K^{\beta} \right)^{\frac{1}{\beta}}. \end{aligned}$$

From these two equations follows, when $\beta < 0$ (i.e., $0 < \sigma < 1$), that

$$\begin{aligned} K &\to (A\alpha)^{-\frac{1}{\beta}} \bar{Y}^{\frac{1}{\gamma}} \text{ for } L \to \infty, \\ L &\to [A(1-\alpha)]^{-\frac{1}{\beta}} \bar{Y}^{\frac{1}{\gamma}} \text{ for } K \to \infty. \end{aligned}$$

When instead $\beta > 0$ (i.e., $\sigma > 1$), the same limiting formulas obtain for $L \to 0$ and $K \to 0$, respectively.

Properties of the CES function in intensive form Given $\gamma = 1$, i.e., CRS, we have $y \equiv Y/L = A(\alpha k^{\beta} + 1 - \alpha)^{1/\beta}$ from (4.46). Then

$$\frac{dy}{dk} = A\frac{1}{\beta}(\alpha k^{\beta} + 1 - \alpha)^{\frac{1}{\beta}-1} \alpha\beta k^{\beta-1} = A\alpha \left[\alpha + (1-\alpha)k^{-\beta}\right]^{\frac{1-\beta}{\beta}}.$$

Hence, when $\beta < 0$ (i.e., $0 < \sigma < 1$),

$$\begin{aligned} y &= \frac{A}{(ak^{\beta} + 1 - \alpha)^{-1/\beta}} \to \begin{cases} 0 \text{ for } k \to 0, \\ A(1-\alpha)^{1/\beta} \text{ for } k \to \infty. \end{cases} \\ \frac{dy}{dk} &= \frac{A\alpha}{[\alpha + (1-\alpha)k^{-\beta}]^{(\beta-1)/\beta}} \to \begin{cases} A\alpha^{1/\beta} \text{ for } k \to 0, \\ 0 \text{ for } k \to \infty. \end{cases} \end{aligned}$$

If instead $\beta > 0$ (i.e., $\sigma > 1$),

$$y \rightarrow \begin{cases} A(1-\alpha)^{1/\beta} \text{ for } k \rightarrow 0, \\ \infty \text{ for } k \rightarrow \infty. \end{cases}$$

$$\frac{dy}{dk} \rightarrow \begin{cases} \infty \text{ for } k \rightarrow 0, \\ A\alpha^{1/\beta} \text{ for } k \rightarrow \infty. \end{cases}$$

The output-capital ratio is $y/k = A\left[\alpha + (1-\alpha)k^{-\beta}\right]^{\frac{1}{\beta}}$ and has the same limiting values as dy/dk, when $\beta > 0$.

Continuity at the boundary of R_+^2. When $0 < \beta < 1$, the right-hand side of (4.46) is defined and continuous also on the boundary of $\mathbb{R}_+^2$. Indeed, we get

$$Y = F(K,L) = A\left[\alpha K^\beta + (1-\alpha)L^\beta\right]^{\frac{\gamma}{\beta}} \rightarrow \begin{cases} A\alpha^{\frac{\gamma}{\beta}}K^\gamma \text{ for } L \rightarrow^+ 0, \\ A(1-\alpha)^{\frac{\gamma}{\beta}}L^\gamma \text{ for } K \rightarrow^+ 0. \end{cases}$$

When $\beta < 0$, however, the right-hand side is not defined on the boundary. We circumvent this problem by redefining the CES function in the following way when $\beta < 0$:

$$Y = F(K,L) = \begin{cases} A\left[\alpha K^\beta + (1-\alpha)L^\beta\right]^{\frac{\gamma}{\beta}} \text{ when } K > 0 \text{ and } L > 0, \\ 0 \text{ when } K \text{ or } L \text{ equals } 0. \end{cases} \tag{4.49}$$

We now show that continuity holds in the extended domain. When $K > 0$ and $L > 0$, we have

$$Y^{\frac{\beta}{\gamma}} = A^{\frac{\beta}{\gamma}}\left[\alpha K^\beta + (1-\alpha)L^\beta\right] \equiv A^{\frac{\beta}{\gamma}}G(K,L). \tag{4.50}$$

Let $\beta < 0$ and $(K,L) \rightarrow (0,0)$. Then, $G(K,L) \rightarrow \infty$, and so $Y^{\beta/\gamma} \rightarrow \infty$. Since $\beta/\gamma < 0$, this implies $Y \rightarrow 0 = F(0,0)$, where the equality follows from the definition in (4.49). Next, consider a fixed $L > 0$ and rewrite (4.50) as

$$\begin{aligned} Y^{\frac{1}{\gamma}} &= A^{\frac{1}{\gamma}}\left[\alpha K^\beta + (1-\alpha)L^\beta\right]^{\frac{1}{\beta}} = A^{\frac{1}{\gamma}}L(\alpha k^\beta + 1 - \alpha)^{\frac{1}{\beta}} \\ &= \frac{A^{\frac{1}{\gamma}}L}{(ak^\beta + 1 - \alpha)^{-1/\beta}} \rightarrow 0 \text{ for } k \rightarrow 0, \end{aligned}$$

when $\beta < 0$. Since $1/\gamma > 0$, this implies $Y \rightarrow 0 = F(0,L)$, from (4.49). Finally, consider a fixed $K > 0$ and let $L/K \rightarrow 0$. Then, by an analogue argument we get $Y \rightarrow 0 = F(K,0)$, (4.49). So continuity is maintained in the extended domain.

4.10 Exercises

4.1 *(the aggregate saving rate in steady state)*

a) In a well-behaved Diamond OLG model let n be the rate of population growth and k^* the steady state capital-labor ratio (until further notice, we ignore technological progress). Derive a formula for the long-run aggregate net saving rate, S^N/Y, in terms of n and k^*. *Hint:* use that for a closed economy $S^N = K_{t+1} - K_t$.

b) In the Solow growth model without technological change a similar relation holds, but with a different interpretation of the causality. Explain.

c) Compare your result in a) with the formula for S^N/Y in steady state one gets in *any* model with the same CRS-production function and no technological change. Comment.

d) Assume that $n = 0$. What does the formula from a) tell you about the level of net aggregate savings in this case? Give the intuition behind the result in terms of the aggregate saving by any generation in two consecutive periods. One might think that people's rate of impatience (in Diamond's model the rate of time preference ρ) affect S^N/Y in steady state. Does it in this case? Why or why not?

e) Suppose there is Harrod-neutral technological progress at the constant rate $g > 0$. Derive a formula for the aggregate net saving rate in the long run in a well-behaved Diamond model in this case.

f) Answer d) with "from a)" replaced by "from e)". Comment.

g) Consider the statement: "In Diamond's OLG model any generation saves as much when young as it dissaves when old." True or false? Why?

4.2 (*increasing returns to scale and balanced growth)*