2. Neoclassical and 'new' growth theory: a critique

Our task in this chapter is to outline formally the assumptions and predictions of neoclassical growth theory as a background to showing, firstly, how the neoclassical production function is used for analysing growth rate differences between countries, and its weaknesses; and secondly, how neoclassical growth theory forms the basis for 'new' endogenous growth theory – the only major difference being that the assumption of diminishing returns to capital is relaxed, so that 'new' growth theory is subject to the same major criticisms as conventional neoclassical theory as far as analysing and understanding growth rate differences between countries is concerned.

The Neoclassical Model

The neoclassical growth model is based on three key assumptions. The first is that the labour force (l) and labour-saving technical progress (t) grow at a constant *exogenous* rate. The second assumption is that all saving is invested: $S = I = sY$. There is no independent investment function. The third assumption is that output is a function of capital and labour, where the production function exhibits constant returns to scale, and diminishing returns to individual factors of production. The most commonly used neoclassical production function, with constant returns to scale, is the so-called Cobb–Douglas production function, named after Charles Cobb, a mathematician, and Paul Douglas, a well-known Chicago economist before World War II (who later became a US senator). The function takes the form:

$$Y = TK^{\alpha}L^{1-\alpha}, \tag{2.1}$$

where Y is output, K is capital, L is labour, T is the level of technology, α is the elasticity of output with respect to capital and $1-\alpha$ is the elasticity of output with respect to labour. Obviously $\alpha + (1-\alpha) = 1$ (the assumption of constant returns to scale), so that a 1 per cent increase in capital and labour leads to a 1 per cent increase in output.

To consider the predictions of the model, it is convenient to transform equation (2.1) into its 'labour-intensive' form by dividing both sides by L, so that the dependent variable is output per head, and the independent variables are the level of technology and capital per head.

$$Y/L = (TK^{\alpha}L^{1-\alpha})/L = T(K/L)^{\alpha}$$

or

$$q = T(k)^{\alpha} \tag{2.2}$$

where q is output per head and k is capital per head.

The basic predictions of the neoclassical model, which can be shown diagrammatically (see below), are as follows:

1. in the steady state, the *level* of output per head (q) is positively related to the savings–investment ratio and negatively related to the growth of population (or labour force);
2. the *growth of output* is *independent* of the savings–investment ratio and is determined by the exogenously given rate of growth of the labour force in efficiency units ($l + t$). This is because a higher savings–investment ratio is offset by a higher capital–output ratio (or a lower productivity of capital) owing to the assumption of diminishing returns to capital;
3. given identical tastes and preferences (that is, the same savings ratio) and technology (that is, production function), there will be an *inverse* relation across countries between the capital–labour ratio and the productivity of capital, so that poor countries should grow faster than rich countries, leading to the *convergence* of per capita incomes across the world.

Figure 2.1 illustrates the first two predictions.

The production function, $q = f(k)$, with diminishing returns to capital, comes from equation (2.2). The ray from the origin with slope $(l + t)/s$ gives points of equality between the rate of growth of

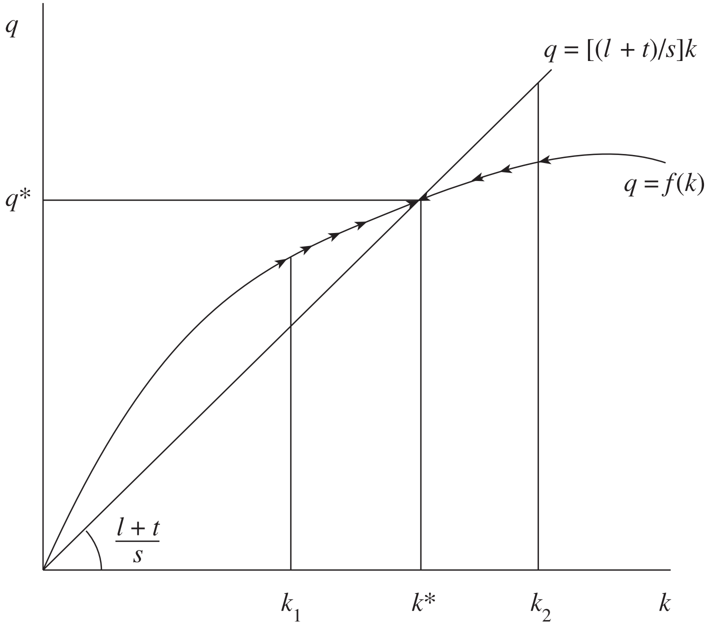

Figure 2.1

capital and labour measured in efficiency units.[1] Only at k^* is the level of output per head such as to give a rate of growth of capital equal to the rate of growth of the labour force. To the left of $k^*(k_1)$, the growth of capital is greater than the growth of labour, and economies are assumed to move along their smooth production function towards k^* using more capital-intensive methods of production. To the right of $k^*(k_2)$, the growth of capital is less than the growth of labour, and economies are assumed to use more labour-intensive techniques of production. At k^*, where the capital to labour ratio is in equilibrium, output per head will also be in equilibrium at q^*. It can be seen from the figure that a rise in the savings ratio (s) pivots downwards the ray from the origin and raises the equilibrium k and raises the *level* of q, but does not affect the growth rate of the economy. It can also be seen that the level of q will be inversely related to the rate of growth of the labour force because a rise in l pivots upwards the ray from the origin.

The explanation for convergence of per capita income across countries can be seen from the formula for the capital–output ratio:

$$K/Y = (K/L)\,(L/Y). \tag{2.3}$$

If there is diminishing returns to capital, a higher K/L will not be offset by a higher Y/L ratio, and therefore K/Y will be higher. Thus, if the savings–investment ratio is the same across countries, rich countries with a higher K/L ratio should grow more slowly than poor countries with a lower K/L because the productivity of capital is lower in the former case than in the latter.

What major criticisms can be made of this model, apart from the empirical fact that across the world we do not observe the convergence of living standards? The fundamental point to be made at this stage is that the neoclassical model is a *supply-oriented* model *par excellence*. First, demand never enters the picture. Saving leads to investment, so that supply creates its own demand. The neoclassical model of growth takes us back to a pre-Keynesian world where demand does not matter for an understanding of the determination of the level of output (and, by implication, the growth of output). Secondly, factors of production and technical progress are treated as *exogenously* determined, unresponsive to demand. But, by and large, the demand for factors of production is a derived demand, derived from the growth of output itself. Much technical progress and labour productivity growth is also induced by the growth of output itself (see later).

The assumption of exogeneity of factor supplies is no more apparent than in the studies that use the aggregate production function for analysing growth rate differences between countries; an approach pioneered by Abramovitz (1956) and Solow (1957) and still widely utilized. Let us consider this approach and comment on its limitations.

Using the Production Function for Analysing Growth Differences

If we go back to the Cobb–Douglas production function in equation (2.1), it is easy to see how it can be used for analysing the sources of growth; that is, decomposing a country's growth rate into the contribution of capital, labour and technical progress. The question is, how useful is it for a proper *understanding* of the growth performance of countries if the main inputs into the growth process are not exogenous but *endogenous*?

The function in equation (2.1) is made operational by taking logarithms of the variables and differentiating with respect to time, which gives:

$$y = t + \alpha(k) + (1 - \alpha)l, \tag{2.4}$$

or in labour-intensive form:

$$y - l = t + \alpha\,(k - l), \tag{2.5}$$

where lower-case letters represent rates of growth of the variables.

Given estimates of α and $(1 - \alpha)$, the contribution of capital growth and labour force growth to any measured growth rate can be estimated, leaving the contribution of technical progress as a residual. For example, suppose $y = 5\%$, $k = 5\%$, $l = 2\%$, $\alpha = 0.3$ and $(1 - \alpha) = 0.7$. The contribution of capital to growth is then (0.3) (5%) = 1.5 percentage points or 30 per cent; the contribution of labour is (0.7) (2%) = 1.4 percentage points or 28 per cent, leaving the contribution of technical progress as 5% – 2.9% = 2.1% or 42 per cent.

Solow (1957) was the first to use the labour-intensive form of the Cobb–Douglas production function in analysing the growth performance of the US economy over the previous 50 years, and concluded that only 10 per cent of the growth of output per man could be 'explained' by the growth of capital per man, leaving 90 per cent of growth to be 'explained' by various forms of technical progress. Denison (1962, 1967) used the same production function approach, or growth accounting framework, to study growth performance in the USA and between the countries of Europe, disaggregating the technical progress term (or residual) into various component parts. Maddison (1970) used the approach to study growth rate differences between developing countries. Since this early research, there has been a mass of other studies too extensive to survey here (however, see Felipe, 1999), but two recent studies may be mentioned as illustrative. The World Bank (1991) did a study of 68 countries showing capital accumulation to be of prime importance, with technical progress minimal. This seems to be the central conclusion for developing countries in contrast to developed countries. Secondly, there is the controversial study by Alwyn Young (1995) of the four East Asian 'dragons' of Hong Kong, Singapore, South Korea and Taiwan which also shows that most of the growth in these countries can be explained by the growth of factor inputs and not technical progress, so that, according to Young, there has been no growth miracle in these countries – contrary to the conventional wisdom.

Before accepting this conclusion, however, the observer still has to explain why there was such a rapid growth of factor inputs, and it is this point which exposes the fundamental weakness of the production function approach to the analysis of growth performance. Inputs are not manna from heaven dropped by God. Something 'miraculous' must have been driving these economies, to which input growth responded. On closer inspection, what distinguishes these countries is their outward orientation and relentless search for export markets, and their remarkable growth of exports which confers benefits on an economy from both the demand and the supply side (see Chapter 4). This exposes another weakness of neoclassical growth theory and that is that the models are closed. There is no trade in these simple models, and no balance of payments to worry about. They are supply-oriented, supply-driven, closed economy models unsuitable for the analysis of open economies in which foreign exchange is invariably a scarce resource acting to constrain the growth process. We return to this topic in Chapters 4 and 5, but first we must look at the challenge of 'new' growth theory.

'New' Endogenous Growth Theory

Since the mid-1980s there has been an outpouring of literature and research on the applied economics of growth, attempting to understand and explain differences in output growth and living standards across countries of the world – most inspired by so-called 'new' growth theory or endogenous growth theory. This spate of studies seems to have been prompted by a number of factors: firstly, by the increased concern with the economic performance of poorer parts of the world, and particularly major differences between continents and between countries, with South East Asia forging ahead, Africa left behind and South America somewhere in the middle; secondly, by the increased availability of standardized data on which to do research (Summers and Heston, 1991); and thirdly, by studies showing *no convergence* of per capita incomes in the world economy (for example, Baumol, 1986), contrary to the prediction of neoclassical growth theory based on the assumption of diminishing returns to capital.

If there are not diminishing returns to capital – but, say, constant returns – a higher capital–labour ratio will be exactly offset by a higher output per head,[2] and the capital–output ratio will not be higher in capital-rich countries than in capital-poor countries, and the savings–investment ratio will therefore matter for long-run growth. Growth is endogenously determined in this sense and not simply determined by the exogenous rate of growth of the labour force and technical progress. This is the starting point for 'new', endogenous growth theory which seeks an explanation of *why* there has not been a convergence of living standards in the world economy.

The explanation of 'new' growth theory is that there are forces at work which prevent the marginal product of capital from falling (and the capital–output ratio from rising) as more investment takes place as countries get richer. Paul Romer (1986) first suggested externalities to research and development (R&D) expenditure. Robert Lucas (1988) focuses on externalities to human capital formation (education). Grossman and Helpman (1991) concentrate on technological spillovers from trade and foreign direct investment (FDI). Other economists have stressed the role of infrastructure investment and its complementarity with other types of investment. In fact, it can be seen from the formula for the capital–output ratio that increasing returns to labour for all sorts of reasons could keep the capital–output ratio from rising.

So now let us turn to 'new' growth theory, see what it has to say, see whether it is saying anything new, and consider some of the problems of interpreting the empirical results from testing new growth theory.

The first crude test of new growth theory is to observe whether or not there is an inverse relation across countries between the growth of output per head and the *initial* level of per capita income of countries. If there is, this would be supportive of the neoclassical prediction of convergence. If not, it would be supportive of 'new' growth theory that the marginal product of capital does not decline. This is referred to as the test for beta (β) convergence. It can be said straight away that no global studies find evidence of *unconditional* beta convergence. Virtually all studies find evidence of divergence. The coefficient linking the growth of output per head to the initial level of per capita income is positive, not negative.

Before jumping to the conclusion that this is unequivocal support for 'new' growth theory, however, it must be remembered that the neoclassical prediction of convergence assumes all other things the same across countries: population growth; tastes and preferences (for example, the savings ratio); technology and so on. Since these assumptions are manifestly false, there can never be the presumption of *unconditional* convergence – only *conditional* convergence controlling for differences in all other factors that affect the growth of living standards, including differences in the ratio of investment to GDP and variables that affect the productivity of capital and labour such as education and training, R&D expenditure, trade, macroeconomic performance and political stability. The question is, what happens to the sign on the initial per capita income variable when these control variables are introduced into the equation? If the sign on initial per capita income turns negative, this is supposed to represent a rehabilitation of the neoclassical model. In other words, living standards would converge *if only* levels of investment, education, R&D expenditure and so on were the same in poor countries as rich countries, but they are not! The argument is reminiscent of the way neoclassical economists continue to work with fictitious models of competitive equilibrium in the presence of increasing returns, by treating the latter as externalities (the device originally adopted by Alfred Marshall in 1890). Indeed, most 'new' growth theorists, and particularly Robert Barro (1991), are clearly neoclassical economists in disguise. We will look at the work of Barro and others later, but first let us consider the 'newness' of 'new' growth theory and the interpretation of results.

First, I find it amusing that it seems to have come as a surprise to many members of the economics profession that living standards in the world have not been converging according to the prediction of neoclassical growth theory. Long before the advent of 'new' growth theory, many 'non-orthodox' economists had been pointing to widening divisions in the world economy, and developed models to explain divergence. That is what the centre–periphery models of Prebisch (1950), Myrdal (1957), Hirschman (1958), Seers (1962) and the neo-Marxist school (for example, Emmanuel, 1972; Frank, 1967) were all about, many based on a combination of international trade and increasing returns.

Secondly, it has to be said that many of the ideas of 'new' growth theory are not new at all. Who, apart from strict adherents to the neoclassical model, ever believed that investment did not matter for long-run growth? Kaldor (1957), with his technical progress function, precisely anticipated new growth theory by arguing that technical progress requires capital accumulation and capital accumulation requires technical progress (it is impossible to have one without the other), and his model of growth gives an explanation of why the capital–output ratio stays constant through time despite a rising ratio of capital to labour (see later). On the origins of increasing returns, we could mention Adam Smith and the division of labour (see Chapter 1), Allyn Young and the idea of increasing returns as a macroeconomic phenomenon related to the interaction between activities (see Chapter 1), Kenneth Arrow's model of learning by doing (Arrow, 1962), the work of Schultz (1961) and Denison (1962) on the social returns to education, and the work of Griliches (1958) on the social returns to R&D. We have an endearing tendency in economics to reinvent the wheel.

Thirdly, when it comes to interpreting the empirical results from testing models of new growth theory and convergence, some care needs to be taken. In particular, great care needs to be exercised in interpreting the negative sign on the initial level of per capita income as necessarily rehabilitating the neoclassical model of growth, as for example, Barro (1991) does, because there are other conceptually distinct reasons for expecting a negative sign. Firstly, outside the neoclassical paradigm, there is a whole body of literature that argues that economic growth *should be* inversely related to the initial level of per capita income because, the more backward a country, the greater the scope for *catch-up*; that is, for absorbing a backlog of technology, which represents a shift in the whole production function. Is conditional convergence picking up diminishing returns to capital in the neoclassical sense, or catch-up? The two concepts are conceptually distinct, but not easy to disentangle empirically. Secondly, the negative term could simply be picking up structural change, with poor countries growing faster than rich countries (controlling for other variables) because of a more rapid shift of resources from low productivity to high productivity sectors (for example, from agriculture to industry). How do we discriminate between these hypotheses?

A fourth point concerns the specification of 'new' growth theory in its simplest form as the so-called *AK* model:

$$Y = AK, \tag{2.6}$$

where A is a constant, which implies a constant proportional relation between output (Y) and capital (K), or constant returns to capital. On close inspection, this specification is none other than the Harrod growth equation $g = s/c$ (see Chapter 1). This can be seen by taking changes in Y and K and dividing by Y, which gives:

$$\Delta Y/Y = A\,\Delta K/Y = A\,(I/Y), \tag{2.7}$$

where $\Delta Y/Y$ is the growth rate (g); I/Y is the savings–investment ratio (s), and A is the productivity of investment, $\Delta Y/I = 1/c$ or the reciprocal of the incremental capital–output ratio. What this means is that, if the productivity of investment (A) was the same across all countries, there would be a perfect correlation between growth and the investment ratio. If there is not a perfect correlation, then *definitionally* there must be differences across countries in the productivity of capital. All that empirical studies of 'new' growth theory are really doing is trying to explain differences in the productivity of capital across countries (provided the investment ratio is in the equation) in terms of differences in education, R&D expenditure, trade and so on, and initial endowments (see Hussein and Thirlwall, 2000, for further elaboration of this point).

As far as the constancy of the capital–output ratio is concerned, it was pointed out by Kaldor (1957) many years ago, as one of his six stylized facts of economic growth, that, despite capital accumulation and increases in capital per head through time, the capital–output ratio has remained broadly unchanged, implying some form of externalities or increasing returns. It is worth quoting Kaldor in full:

> As regards the process of economic change and development in capitalist societies, I suggest the following 'stylised facts' as a starting point for the construction of theoretical models – (4) steady capital–output ratios over long periods; at least there are no clear long-term trends, either rising or falling, if differences in the degree of capital utilization are allowed for. This implies, or reflects, the near identity in the percentage growth of production and of the capital stock i.e. for the economy as a whole, and over long periods, income and capital tend to grow at the same rate.

Kaldor's explanation lay in his innovation of the technical progress function (TPF) relating the growth of output per man ($\dot{q}$) to the growth of capital per man ($\dot{k}$), as in Figure 2.2.

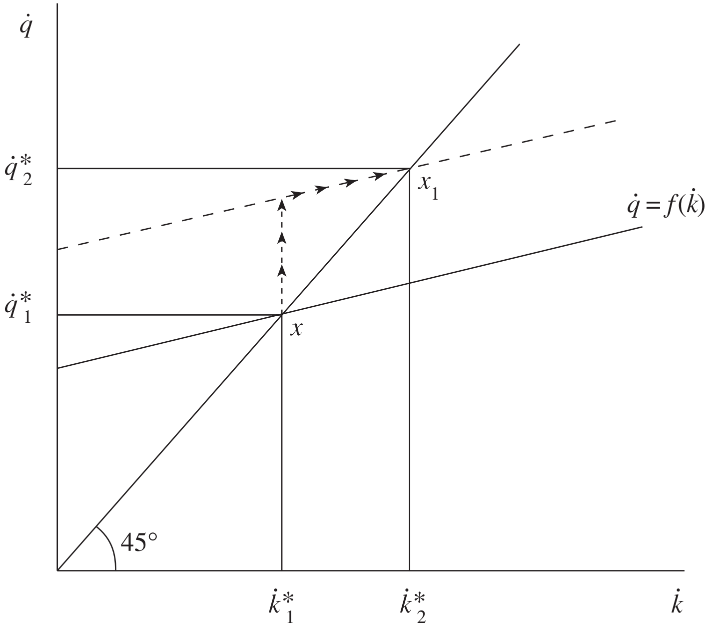

Figure 2.2

The position of the (linear) TPF drawn in Figure 2.2 depends on the exogenous rate of technical progress, and the slope of the function depends on the extent to which technical progress is embodied in capital. Along the 45° line, the capital–output ratio is constant, and the equilibrium growth of output per head is $\dot{q}^*_1$. An upward shift of the function associated with new discoveries, technological breakthroughs and so on will cause the growth of output to exceed the growth of capital, raising the rate of profit and inducing more investment, to give a new equilibrium growth of output per head at $\dot{q}_2^*$ (follow the arrows). An increase in capital accumulation not accompanied by technical progress will simply cause the capital–output ratio to rise. If the capital–output ratio is observed to be constant there must be technological forces at work shifting the function upwards. 'New' growth theory is precisely anticipated.

What applies to countries through time applies *pari passu* to different countries at a point in time, with differences in country growth rates at the same capital–output ratio associated with different technical progress functions. To quote Kaldor again:

> A lower capital–labour ratio does not necessarily imply a lower capital–output ratio – indeed, the reverse is often the case. The countries with the most highly mechanised industries, such as the United States, do not require a higher ratio of capital to output. The capital–output ratio in the United States has been falling over the past 50 years whilst the capital–labour ratio has been steadily rising; and it is lower in the United States today than in the manufacturing industries of many underdeveloped countries. (Kaldor, 1972)

In other words, rich and poor countries are simply not on the same production function.

A final point concerns the way that new growth theory models trade. First of all, some of the models and empirical studies do not consider the role of trade at all, as if economies are completely closed. It is hard to imagine how it is possible to explain growth rate differences between countries without reference to trade, and particularly without reference to the balance of payments of countries which constitutes for many developing countries the major constraint on the growth of demand and output (which will reduce the productivity of capital). When a trade variable is included in the model, it is invariably insignificant, or loses its significance when combined with other variables. On the surface, this is a puzzle. It would conflict with the rich historical literature that exists on the relation between trade and growth (Thirlwall, 2000). It would conflict with the voluminous work of the World Bank and other organizations showing the beneficial effects of trade liberalization, and it would undermine the whole thrust of international policy making since World War II, which has been to free up markets and to promote trade in the interests of economic development.

There may be several explanations for the weak results, but I believe the major one is that the trade variable normally taken is the *share* of exports in GDP as a measure of 'openness' which may pick up the static gains from trade and technological spillovers, but not the *dynamic* effects of trade which can only be properly captured by the *growth* of exports which affects demand, both directly and indirectly (by relaxing a balance of payments constraint on demand), and also the supply side of the economy by permitting a faster growth of imports. This point relates to my general criticism of 'new' growth theory that it neglects demand-side variables. When an export growth variable is included in a 'new' growth theory equation, it is highly significant (see Thirlwall and Sanna,1996).

When it comes to evaluating the empirical evidence, only four variables in 'new' growth theory equations appear to be robust (see Levine and Renelt, 1992): the initial level of per capita income, the savings–investment ratio, investment in human capital, and population growth (usually). All other variables are fragile in the sense that, when they are combined with other variables, they lose their significance. The robust variables are ones that growth analysts have stressed for many years, long before the advent of 'new' growth theory. *Plus ça change, plus c'est la même chose*.

Notes

1. This can be seen by rearranging the equation $q = [(l + t)/s]\,k$ to $qs/k = l + t$, where $q = Y/L$; $s = S/Y = \Delta K/Y$ (since all saving leads to capital accumulation) and $k = K/L$. Therefore $(Y/L)\,(\Delta K/Y)\,(L/K) = \Delta K/K = l + t$.
2. Remember $K/Y = (K/L)/(Y/L)$.