Ronald Schettkat November 2004

Two Blades of the Scissors
The Interaction between Demand and Supply in Market Economies

5. The Interaction of Demand and Supply: Evolutionary Growth Theory

5.1 Neoclassical versus Evolutionary Growth Theory

The steady state and an economy's tendency to move back to equilibrium after any 'exogenous shocks' are the hallmarks of neoclassical economic theory. The ingredients for this perfect market scenario are diminishing returns to individual factor inputs and constant returns to scale. In this model a deviation from equilibrium will immediately leash forces, which shift the market back to equilibrium. Negative feedback effects dominate neoclassical models. In essence, the neoclassical world is a static world in which capital owners only shift existing resources to the most profitable locations. Permanent change only occurs as continuous shift of an equilibrium or as steady-state growth (see Chapter…..).

In retrospect, however, economic development looks more like a story of substantial changes rather than one of a smoothly moving equilibrium. Along the way, totally new products (1), radically different modes of production (2), and new markets were created (3), new materials were used (4) and industry structures changed totally (5) (see Chapter 2). These are also the five developments Joseph Schumpeter (1911) listed a responsible for discontinuous changes in economic development. He grouped them together under the term ‘technological progress’ or innovation, which he believed to be the result of entrepreneurial activity.

Although every economist is aware of the tremendous changes that have been taking place, and recognizes fluctuations in economic activity, most traditional economic models ignore such developments. They approximate growth processes by steady states, and they blame deviations from the steady state, that is, from the equilibrium growth path, on exogenous shocks, and they consider a distortion of the equilibrium to be a short-run phenomenon. And although they recognize that frictions or market imperfections (which are thought to be especially relevant to labor markets) may slow down adjustments, they maintain that in the long –run, the equilibrium assumption offers an adequate representation of actual economic trends.

Schumpeter, the eminent evolutionary economist, argued that economic development is mainly caused by distortions of the equilibrium. In his view, ‘creative destruction’, that is, the creation of new products and production techniques (technological progress), is the driving force behind economic progress. Unlike traditional theories, Schumpeter’s theory is dynamic; it is based on the assumption that positive feedback effects will push the economy to a new structure, rather than on the assumption that negative feedback effects will drive the economy back to its old equilibrium. It assumes that the structure of the economy will evolve and that economic development is a supply-side driven evolutionary process. Entrepreneurs innovate and although Schumpeter saw that the demand side also needs to adjust (see Chapter 7), it is the supply side, which gives the initial impetus to change.

The Schumpeterian growth model emphasizes the heterogeneity of firms and takes differences between firms to be the major source of change and technological progress, a process commonly known as creative destruction. Meanwhile there is strong empirical evidence that the idea of 'creation and destruction' better describes capitalist economic processes than the concept of the representative firm. That is in every industry (every market) a huge variation in economic performance occurs not adequately captured by the average firm in an industry (the representative firm). Analysis of job-turnover, that is, the creation and destruction of jobs in establishments in an industry, clearly shows that both during expansionary and during contractionary periods, and both in growing as well as in shrinking industries, some firms grow while others shrink (Leonard 1987, 1996, and also Davis/ Haltiwanger 1990, Schettkat 1992, 1996). Successful firms are the ones that innovate their products and processes and it is clearly shown that the innovators gain part of their market share at the expense of non-innovating firms (Scholz/ Penzkofer/ Schmalholz/ Beutel 1990, Matzner/ Schettkat/ Wagner 1990, Greenan/ Guellec l996). There are obviously some firms, which develop a culture of innovation, or routines of innovations (Nelson/ Winter 1974, 1982), and these are the most successful competitors. Passive firms, that are firms that merely react to changes in the environment, become the losers.

Even in shrinking markets (which may be represented by an industry), some firms grow while others decline, which is possible only if some firms gain market share at the expense of others. Losing firms are usually the ones that do not innovate, indicating that the conditions allowing an individual firm to expand differ from those determining the expansion of the market as a whole. Even an expanding industry, which allows a greater number of firms to expand, cannot prevent some firms from losing demand to their competitors although survival is easier in growing markets.

Schumpeter's main point is that it requires an entrepreneur rather than an administrator to deal with the uncertainty[1] inherent in the innovation process. This uncertainty varies from one type of innovation to another. In the case of fundamental research there is ‘true uncertainty’. For example, when the laser was first invented, no one had a clue what to use it for. In the case of minor technical improvements or product differentiation, there is relatively little uncertainty. Table 5.1 provides an

[1] 'True uncertainty' exists when the probability distribution is unknown. This distinguishes it from risk. (Knight 1921).

overview of various types of innovations and their corresponding degrees of uncertainty.

Table 5.1: Types of Innovation and their Corresponding Degree of Uncertainty

Type of Innovation	**Degree of Uncertainty**
Fundamental research Fundamental invention	True uncertainty
Radical product innovations Radical process innovations outside the firm	Very high degree of uncertainty
Major product innovations Radical process innovations inside the firm	High degree of uncertainty
New ‘generations’ of established products	Moderate uncertainty
Licensed innovations Limitations to product innovations Modification of products and processes Early adoption of established processes	Little uncertainty
New ‘model’ Product differentiation Agency for established product innovation Late adoption of established processes Innovation and franchised operations inside the firm Minor technical improvements	Very little uncertainty

Source: Freeman/Soete 1997:244.

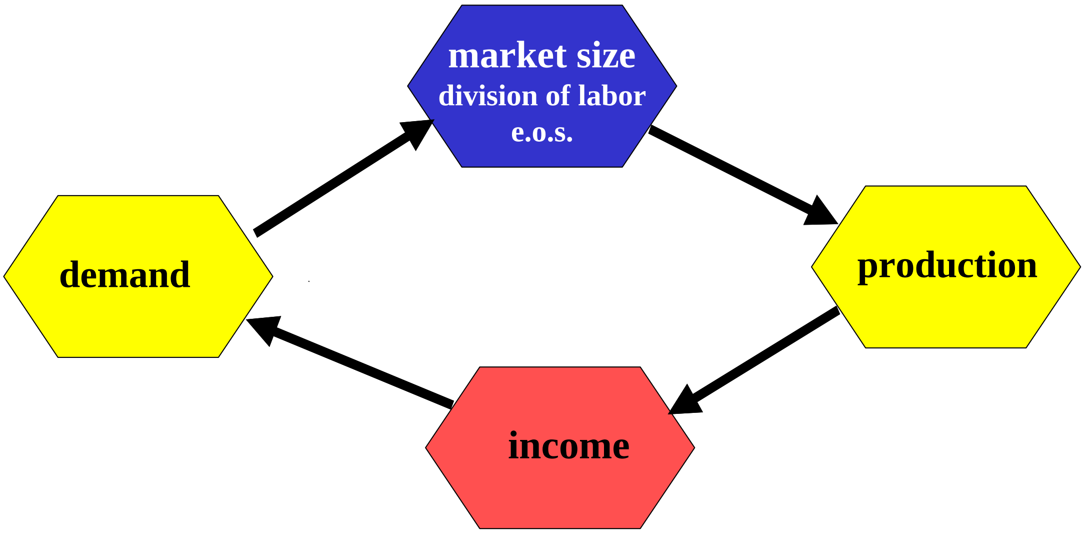
Although Schumpeter is the most renowned evolutionary economist, Adam Smith (1776) was the first to identify its key ingredient: increasing returns to scale. Ever since Adam Smith, productivity growth, or greater efficiency in production, has been associated with specialization (the division of labor) and learning, which are both dependent on the size of the market. That is, the bigger the market, the more profitable is specialization and the higher the learning component of specialized activities. In Adam's Smith 'Wealth of Nations', this reasoning is illustrated by the famous pin example. Two centuries later, Smith's reasoning was part of the economic rationale for establishing the European Union.[2] As Adam Smith already postulated, increasing economies of scale are characteristic of manufacturing production and are related to the accumulation of knowledge, capital, and technology. Mass production processes and production processes with high fixed costs such as, for example, the production process of software, typically entail increasing returns to scale, and thus, declining marginal and average costs. The division of labor makes it possible to simplify tasks so that they can be taken over by machines.

According to Adam Smith, the size of a market is largely determined by the size of its geographical unit, or by the size of the population it covers. Allyn Young (1928) argued that a population's purchasing power, which determines its capacity to absorb large amounts of goods, is an equally, or probably even more important[3], determinant of market size. In Young's conception, the division of labor determines in large parts the division of labor; for it is the division of labor that allows for greater efficiency and more production, and thus for higher income (Young 1928, 533). The capacity to absorb production, that is demand, is in turn highly dependent on income and thus on the capacity to produce. Market expansion, in other words, is an interactive process, in which demand depends on supply, and supply depends on demand. As illustrated in Figure 5.1, this process entails positive feedback effects, in the sense that both sides of the market are mutually dependent.

[2] This argument has been used to explain the productivity lead of the US, which has a bigger market and therefore greater economies of scale (for example, Young 1928: 532).

[3] Allyn Young modestly presents his extension of Smith's ideas as no more than a minor amendment (see Young 1928: 529).

Figure 5.1: Dynamic Economies of Scale (E.O.S.)

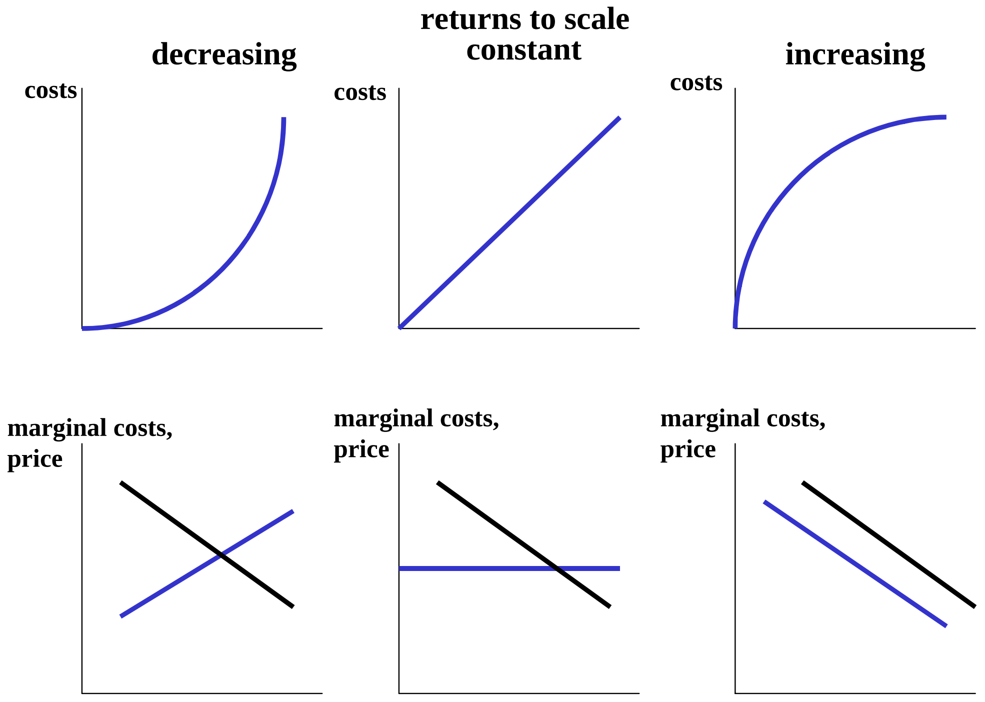


The impact of an increase in production capacity on growth - if employment and capital are held constant and the rate of capacity growth equals the rate of productivity growth – depends on the reaction of the demand side. The balance between demand and supply is precarious: if additional production does not translate into additional income, that is, if there is no demand for additional production, the virtuous circle of demand and supply can turn into a vicious circle since positive feedback, or the amplification of effects, can work in two directions. Although Young argues that production will only be absorbed if product demand is elastic, he assumes that there are no demand restrictions at the level of the aggregate economy: '… there are no limits beyond which demand is not elastic and returns do not increase' (Young 1928: 534). Young makes the assumption that Say's law (or the full employment assumption) holds, meaning that there is a demand elasticity of at least one at the aggregate level. Under such conditions, supply can find its own demand, and additional production would be fully absorbed. Thus, if Say's law would hold, there would be no need to investigate the demand side of the story.[4] Kaldor (1985) points out that Young wrote his path- breaking paper at a time when the US economy was experiencing a long period of economic expansion, which probably led Young to focus only on the bright side of feedback effects.

If demand does not keep pace with capacity growth, positive feedback effects may, however, create a vicious circle. Although Young acknowledged that this could occur at the industry level, he still assumed a demand elasticity of one at the aggregate level. 'The rate at which any one industry grows is conditioned by the rate at which other industries grow, but since the elasticities of demand and supply will differ for different products, some industries will grow faster than others.' (Young 1928: 534, see also Chapter….). In other words, elasticities of demand are considered important for explaining structural change, or the changing composition of the economy by industry, which may be slowed down by frictions (Young 1928: see Chapter 6), but at the aggregate level the weighted industry-specific demand elasticities are assumed to add up to one.

[4] However, according to Allyn Young (1928: 534) there are some sources of friction: 'human material', as human capital was called in the 1920s, is resistant to change, and the accumulation of the necessary capital may be problematic.

Box 5.1: Cost Functions and Derived Product Supply; Diminishing, Constant and Increasing Returns to Scale



Decreasing returns to scale [f(aK, aL) > aY]
Production functions with decreasing returns to scale were derived from agricultural production, in which land is the limiting input factor and which may prevent the simple duplication of the production process. To increase production, substitutes for land need to be found. This implies that productivity growth depends on the availability of substitutes, rather than on scale as such.

Constant returns to scale [f(aK, aL) = aY]
Constant returns to scale occur if none of the input factors is limited. Under such conditions, production processes can simply be replicated

Increasing returns to scale [f(aK, aL) < aY]
Increasing returns to scale arise from technology in the wider sense of the word, that is, from increases in the efficiency of production through the division of labor, from the spread of fixed costs, the use of more efficient technology. Increasing returns to scale are inherently technological.

Why do economies of scale actually occur? In general, the unit cost of any product with a fixed cost component such as, for example, high R&D costs, decreases as long as its variable input factors are unlimited. Software programs are probably the best example of products with decreasing average costs at the product level. Whereas it is very costly to develop new software, producing an additional copy of an existing program can be done at almost zero costs (for an investigation of the information economy see Shapiro/ Varian 1999). However, there are other sources of economies of scale. As mentioned above, Young saw economies of scales to be the result of an increasing division of labor and improvements in knowledge. ‘In this circumstance lies the possibility of economic progress, apart from the progress which comes as a result of the new knowledge which men are able to gain, whether in the pursuit of their economic or their non-economic interests.' (Young 1928: 540). The division of labor can occur at the firm level, but may also result from specialization at the industry level, that is, from the inter-industry division of labor. Such ‘roundabouts’ make ‘technological progress’, and the accompanying rise in efficiency, a macro-economic phenomenon (for an analysis see Russo/ Schettkat 2001, Gregory/ Grenagh 2001, ten Raa/ Wolf 2001).

Table 5.2 provides a summary of the main differences between neoclassical and evolutionary economics.

Table 5.2: Neoclassical versus Evolutionary Economics

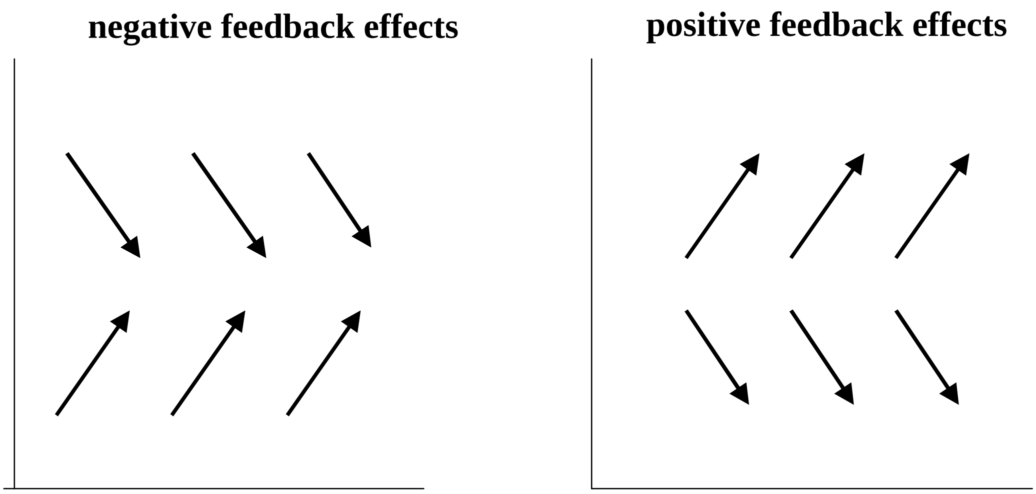
Neoclassical Economics	**Evolutionary Economics**
Decreasing returns to scale Constant returns to scale	Increasing returns to scale
Negative feedback	Positive feedback
Unique equilibrium	Multiple equilibria
Static (timeless)	Dynamic (sequential)
Known alternatives	Learning
Perfect rationality	Bounded rationality
Risk	Uncertainty
Exogenous technological progress	Cumulative technological progress
Homogeneity	Heterogeneity, variety, selection
Analysis of equilibrium behavior	Analysis of ‘disequilibrium behavior’
Individually independent behavior (methodological individualism)	Individually interdependent behavior

5.2 Increasing Returns to Scale and Cumulative Causation

Cumulative causation occurs when positive feedback effects reinforce economic trends. The term ‘cumulative causation’ was coined by Gunnar Myrdal (ig..), a Swedish economist and Nobel Prize winner, who dedicated a substantial part of his work to development economics. - Myrdal used the concept to refer to uneven economic development, and to explain why there had been no convergence between developed and developing countries (see Chapter 4). ‘Cumulative causation’ is a circular process that is set in motion by increasing returns to scale. Whenever costs and prices fall, and quality improves with the number of units produced, cumulative causation, or the self-enforcing mechanism of positive feedback effects, occurs, that drives the market away from its equilibrium. Paul Krugman (1990) uses the idea of cumulative causation in his ‘new trade theory’ to explain why international trade mainly occurs between developed countries, and why it often occurs within the same industries. In his view, the existence of increasing returns forces countries to specialize and leads to certain fixed patterns of trade. Krugman also uses cumulative causation to explain differences in regional development and the formation of industrial clusters (Krugman 1991).

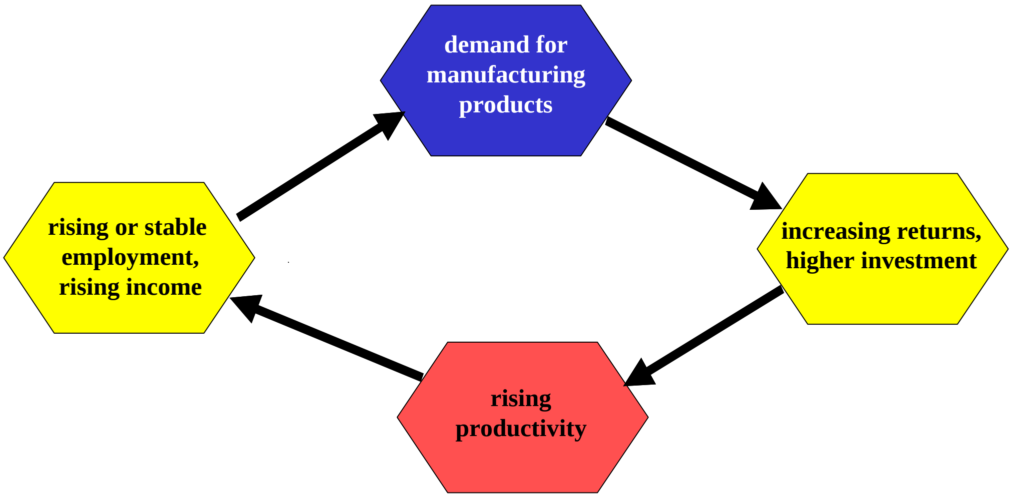
The concept of cumulative causation is also particularly useful when trying to explain phenomena in the information economy, in which fixed costs are important and technological developments are the result of learning (see Shapiro/ Varian 1999). At the firm level, increasing returns to scale lead to the creation of monopolies and ‘winner-takes-all’ markets (Frank/ Cook 1993). Firms that take a lead in output will be able to lower their production costs, and thus their price, which will allow them to attract even more demand. Because of economies of scale, this will enable them to reduce the costs per unit even further. Under such conditions, the market will end up in a ‘corner solution’, and only one firm will be able to stay in the market. Small differences may determine whether firms shift from a virtuous to a vicious circle, or vice versa. As illustrated in Figure 5.2, in the case of negative feedback effects, a market will always convergence to its equilibrium, whereas positive feedback effects will push a market away from its initial position into a corner solution. One may say that processes with positive feedback effects “explode”.

Figure 5.2: Negative and Positive Feedback Effects



Positive feedback effects and cumulative causation occurs not only at the firm level, but also at the level of the macro economy, where their impact is probably even stronger. Given increasing returns to scale, which may result from industry specialization, or the inter-industry division of labor, powerful interactions between demand and supply may occur. Since manufacturing industries are characterized by increasing returns to scale, they were identified by Kaldor (1966) as the 'engine of growth'. Meaning that, if the manufacturing industry experiences a positive demand shock, its productivity will rise, which will lead to higher incomes in the manufacturing sector. This increase in income will, in turn, be spent on manufacturing products, and will create a virtuous circle. If the share of manufacturing income rises, economy-wide productivity will also rise, particularly if employment is reallocated from less productive industries to the more productive manufacturing industry (see Figure 5.3) {Kaldor, 1966 #122}. In other words, the demand for manufacturing goods will create positive feedback effects, which will raise productivity by allowing increasing economies of scale. The rise in productivity will, in turn, be translated into higher incomes, which will create an even higher demand for goods, which will again lead to a rise in productivity, and so on and so forth. Thus, the causes of productivity growth have cumulative effects, which will eventually drive the economy away from its initial position.

Figure 5.3: Manufacturing: Kaldor's Demand-Driven Engine of Growth



Analyzing Britain's comparatively low rates of economic growth after WWII (see Chapter ..), Kaldor (1966) concludes that this is mainly due to the slow expansion of its manufacturing industries.[5] According to Kaldor, limited labor reserves were hampering the growth of Britain's manufacturing industries and, as a consequence, the economy's overall productivity growth. Originally, Kaldor thought that this process of cumulative causation resulted from the lack of adequate 'labor reserves' to feed the expanding manufacturing industries.[6] He assumed that these reserves consisted mainly of the labor forces employed in agriculture. Since Britain had access to agricultural products from its colonies, its agricultural sector was relatively small,

---

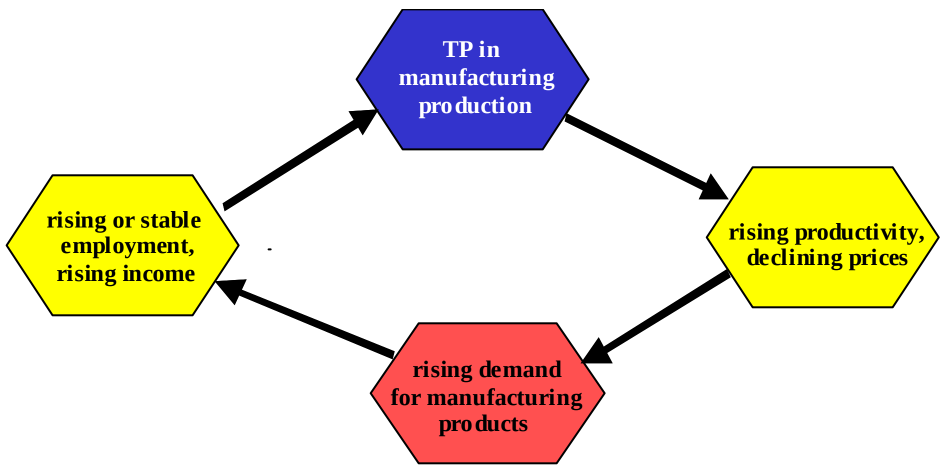
5 Like Allyn Young, Kaldor also assumes 'full employment' or insatiable demand (Kaldor 1986).

6 Later Kaldor (1996) adopts a different approach in which effective demand is the driving force of cumulative causation.

and so were its labor reserves. According to Kaldor, this was the bottleneck for the expansion of the country's manufacturing industries, which slowed down Britain's productivity growth.

Kaldor's cumulative growth process is set off by the demand for manufacturing products and is then driven by positive feedback effects. Cumulative causation can, however, also be explained from the supply-side as is done by Salter (1960). For Salter, the virtuous circle starts with technological progress in the manufacturing industry. This allows prices to fall, which, in turn, stimulates the demand for manufacturing goods, and this leads eventually to an expansion of the manufacturing industry. Thus, although Salter and Kaldor both explain a rise in the efficiency of the overall economy by a rise in productivity in the manufacturing industry, and by a greater weight of manufacturing industries, their starting points differ (compare Figure 5.4 with Figure 5.3).

Figure 5.4: Salter's Growth Process



Both Salter's and Kaldor's growth models center on changes in the structure of the economy, which are treated as endogenous in both models. However, even when abstracting from the underlying structure, there are arguments to support the idea of positive feedback effects or cumulative causation at the macroeconomic level. Schumpeter thought of recessions as the periods in which innovations break through and technological progress is rapid. However, as neo-Schumpeterian economists argue, it may also be the case that periods of high investments are conducive to innovations (Schmookler 1962). If investments stimulate innovations and, if innovations, in turn, stimulate new investments, and learning by doing (Arrow), a virtuous circle will develop, which will be driven by dynamic rather than static economies of scale.

Box 5.2: Static versus Dynamic Returns to Scale

In the case of static returns to scale, productivity is dependent on output. In this mechanical relationship, productivity and output move in conjunction: an expansion of output raises productivity, and a contraction lowers productivity. The effects are symmetrical.

Dynamic returns to scale, in contrast, are asymmetrical and result in permanent changes in production methods. That is, if during an expansionary period output growth and higher investments stimulate learning, this increase in knowledge will remain, even if a recession follows, and thus the path of economic growth has been changed irreversibly. Times have changed.

In the growth accounting literature, increases in investment are assumed to lead only to transitory economic expansions, and not to permanent changes in an economy's long-run growth rate. Hence, in the growth accounting equation, the effects of variations in the capital stock are captured by the capital variable. However, as discussed in Chapter 3, investments in new equipment will be made, which will most likely be technologically superior to older generation equipment. This so-called 'embodiment effect', is included in augmented growth equations, in which capital, and labor, (see Chapter 3.3) are 'quality corrected', and new-generation capital is given additional weight.[7] According to the Solow growth model, however, changes in the capital stock will affect an economy's growth rate only temporarily, i.e. increased capital per worker leads to a move along the production function. If technological progress is embodied in new equipment, high rates of productivity growth (such as occurred during the immediate post-war decades, for example) may be the result of high rates of investment (Nelson 1964). In a similar vein, reductions in investment can help to explain the slower pace of technological change. This is the favorite hypothesis of the late Zvi Griliches, who used it to explain the productivity slowdown that occurred in the industrialized world after 1973 (Griliches 1988). The productivity growth and investment hypothesis (Gordon 1999, 2000) was also used to counter the voluminous and fashionable writing on the new economy, which was highly popular in the late 1990s.

Nelson, in his comprehensive 1964-article on 'Aggregate Production Functions and Medium Range Growth Projections', draws attention also to the positive feedback effects of incentives {Nelson, 1964 #293} (Nelson 1964: 595). If, as the embodiment story suggests, the growth of 'total factor productivity' and the growth of capital complement each other, then improvements in the design of new equipment could stimulate investments. Technological obsolescence turns out to be indeed one of the major reasons for writing off old equipment and for investing in newer models, as is most clearly demonstrated by the short life span of computer equipment.

[7] Another possibility would, of course, be raising α instead of using α*gk*q. Although the residual would in both cases be reduced, in the latter case the theoretical justification of α as a measure of the marginal product is preserved. However, as Nelson (1964) justly observes, this adjustment does not solve the problem, but merely moves it to an earlier stage: for what factor other than technological progress could explain an increase in efficiency?

Figure 5.5: A Virtuous Circle of Investment, Innovation, Investment, and Learning