Until recently black workers have not enjoyed job security in the coal mining industry. After the expiry of their contract they had to "take their chance", along with other black workers, at receiving a further contract after they had returned home. Although a man with experience was always preferred to a novice his re-engagement depended upon the state of demand for coal at that time and there was no guarantee that he would return to the same colliery, or the same job, or receive the same rate of pay. In effect, therefore, migrant blacks enjoyed job security only for the duration of their contract, not permanent job security. However, the recent innovation of the re-engagement certificate now provides some measure of job continuity to more-skilful black migrants who are beginning to regard mining as a career. Some mining groups are increasingly moving towards the situation where they regard their black workforce as a permanent one, subject to annual leave conditions.

Section 10 urban blacks do not work on contracts. However, prior to 1971 such blacks, if made redundant, could only seek further work within the boundaries of the local-authority area in which they were registered. Between 1971 and 1980 this mobility was extended to the boundaries of the administration-board area in which they were registered and since 1980 they have enjoyed nationwide mobility in search of further employment, and can settle in any urban township provided accommodation and a job are available for them. This mobility has aided the industry in its efforts to encourage blacks to regard coal mining as a career and to train for higher skills and is instrumental in the objective of moving away from rural, single, unskilled migrants to a more-local, stable workforce based on urban, married, skilled workers.

Over the study period of 1950-80 the vast majority of blacks becoming unemployed in the coal mining industry could not claim unemployment benefits in terms of the 1966 Unemployment Insurance Act. "Blacks employed on any gold or coal mine and who are provided by their employers with both food and quarters" were excluded from the definition of a "contributor". This provision was only repealed in 1981. However, migrants from independent states, such as Lesotho, Mozambique and Swaziland, are excluded under a separate provision of the Act, which now includes Transkei, Bophuthatswana, Venda and Ciskei. Section 10 blacks living in their own accommodation in a township are classed as "contributors", but their numbers have traditionally been small.

White workers receive subsidised housing, either on the mine or in a nearby town, with free water and electricity. Assistance with removal expenses is often available. Some groups provide a house-purchase scheme for key workers whereby they are enabled to own their own home under favourable conditions. Primary-school facilities are available on some mines. For those which do not have a mine school, free transport is provided for children to school in town. Workers who live in town are also provided with free bus transport to and from the colliery. All white workers enjoy the protection of a pension and medical-aid scheme. The Rand Mutual provides workmens' compensation and some groups provide free life-insurance cover. Social, sporting, and recreational facilities are available on all collieries. The social club is a favourite meeting place of white workers and their families, and different mines variously boast the following sporting facilities: golf, swimming, angling, bowls, tennis, squash, badminton, soccer, cricket, and rugby.

In addition to production and incentive bonuses, an annual bonus of one month's salary is offered by different collieries. Generous leave allowances are also available. Top employees receive a company car. White workers are classed as "contributors" to the Unemployment Insurance Fund if they earn under the statutory ceiling. If not, they are expected to make their own private arrangements for unemployment insurance cover.

REFERENCES

(1) Chamber of Mines, 1978 Year-End Review, December Report, "The Human Factor", page 9.

(2) Breytenbach, W.J., Migratory Labour Arrangements in Southern Africa, Africa Institute of South Africa, 1979.

(3) Jones, R.A., and Griffiths, H.R., Labour Legislation in South Africa, McGraw-Hill, Johannesburg, 1980, page 1.

(4) Department of Mines, Annual Report, Report of the Government Mining Engineer, published annually; and Mining Statistics, tables 1 and 2, published annually.

(5) Graham, A.C., The Coals of the Witbank District (Transvaal), Transvaal Chamber of Mines, Collieries Committee, 1931, page 69.

(6) Department of Mines, Report of the Commission of Inquiry Into the Possible Introduction of a Five-Day Working Week in the Mining Industry of the Republic of South Africa, (Franzsen Commission), Government Printer, Pretoria, 1977, page 9.

(7) Griffiths, H.R., and Jones, R.A., South African Labour Economics, McGraw-Hill, Johannesburg, 1980, chapter 1.

(8) Hall, P.E., "The Coal Industry in the Union of South Africa", South African Journal of Economics, September 1948, page 242.

(9) Wilson, F., Labour in the South African Gold Mines, 1911-69, Cambridge University Press, 1972; and Migrant Labour in South Africa, Report to the South African Council of Churches, published jointly by SACC and SPRO-CAS, Johannesburg, 1972.

(10) Merle Lipton, "Men of Two Worlds: Migrant Labour in South Africa", Optima, Volume 29, Number 2/3, November 28 1980, pages 72-201.

(11) Graham, A.C., The Coals of the Witbank District (Transvaal), op cit., page 74.

(12) ibid.

(13) Hall, P.E., "The Coal Industry in the Union of South Africa", op cit., page 242.

(14) Chamber of Mines, "A Comfortable Environment", 1979 Year-End Review, page 14; and "Accommodation: the Transformation Continues", 1980 Year-End Review, page 16.

(15) see Supplement to the Financial Mail, September 28 1979, "Looking at Rietspruit".

(16) see "Housing Black Miners: One Great Leap Forward", Star, March 22 1979.

(17) Graham, A.C., The Coals of the Witbank District (Transvaal), op cit., pages 75-6.

(18) Fraser, A.E., "Coal Mining in South Africa", Transactions of the Seventh Commonweath Mining and Metallurgical Congress, 1961, volume 2, page 907.

(19) Financial Mail, Special Survey, "Lean Years for Coal", December 13 1968.

(20) Chamber of Mines, "Brigadesmen - Mining's Quiet Heroes", in The Chamber of Mines of South Africa: Ninety Years of Achievement, July 1979, page 45.

(21) see "Safety - Solid Foundation for Steady Productivity", Coal, Gold and Base Minerals of Southern Africa, April 1982.

(22) see "Medical Services on the Mines", Chamber of Mines, Mining Survey, S.A.'s New Path of Progress, No. 93, No. 2 of 1979, pages 15-24.

(23) Chamber of Mines, "The Rand Mutual - a Pioneer of Workmens' Compensation", in The Chamber of Mines of South Africa: Ninety Years of Achievement, op cit.

(24) Mr. Rensburg, Welfare Officer, Mineworkers' Union, in communication with H.R. Griffiths, August 28 1979, as reported in Griffiths, H.R., and Jones, R.A., South African Labour Economics, op cit., page 240.

(25) see "Petrol Blasts as Paulus Slates Report", Citizen, May 17 1979.

(26) White Paper on the Committee on the Decentralisation of Industries, Government Printer, Pretoria, May 1971, paragraph 18.

(27) Jones, R.A., and Griffiths, H.R., Labour Legislation in South Africa, op cit., pages 61-3.

(28) Griffiths, H.R., and Jones, R.A., South African Labour Economics, op cit., page 190.

(29) Witbank Colliery Limited, Annual Report,1982, page 6.

(30) Spandau, A., "Mechanisation and Labour Policies on South African Mines", South African Journal of Economics, June 1980, page 173.

(31) Spandau, A., "Mechanisation and Labour Policies on South African Mines", Labour Research Unit, University of the Witwatersrand, Report No. 22, September 1979, page 25.

(32) Thompson, G.C., and Henderson, P.G., "The Organisation and Management of a Large Mechanised Colliery", Journal of the South African Institute of Mining and Metallurgy, March 1975, pages 200-3.

(33) Spandau, A., Labour Research Unit, op cit., pages 26-7.

(34) ibid.

(35) Chamber of Mines, "Job Patterns Change as Coal Mining Industry Grows", in Mining Survey, Number 1/2, 1981, pages 23-5.

(36) Anglo-American Corporation and General Mining, A Career in Coal Mining, no date, Penrose Press.

(37) Chamber of Mines, Careers in Mining, P.R.D. Series No. 231, 1979, pages 8-11.

(38) see "How Amcoal Utilises its People", Mining Week, March 5 1980.

(39) see "No Race Barriers in New Training Centre", Star, August 17 1981.

(40) Witbank Colliery Limited, Annual Report, 1981, page 9.

(41) see "R15 million for Training Facilities", Coal, Gold and Base Minerals of Southern Africa, October 1981.

(42) see "Opencast Colliery School", Coal, Gold and Base Minerals of Southern Africa, June 1981.

(43) Oppenheimer, H., Anglo-American Corporation of South Africa Ltd., Chairman's Statement, 1979, July 26 1979, pages 7-8.

(44) see Coal, Gold and Base Minerals of Southern Africa, August 1979, page 15.

(45) Griffiths, H.R., and Jones, R.A., South African Labour Economics, op cit., pages 152-169.

(46) Jones, R.A., Collective Bargaining in South Africa, Macmillan, 1982, pages 26-33.

(47) see "Black Mining Unions", Financial Mail Supplement, October 22 1982.

(48) see "A New Force Emerges", Financial Mail, December 3 1982.

(49) Department of Manpower Utilisation, Report of the Commission of Inquiry into Labour Legislation, Part 6, November 1980, Government Printer, Pretoria.

(50) Financial Mail, Special Survey, August 15 1980, "Coal-Warming Up".

(51) see "Wiehahn and the Mining Industry", Coal, Gold and Base Minerals of Southern Africa, May 1979.

(52) ibid.

(53) see Coal, Gold and Base Minerals of Southern Africa, August 1979, pages 14-17; and Chamber of Mines, Annual Report, 1978, Presidential Address, page 12.

(54) Brits, R.N., and Reese, K., Business Economics for South African Students, McGraw-Hill, 1978, page 196.

(55) see Spandau, A., Economic Boycott Against South Africa, Juta, 1979, chapter 3; and Griffiths, H.R., and Jones, R.A., South African Labour Economics, op cit., chapter 5.

(56) Graham, A.C., The Coals of the Witbank District (Transvaal), op cit., page 70.

(57) Hall, P.E., "The Coal Industry in the Union of South Africa", op cit., page 244.

(58) Financial Mail, Special Survey, "Lean Years for Coal", December 13 1968.

(59) Chamber of Mines, Mining Survey, Oil Crisis Prompts New Look at S.A., No. 92, No. 1 of 1979, page 10.

(60) Mining Survey Supplement to the Financial Mail, "Coal-Warming Up", August 15 1980.

(61) see "Mine Wages: The Gap Narrows", Financial Mail, April 4 1980.

(62) Anglo-American Corporation, Annual Report, 1979.

(63) see "Black Pay Rises on Mines for Racial Peace", Star, July 16 1981.

(64) see Anglo-American, Annual Statement on Wages Policy, 1981, as printed in Financial Mail, July 18 1981.

(65) Chamber of Mines, Mining Survey, Oil Crisis Prompts New Look at S.A., op cit., page 11.

(66) Spandau, A., "A Note on Black Employment Conditions on South African Mines", South African Journal of Economics, June 1980.

(67) Spandau, A., Labour Research Unit Report, op cit., page 18.

(68) see, for instance, Chamber of Mines, "Huge Improvements for Black Mineworkers", Mining Survey, No. 92, No.1 of 1979, op cit., pages 5-16; "The Best Cared-For Workers in Africa", 1978 Year-End Review; and 1980 Year-End Review, pages 4-16.

(54) Brits, R.N., and Reese, K., Business Economics for South African Students, McGraw-Hill, 1978, page 196.

(55) see Spandau, A., Economic Boycott Against South Africa, Juta, 1979, chapter 3; and Griffiths, H.R., and Jones, R.A., South African Labour Economics, op cit., chapter 5.

(56) Graham, A.C., The Coals of the Witbank District (Transvaal), op cit., page 70.

(57) Hall, P.E., "The Coal Industry in the Union of South Africa", op cit., page 244.

(58) Financial Mail, Special Survey, "Lean Years for Coal", December 13 1968.

(59) Chamber of Mines, Mining Survey, Oil Crisis Prompts New Look at S.A., No. 92, No. 1 of 1979, page 10.

(60) Mining Survey Supplement to the Financial Mail, "Coal-Warming Up", August 15 1980.

(61) see "Mine Wages: The Gap Narrows", Financial Mail, April 4 1980.

(62) Anglo-American Corporation, Annual Report, 1979.

(63) see "Black Pay Rises on Mines for Racial Peace", Star, July 16 1981.

(64) see Anglo-American, Annual Statement on Wages Policy, 1981, as printed in Financial Mail, July 18 1981.

(65) Chamber of Mines, Mining Survey, Oil Crisis Prompts New Look at S.A., op cit., page 11.

(66) Spandau, A., "A Note on Black Employment Conditions on South African Mines", South African Journal of Economics, June 1980.

(67) Spandau, A., Labour Research Unit Report, op cit., page 18.

(68) see, for instance, Chamber of Mines, "Huge Improvements for Black Mineworkers", Mining Survey, No. 92, No.1 of 1979, op cit., pages 5-16; "The Best Cared-For Workers in Africa", 1978 Year-End Review; and 1980 Year-End Review, pages 4-16.

(54) Brits, R.N., and Reese, K., Business Economics for South African Students, McGraw-Hill, 1978, page 196.

(55) see Spandau, A., Economic Boycott Against South Africa, Juta, 1979, chapter 3; and Griffiths, H.R., and Jones, R.A., South African Labour Economics, op cit., chapter 5.

(56) Graham, A.C., The Coals of the Witbank District (Transvaal), op cit., page 70.

(57) Hall, P.E., "The Coal Industry in the Union of South Africa", op cit., page 244.

(58) Financial Mail, Special Survey, "Lean Years for Coal", December 13 1968.

(59) Chamber of Mines, Mining Survey, Oil Crisis Prompts New Look at S.A., No. 92, No. 1 of 1979, page 10.

(60) Mining Survey Supplement to the Financial Mail, "Coal-Warming Up", August 15 1980.

(61) see "Mine Wages: The Gap Narrows", Financial Mail, April 4 1980.

(62) Anglo-American Corporation, Annual Report, 1979.

(63) see "Black Pay Rises on Mines for Racial Peace", Star, July 16 1981.

(64) see Anglo-American, Annual Statement on Wages Policy, 1981, as printed in Financial Mail, July 18 1981.

(65) Chamber of Mines, Mining Survey, Oil Crisis Prompts New Look at S.A., op cit., page 11.

(66) Spandau, A., "A Note on Black Employment Conditions on South African Mines", South African Journal of Economics, June 1980.

(67) Spandau, A., Labour Research Unit Report, op cit., page 18.

(68) see, for instance, Chamber of Mines, "Huge Improvements for Black Mineworkers", Mining Survey, No. 92, No.1 of 1979, op cit., pages 5-16; "The Best Cared-For Workers in Africa", 1978 Year-End Review; and 1980 Year-End Review, pages 4-16.

PART II

THE CONCEPT AND MEASUREMENT OF PRODUCTIVITY

CHAPTER SIX

THE CONCEPT AND THEORY OF PRODUCTIVITY CHANGE

Productivity is a concept which is subject to much discussion. Unfortunately it is surrounded by considerable confusion and, consequently, is frequently misunderstood. It is common for the same term to be used when referring to different things and for this reason a close examination of the meaning of productivity is desirable.

Definition of Productivity

On a generalised, macro-economic level, there are certain categories of economists who regard labour to be, if not the only true factor of production, then, at least, the driving force amongst the factors. Without labour, raw materials, land and capital would remain passive elements. They are activated, co-ordinated, and placed into productive use only through the intervention of man, his efforts, and his ingenuity. The object steadily pursued by man is to produce more, to achieve an even greater output from his labour (1). This can only be achieved by utilising more thoughtfully the otherwise passive factors which are available to him. In this context the International Labour Organisation definition is relevant (2);

> "higher productivity means, in the most general terms, an increase in the ratio of the output of wealth (goods and services) to the corresponding input of labour, an increase in the production of wealth per unit of labour".

This definition does not in any way imply harder physical work. Productivity is not necessarily measured in terms of effort and "sweat". Rather it also involves elements of "mental" advances, or the notion of "getting smarter" in common parlance. The concept of productivity being advanced by labour means obtaining better results and increasing available resources by using a larger number of modern machines, by standardising and simplifying, by planning production, and by providing workers with incentives to obtain better results, amongst other factors. Machinery is regarded as the hallmark of progress. An indicator of man's ingenuity is the advantage he is taking of expanding technical facilities (3).

Whilst this definitional approach may be relevant on the generalised level, it is of little use when measuring productivity at the level of the individual sector, industry, or firm. On this micro level, if an enterprise produces more goods per man simply as a result of replacing labour by machinery, this can hardly be regarded as productivity improvement for that undertaking. Rather it is merely incorporating into its production structure the end result of an outside-agent's ingenuity, in this case, the capital-goods industry. Thus, on a micro level we are more concerned with what the ILO calls "real productivity" (4), that is, the relationship between output and the combined use of all separate inputs.

By concentrating on the concept of output per unit of all inputs, the economic approach to productivity can now be refined as,

> "the ratio of what is produced to what is required to produce it" (5).

Any input can be used in the denominator of the productivity ratio so that there are as many productivity measures as the number of classes of inputs we choose to distinguish. Such measures are known as partial productivity ratios, and although they are interesting when correctly analysed in context, they can be a dangerous element in that we cannot divorce changes in the productivity of one factor from that of other factors. Partial ratios should be analysed in this spirit to avoid the risk of attributing to them an undeserved significance. A conceptually better approach is to ignore the usual divisions between land, labour, capital, raw materials and entrepreneurial ability, and instead to regard them as a combined agglomerate to be called "inputs". The concept of productivity as the ratio of output to input is then inalienable. The practical reality of statistically measuring this ratio, however, involves numer and complex problems which are examined in chapters 7 and 8.

Significance of Productivity

Problems of definition and measurement of productivity are one aspect. These are, however, quite different from problems of interpretation. It is important not to confuse issues of measurement with issues of interpretation. Failure to do so has resulted in fruitless discussion and has been responsible for much of the confusion which has arisen. Salter recognises the multitude of interpretations to which productivity is subjected;

> "to some it measures the personal efficiency of labour; to others, it is the output derived from a composite bundle of resources; to the more philosophic it is almost synonymous with welfare; and in one extreme case it has been identified with time" (6).

In this aspect Salter is associated with a school of thought expressing dissatisfaction with the conventional interpretation of productivity measures. This conventional approach is well expressed in the dictionary definitions of productivity revolving around the notion of "efficiency in industrial production" (7), although the word "efficiency" in this context is not the same as the economist's interpretation as a ratio of an actual level to an optimally-efficient level. Rather, Fabricant indicates the proper perspective of the word by reformulating the dictionary definition, and interpreting productivity as,

> "a measure of the efficiency with which resources are converted into the commodities and services that men want" (8).

In this context, Fabricant is able to interpret the significance of higher productivity as,

> "a means to better levels of economic well-being and greater national strength ... a major source of the increment in income over which men bargain and sometimes quarrel ... (it) affects costs, prices, profits, output, employment, and investment and thus plays a part in business fluctuations, in inflation, and in the rise and decline of industries ... Indeed in one way or another productivity enters virtually every broad economic problem, whatever current form or new name the problem takes - industrialisation, or research and development, or automation, or tax reform, or cost-price squeeze, or improvement factor, or wage inflation, or foreign

fruitless discussion and has been responsible for much of the confusion which has arisen. Salter recognises the multitude of interpretations to which productivity is subjected;

> "to some it measures the personal efficiency of labour; to others, it is the output derived from a composite bundle of resources; to the more philosophic it is almost synonymous with welfare; and in one extreme case it has been identified with time" (6).

In this aspect Salter is associated with a school of thought expressing dissatisfaction with the conventional interpretation of productivity measures. This conventional approach is well expressed in the dictionary definitions of productivity revolving around the notion of "efficiency in industrial production" (7), although the word "efficiency" in this context is not the same as the economist's interpretation as a ratio of an actual level to an optimally-efficient level. Rather, Fabricant indicates the proper perspective of the word by reformulating the dictionary definition, and interpreting productivity as,

> "a measure of the efficiency with which resources are converted into the commodities and services that men want" (8).

In this context, Fabricant is able to interpret the significance of higher productivity as,

> "a means to better levels of economic well-being and greater national strength ... a major source of the increment in income over which men bargain and sometimes quarrel ... (it) affects costs, prices, profits, output, employment, and investment and thus plays a part in business fluctuations, in inflation, and in the rise and decline of industries ... Indeed in one way or another productivity enters virtually every broad economic problem, whatever current form or new name the problem takes - industrialisation, or research and development, or automation, or tax reform, or cost-price squeeze, or improvement factor, or wage inflation, or foreign

dollar shortage" (9).

The importance of changes in productivity levels as a major influence on wide-ranging social and economic considerations is also emphasized by Eilon and Soesan who pinpoint "rapid growth, higher standards of living, improvements in the balance of payments, inflation control, and leisure". Indeed they are prepared to generalise even further to state that "the welfare of individual enterprises, and even of entire national economies, is widely regarded as dependent on their comparative productivity" (10).

Kendrick adopts virtually the same theme, considering the story of productivity to be "at heart the record of man's efforts to raise himself from poverty". Historical productivity advances have meant,

> "not only a large gain in the plane of living, but an increase in the quality and variety of goods and an expansion of leisure time, while increasing provision was made for future growth and for national security" (11).

Salter, in expressing dissatisfaction with the conventional approach, considers the interpretation of productivity measures to be an extremely difficult exercise because it raises a host of complex and varied problems (12). It essentially involves an appreciation of the workings of a moving economic system. All the dynamic forces of economic life lie behind productivity, e.g. technical progress, accumulation, enterprise, and the institutional pattern of society. The movement from productivity measurement to interpretive analysis demands an understanding of the relationships between productivity and these dynamic forces for change. In his classic book Salter attempted to analyse one part of this problem: the relationship between productivity and technical change and to fit this into a context of prices and costs, via conventional concepts rooted in economic theory. This he believed to be more meaningful than "some more superficially-appealing concept which has pretensions to measuring economic efficiency".

Gold (13) expresses similar reservations to Salter in criticising the conventional interpretation of productivity. This, he claims, has given rise to a "mythology" of four elements, namely,

- that productivity measures reflect changes in the "efficiency" of production,

- that output per manhour reasonably measures productivity changes,

- that increases in output per manhour are desirable because they decrease unit costs,

- that increases in output per manhour warrant parallel increases in wages per manhour.

Yet Gold shows that none of these contentions is sustainable either on theoretical or empirical grounds.

Early productivity analysis concentrated on agriculture and simple manufacturing processes but shifted without change from these relatively primitive operations to highly-complex activity systems, and from the context of engineering measurements of physical relationships to managerial appraisals of economic relationships. However, the necessary far-reaching readjustments in purposes, concepts and methods were largely overlooked. As a result the concept of productivity came to reflect a mixture of "input creativity" and "conversion efficiency", and in so doing encouraged a myopic concentration on one component of a complex of relationships. According to Gold, merely juxtaposing the comparative magnitudes of specified inputs and outputs reveals nothing more than the level of, and changes in, the given ratio. Interpretive analysis only becomes possible when the variables are derived from an analytical framework encompassing all the system's inputs and outputs and providing a theory of how it functions. This then permits working backward from specified objectives to determine which variables should be studied, how they should relate to one another and what measurements should be used to facilitate effective use of the findings.

Such a re-direction of productivity analysis necessitates the following;

- clarification of the nature of productivity adjustments,

- development of more-effective measures of productivity changes,

- exploration of the sources of significant productivity changes,

- tracing the successive links between productivity and costs, prices, and profitability,

- integrating the above into a managerial control system designed to enable mangement to (i) appraise alternative means of changing productivity, (ii) appraise alternatives in the application of such innovations, (iii) determine the effects of past and prospective innovations.

Gold emphasizes that productivity adjustments can only be appraised within a specified framework encompassing all the system's input-output flows and specifying the criteria in terms of which alternatives are considered and performance evaluated.

These more-sophisticated approaches of Salter and Gold have by no means acquired wide acceptance. The conventional interpretations are still generally accepted on the grounds that simplicity is a virtue. The Salter-Gold approach can become unnecessarily cumbersome. Solow expresses the view of the conventionalists,

> "the economist really need not know at all what it feels like to be inside a steel plant ... he quantifies technological change by making measurements of ... output per unit of this, or input per unit of that" (14).

The Theory of Productivity Change

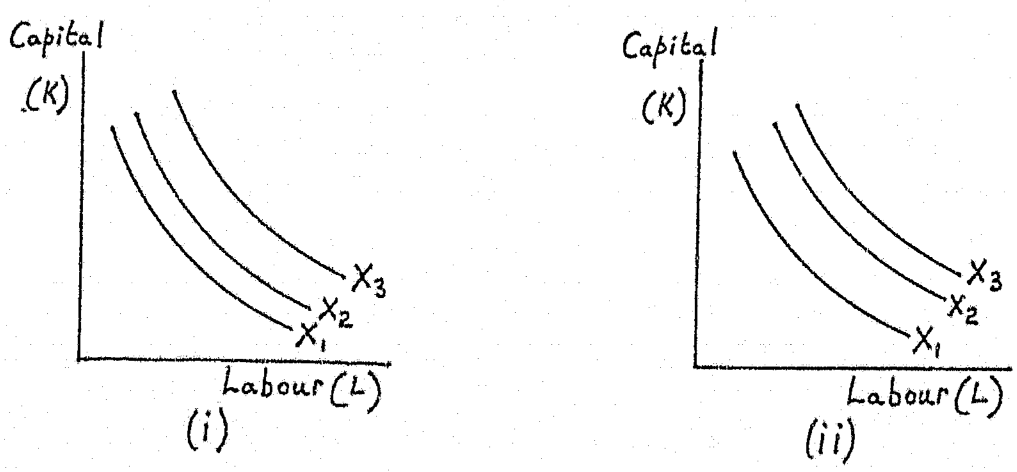
The introduction of any innovation into the production process allows a firm to produce an unchanging level of output by utilising smaller levels of input or to produce an increased level of output by utilising an unchanged level of inputs. These processes can be described as "input saving" or "output expanding". Diagramatically both processes are represented by a shift towards the origin of the firm's entire isoquant map. Such a map depicts the transformation of inputs into output and the way inputs co-operate in varying proportions to produce given levels of output. These relationships are determined by the ruling technology which is embedded in the isoquant map. When technology changes the impact on the isoquants can be reflected in four ways which determine the nature of the shifts towards the origin (15).

These are briefly discussed below;

(a) The efficiency of a technology: this affects the relationship between inputs and output but not between the inputs. It depicts the efficiency with which inputs are transformed into output and is, therefore, reflected as a scale transformation.

(b) Technologically-determined economies of scale: these are reflected when a given proportional increase in inputs generates a larger proportional increase in output. The scale of operations of the enterprise determines the magnitude of the benefit to be derived from such economies. Diagram (i) below represents decreasing returns to scale whilst diagram (ii) shows increasing returns to scale at the same input levels following a technological change.

FIGURE 6 1 TECHNOLOGICALLY-DETERMINED CHANGES IN RETURNS TO SCALE



(c) Capital intensity: this aspect relates the quantity of capital to the quantity of labour used in the production process. Different firms require different capital-labour ratios due to the particular technology of the enterprise, i.e. capital intensity is part of the structure of the ruling technology.

(d) Capital-labour substitution: this is measured by the elasticity of substitution, ε , or the percentage change in the capital-labour ratio in response to a given percentage change in the marginal rate of substitution of labour for capital (or the proportional change in the relative factor-price ratio). A high value of ε signifies easy substitution between factors indicating that they are technologically similar, and vice versa for a low ε.

Technological change in cases (a) and (b) above is neutral in the sense that it is depicted by parallel inward shifts of the isoquants, whereas cases (c) and (d) represent non-neutral changes depicted by twisting inward shifts of the isoquants. These terms require more explanation.

Neutral and Non-Neutral Shifts

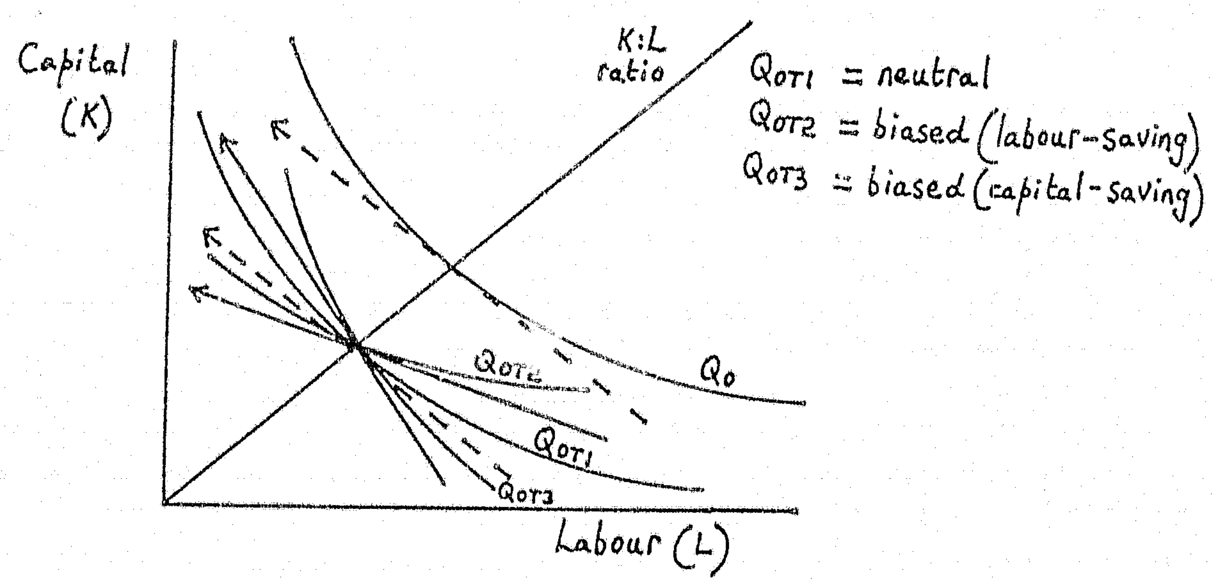
The nature of such shifts helps to determine the labour-capital mix over time. Changing factor productivities will lead to changing optimal factor proportions. Accordingly, the time path of technology is an important element in the economic-growth process. Such technology shifts may be either neutral or non-neutral (biased) in nature. Gehrig (16) states,

> "technical progress is called neutral with respect to certain economic variables if it does not affect these variables or functional relationships between them. Each type of these relationships then forms a certain type of neutral technical progress."

Hicks (17) defined a neutral shift as occurring when the marginal rate of technical substitution of labour for capital remains unchanged at the original capital-labour ratio prevailing before the innovation occurred. Hence, it is characterised by parallel shifts of the isoquants towards the origin. The innovation raises the marginal products of all factors by the same proportion thus leaving relative marginal products unchanged. There is a uniform improvement in the quality of all factors, so consequently this does not encourage factor substitution. Factor proportions remain unchanged. Thus a neutral change neither saves nor uses any factor.

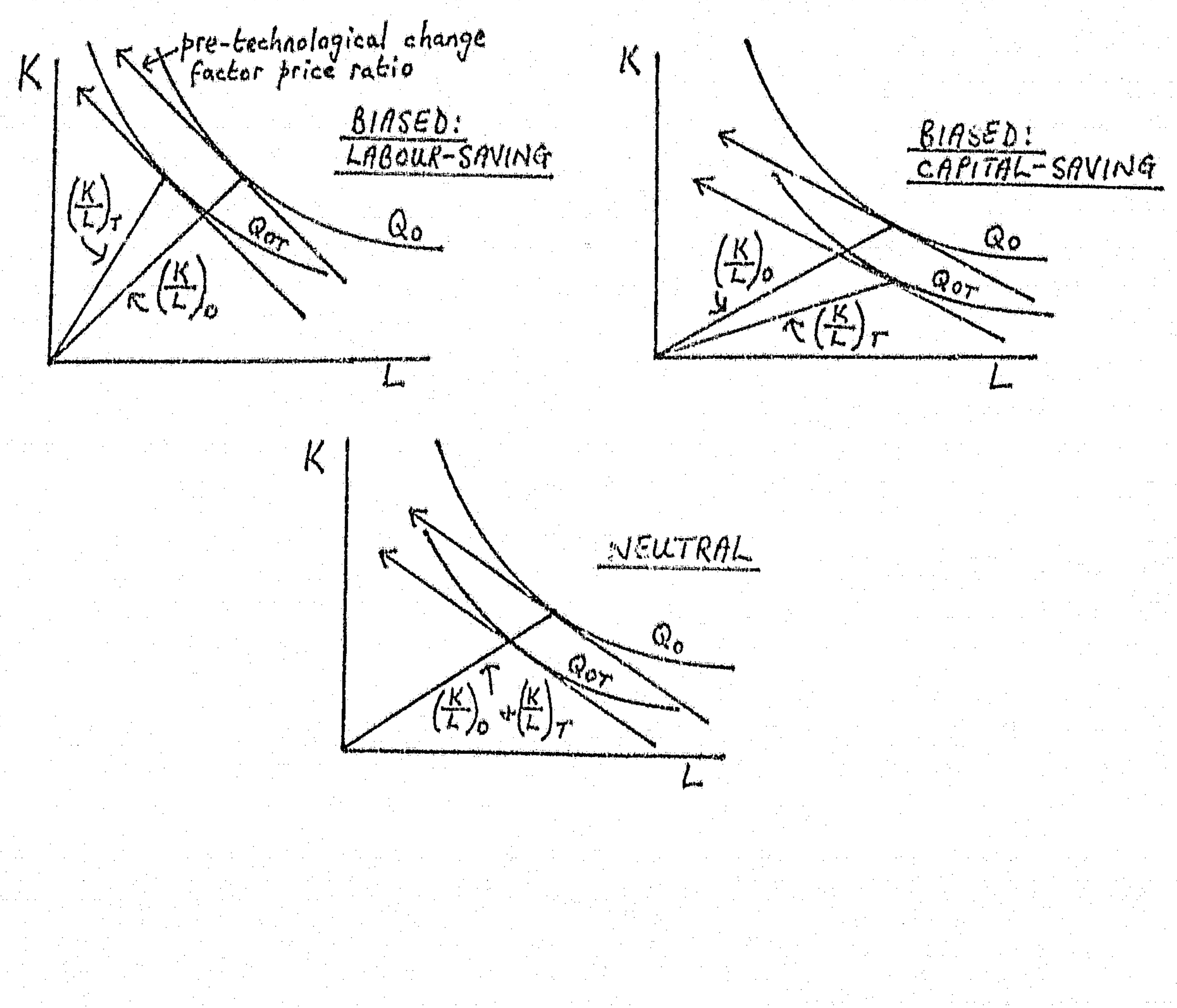
On the other hand, a biased shift occurs if the $MRTS_{LK}$ changes along the constant K:L ratio. This is characterised by non-parallel shifts of the isoquants so that they progressively "twist" in a certain direction. Non-neutral innovations cause a non-proportional change in the marginal productivities of the factors at the given K:L ratio, encouraging factor substitution to take place. If the innovation increases MP_K more than MP_L then the firm is prompted to employ more capital relative to labour. This results in a labour-saving shift of the production function, giving a shallower slope of the isoquant at the original K:L ratio (i.e. the $MRTS_{LK}$ diminishes) implying that because of the increased MP_K it now takes a proportionately larger amount of labour to compensate for a unit fall in capital employment to maintain output unchanged along a given isoquant. The opposite argument applies in the case of a capital-saving shift of the production function. Here the isoquants become steeper at the original K:L ratio indicating that because the MP_L has risen relative to the MP_K it takes a proportionately smaller increase in labour to compensate for a unit fall in capital to maintain output unchanged along a given isoquant. These concepts are illustrated in the diagram below.

FIGURE 6.2 HICKS NEUTRAL AND BIASED TECHNOLOGICAL PROGRESS



The Hicksian criterion of classifying biasedness according to the degree of change of the $MRTS_{LK}$ at a given K:L ratio, contrasts with the Salter criterion which classifies biasedness according to the degree of change in the K:L ratio for a given $MRTS_{LK}$ (18). An unchanging $MRTS_{LK}$ as technological progress occurs implies that the ratio of factor prices is maintained unchanged at its pre-technological-change level. Salter's criterion for emphasising this condition is that the successive combinations of capital and labour chosen by the firm and their transformation into output depends upon both the advance of technological knowledge and the ratio of factor prices. Only if the technological advances are "economic" will they be incorporated into production, and this depends upon factor prices. Thus, many technological advances are not immediately put into use if the factor prices are uneconomic. There is a delay in their incorporation into production. After many years lag, factor prices may change and make economic the introduction of the "old" technology. Accordingly, a situation arises where changing prices are introducing a flow of new techniques although these "new" techniques are "old". It may be the position that at the moment new technological advances are stagnant. As a result the process of "best-practice" productivity implies that the actual incorporation of technological advance depends upon the interaction of technological advances and factor prices. It is difficult to distinguish how far new techniques are attributable to new knowledge or to changing factor prices. Analytically the distinction can be made by holding factor prices constant and observing the sequence of best-practice productivity as the isoquants shift, i.e. it involves asking what changes in technique would take place if relative factor prices were constant. Thus, for Salter, the labour-or capital-saving biases of technological progress are measured by the relative change in capital per labour unit when relative factor prices are constant. Graphically, it measures the extent to which points on each isoquant with the same slope move closer to one axis than another. These concepts are illustrated below.

FIGURE 6.3 SALTER NEUTRAL AND BIASED TECHNOLOGICAL PROGRESS



It will be seen that since the Hicks and Salter definitions are the reverse of each other, they both imply the same definition of neutral technological progress. The two definitions differ in the measure of the extent to which technological progress is labour or capital saving (and then the difference may not be great).

The difference in emphasis between Hicks and Salter lies in their objectives. Hicks' definition has been framed for aggregative analysis, particularly the question of relative shares. Thus, it is appropriate to assume factor supplies constant and classify advances according to their effects on marginal products. But this definition cannot provide any guide to the effects of technological progress on productivity at the micro level, for by assuming fixed factor proportions they automatically assume that the productivity of all factors changes in equal proportions. When we are concerned with productivity, price, and cost movements at the micro level, where relative marginal products are determined by factor prices external to the industry, it is appropriate to reverse the definition by assuming marginal products constant and examine the effects on factor requirements as in the Salter definition.

Harrod (19) approached the question of neutral shifts from a different standpoint than that of Hicks. He had a special interest in the capital-output ratio stemming from his concern with growth theory and defines a neutral shift as one that leaves the capital-output ratio unchanged at a constant rate of interest. Mrs. Robinson (20) also employs the same definition. Harrod neutrality, therefore, involves a technological shift which leaves the MP_K unchanged at a given capital-output ratio and hence can be characterised as a uniform improvement in the efficiency or quality of the labour force. This can be due to increased education of the labour force but is assumed to apply to all workers regardless of the length of time they have been employed. Harrod-neutral shifts are said to be labour-augmenting. In the words of Hahn and Matthews (21),

> "population growth causes there to be two men where there was previously one; Harrod-neutral technological progress causes one man to be able to do twice what he could have done previously".

Thus, as technological progress continues, the K:L ratio must increase, i.e. there would be fewer workers per unit of capital. The output-labour ratio also reflects the same characteristic. It is intuitive that output per man will increase as labour-augmenting shifts occur. However, the capital-output ratio remains undisturbed. This is the reason why Harrod specified labour-augmenting shifts in his growth model for it allowed him to ignore the impact of technological progress on the capital-output ratio.

The seeming dissimilarity between the Hicks and Harrod definitions of neutrality led to an interesting debate in the literature with the main protagonists being Kennedy (22) (24) (25) and Harrod (23). The conclusion emerged that the two definitions were equivalent provided they were applied to a single improvement or alternatively at the economy level if the elasticity of substitution is unity.

Solow employs yet another approach in defining a neutral shift as one which leaves the output-labour ratio unchanged at a constant wage rate. Solow neutrality, therefore, involves a tecnnological shift which leaves the MP_L unchanged at a given output-labour ratio and hence can be characterised as a uniform improvement in the efficiency of the existing capital stock regardless of its age. Solow-neutral shifts are said to be capital-augmenting.

However, the introduction into the literature of neutrality definitions does not constitute a theory of neutral technological progress. Such a theory was only developed at a later stage by Sato and Beckmann (26) (27). Gehrig (28) states that the usual procedure in defining concepts of neutral technological progress is to formulate invariant relationships between economic variables - such as the interest (wage) rate, the elasticity of substitution, etc. - which are either empirically tested or seem to be significant. He considers seven neutrality concepts:

(1) HICKS: $MRTS_{LK}$ remains constant at a constant K:L ratio (output-augmenting).

(2) HARROD: interest rate is constant at a constant output-capital ratio (labour-augmenting).

(3) SOLOW: wage rate is constant at a constant output-labour ratio (capital-augmenting).

Thus, as technological progress continues, the K:L ratio must increase, i.e. there would be fewer workers per unit of capital. The output-labour ratio also reflects the same characteristic. It is intuitive that output per man will increase as labour-augmenting shifts occur. However, the capital-output ratio remains undisturbed. This is the reason why Harrod specified labour-augmenting shifts in his growth model for it allowed him to ignore the impact of technological progress on the capital-output ratio.

The seeming dissimilarity between the Hicks and Harrod definitions of neutrality led to an interesting debate in the literature with the main protagonists being Kennedy (22) (24) (25) and Harrod (23). The conclusion emerged that the two definitions were equivalent provided they were applied to a single improvement or alternatively at the economy level if the elasticity of substitution is unity.

Solow employs yet another approach in defining a neutral shift as one which leaves the output-labour ratio unchanged at a constant wage rate. Solow neutrality, therefore, involves a tecnnological shift which leaves the MP_L unchanged at a given output-labour ratio and hence can be characterised as a uniform improvement in the efficiency of the existing capital stock regardless of its age. Solow-neutral shifts are said to be capital-augmenting.

However, the introduction into the literature of neutrality definitions does not constitute a theory of neutral technological progress. Such a theory was only developed at a later stage by Sato and Beckmann (26) (27). Gehrig (28) states that the usual procedure in defining concepts of neutral technological progress is to formulate invariant relationships between economic variables - such as the interest (wage) rate, the elasticity of substitution, etc. - which are either empirically tested or seem to be significant. He considers seven neutrality concepts:

(1) HICKS: $MRTS_{LK}$ remains constant at a constant K:L ratio (output-augmenting).

(2) HARROD: interest rate is constant at a constant output-capital ratio (labour-augmenting).

(3) SOLOW: wage rate is constant at a constant output-labour ratio (capital-augmenting).

(4) Wage rate is constant at a constant output-capital ratio (labour-combining).

(5) Interest rate is constant at a constant output-labour ratio (capital-combining).

(6) Wage rate is constant at a constant capital-labour ratio (capital-additive).

(7) Interest rate is constant at a constant capital-labour ratio (labour-additive).

The question as to which of these neutrality concepts best approximates reality was tackled by Sato and Beckmann applying regression analysis to USA, Japan, and Germany's private non-farm sector time-series data. On a log-linear basis they found,

	USA	JAPAN	GERMANY
1.	SOLOW	HARROD	SOLOW
2.	HARROD	HICKS	CAPITAL-COMBINING
3.	LABOUR-COMBINING	SOLOW	LABOUR-COMBINING

These results provide some justification for formulating more neutrality concepts than the Hicks, Harrod and Solow definitions.

Disembodied and Embodied Technological Changes

The manner in which technological progress is incorporated into the production process is now in need of explanation. Technological change is called "disembodied" when it arrives in the form of "manna from heaven" (29) or "floats down from the outside" (30) making its benefits freely available to all factors. The previous discussion of neutral shifts implicitly made this assumption. Improvements in the quality of labour and/or capital applied equally to old and new units of the factor(s) concerned, i.e., old or new capital equipment participate equally in technical change, as also do old and new members of the workforce. Labour-augmenting shifts assumed the quality of each worker to improve uniformly over time. Capital-augmenting shifts assumed the quality of each piece of equipment to improve uniformly (31). In other words, improvements in technology can be incorporated into production independently of the purchase of new machinery or the hiring of new workers. If capital and labour are held constant, output will still increase. The pace of investment and the rate of hiring new workers has no influence on the rate at which technique improves.

Stemming from such an analysis is the obvious conclusion that knowledge is growing with time so that the shifts of the production function towards the origin can be represented by including a time variable,

i.e. $X = F(K,L,t)$.

However, using time as an explanatory exogenous variable to measure the growth of knowledge is far from satisfactory and basically a confession of ignorance. What is required is an examination of how knowledge is acquired making possible an endogenous theory of changes in knowledge. The acquisition of knowledge is usually called "learning", reflecting itself in the phenomenon of performance over time. The processes of learning are not mutually agreed upon by psychologists but there is a quantum of agreement that learning is the product of experience - taking place through the attempt to solve a problem. It, therefore, only occurs during activity. Learning through repetition tends to be associated with diminishing returns and to postulate steadily increasing performance implies evolving, rather than repeating, stimulus situations.

Arrow (32) advanced the hypothesis that technological change in general can be ascribed to experience. The very process of production gives rise to problems for which favourable responses are selected over time. This is known as "learning by doing". Brown (33) cited "managerial and/or organisational changes" as reflecting such experience, whilst Alchian (34) was more specific and advanced "job familiarisation, general improvement in co-ordination, shop organisation and engineering liaison and more-efficient subassembly production". Arrow formulated the hypothesis precisely in mathematical terms in order to draw economic implications, in contrast to other researchers who merely observed and documented the role of experience in increasing productivity. Aeronautical engineers were amongst the first to observe this process. Wright (35) calculated that the number of manhours expended in airframe production was a remarkably precise decreasing function of the total number of similar airframes previously produced. Counting from the production inception, the amount of labour required to produce the n^{th} airframe of a given type was proportional to $n^{-1/3}$. This is known as the "learning curve" or the "progress curve". Alchian calls it an 80-per-cent progress curve in airframe production because the cost of the $2n^{th}$ item was calculated to be approximately 80 per cent of the cost of the n^{th} item, i.e., the fortieth plane (say) would involve only 80 per cent of the direct manhours and materials that the twentieth plane did. Such cost-quantity relationships are well documented by Asher (36) for the airframe industry, and by Hirsch (37) and (38) for the production of other items.

Lundberg (39) observed a steadily increasing performance which could only be ascribed to learning from experience at the Horndal iron works in Sweden (the "Horndal effect"). This plant experienced an average increase of 2 per cent per annum in output per manhour over a 15- year period and yet no new investment occurred at the plant during this time.

An alternative way of conceptualising the incorporation of technological progress into the production process is through its embodiment in the latest factors of production. Technological improvements are embodied in newly-trained workers and newly-produced capital, and in order to reap the benefits of these improvements a firm must install new machinery and hire new workers. This is in contrast to disembodied progress which could be enjoyed independent of new units of capital and labour. Under embodied progress it is necessary to distinguish between different vintages, or ages, of capital and labour, since those obtained most recently embody the most advanced technology whereas those obtained in the more-distant past reflect an increasingly inferior state of technology. This again contrasts with disembodied progress which was spread evenly over all ages of capital and labour.

The pioneering models of embodied technological progress belong to Johansen (40), Solow (41), and Salter (42). Their models tended to emphasize the crucial roles of net capital formation and the replacement of old-fashioned equipment by the latest vintages in carrying improvements in technology into practice. Thus, they implicitly specify an aggregate production function of the form,

$$X = F\ (L, K_1, K_2, K_3, \ldots K_T)$$

where X,L and K refer to output, labour and capital respectively and the subscripts on K denote the capital vintage, i.e. K_1 is the current capital stock with an age of one period, K_2 is the current capital stock with an age of two periods, and so on.

An important implication of the embodied technological-progress hypothesis concerns the impact of investment on economic growth (43). Under disembodied progress output will continue to grow even in the absence of capital accumulation and population growth but will not be greatly influenced by an increase in the capital stock. It would, therefore, take unrealistically large expenditures on capital equipment to have a significant effect on economic growth. On the other hand, Solow shows (44) that under embodied progress where capital equipment encompasses the latest technology, relatively small investment expenditures have a significant impact on economic growth. Thus greater emphasis is placed on investment as a vehicle towards ensuring economic growth. These assertions have generally been accepted into economic thinking, apart from a convincing challenge by Brown (45).

The Concept of a Production Function

A production function is an a priori functional relationship between the maximum quantity of output and the inputs required to produce it, and the relationship between the inputs themselves. The function is a datum, given by technological or "extra-economic" considerations. It is the ruling technology which acts as a constraint on decision-making and this is embedded in the function. Of course, with complete knowledge of the production process, it would be possible to specify exactly the functional form relating inputs to output. However, in the real world of imperfect knowledge, production functions have to be specified which only approximate the actual functional form. The two best-known and extensively-utilised production functions are the Cobb-Douglas and Constant-Elasticity-of-Substitution functions.

The former function was introduced and tested empirically in 1928 by C.W. Cobb and P.H. Douglas (46) and for two factor inputs takes the form of

$$X = A\,K^{\sigma}L^{\gamma}$$

where X is output and K and L measure capital and labour services respectively. A is an output scaling co-efficient; σ and γ are distribution co-efficients. A, σ, and γ are constants to be determined empirically. The properties of the Cobb-Douglas function are examined in most applied econometrics textbooks (47). The most important ones are summarised below;

- the marginal product of capital MP_K is represented by the parameter σ multiplied by the average product of capital (i.e. $\sigma. X/K$) ; and similarly the MP_L is represented by $\gamma.X/L$,
- the marginal rate of technical substitution of labour for capital ($MRTS_{LK}$) is equivalent to MP_K/MP_L, i.e. $\sigma/\gamma.L/K$,
- σ is the partial elasticity of output with respect to capital. It denotes the percentage change in output attributable to a percentage change in capital input, with labour constant. γ is the partial elasticity of output with respect to labour, which has an equivalent meaning,
- $\sigma + \gamma$ represents the total percentage change in output for a given percentage change in capital and labour. Thus $\sigma + \gamma$ represents the degree of homogeneity of the function. A sum equal to unity represents constant returns to scale, whilst a sum less than or greater than unity represents decreasing and increasing returns to scale, respectively,
- the elasticity of substitution is unity everywhere along the function,
- σ is the share of capital in total revenue (value of output); similarly γ is the share of labour,
- the level of output is indeterminate under conditions of perfect competition. All that is required is that $\sigma + \gamma = 1$, for all income to be distributed and zero profit incurred. Thus Cobb-Douglas describes the production function only within the range immediately adjacent to the equilibrium output at the trough of the long-run average cost curve.

The Constant-Elasticity-of-Substitution production function (CES) was derived independently by two groups of researchers in the early 1960's. The original formulation for constant returns to scale was derived by Arrow, Chenery, Minhas and Solow (SMAC) (48) and this was followed by the general formulation for any degree of returns to scale, derived by Brown and De Cani (49). The CES general formulation can be represented by

$$X = \kappa[\delta K^{-\rho} + (1 - \delta) L^{-\rho}]^{\frac{-v}{\rho}}$$

where

κ is an output scaling co-efficient denoting the efficiency of a technology (analogous to A in Cobb-Douglas),

δ is a distribution co-efficient indicating the degree to which technology is capital-intensive,

ρ is a substitution parameter,

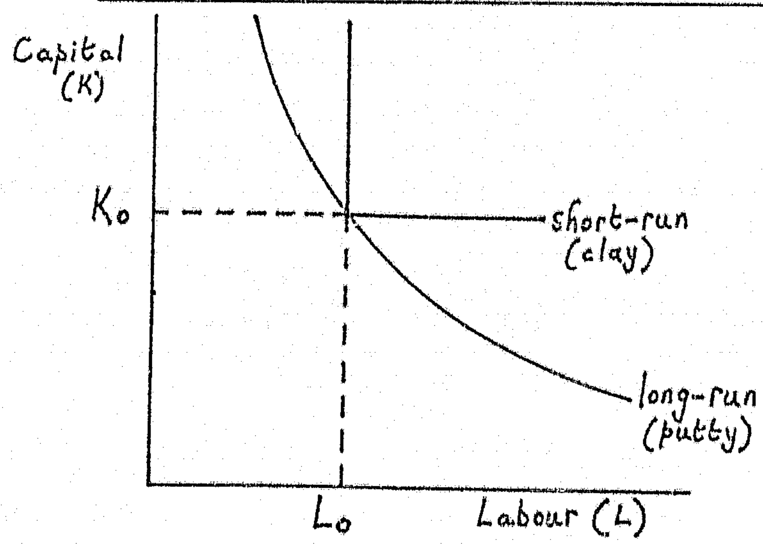
v represents the degree of returns to scale.

The properties of the CES function are again examined in most of the recent applied econometrics textbooks (50), but the more-important ones concern the degree of returns to scale and the elasticity of substitution. $v=1$ represents constant returns to scale (the SMAC formulation), whilst if v is less than or greater than unity this represents decreasing or increasing returns to scale, respectively. In regard to the elasticity of substitution, whereas the Cobb-Douglas function constrained this value to unity (which may or may not be a good approximation), the CES function constrains the value to be a constant, but not necessarily equal to unity. It can be shown that the elasticity of substitution is represented by $\varepsilon = 1/1+\rho$, so that as the value of ρ approaches $0, \infty$, or -1, the value of ε approaches 1, 0 or ∞ respectively. The former is the Cobb-Douglas case; the second is the Marx-Leontief case of zero substitutability; the latter is the case of perfect substitution. The concept of a constant elasticity of substitution assumes only that changes in relative factor inputs and prices do not affect the elasticity. However, its value is determined by the underlying technology and hence does react to changing technological conditions.

This is a convenient point at which to introduce the extra dimension of time. A firm can substitute capital and labour up to the value of the technologically-determined elasticity of substitution. However, once capital is in place it may be very difficult to substitute any further. The capital has been designed to produce an optimum output in co-operation with a certain amount of labour. Equipment is usually designed with strict specifications as to the number, skill levels, and time of worker attendants. Therefore, the elasticity of substitution between capital in place and labour may be very small, and, in many cases, zero. In Hicksian terminology, an enterpreneur, by investing in fixed plant, "gives hostage to fortune" as long as that plant is in existence. However, when plant comes to be renewed his substitution possibilities are once more opened up to him (51). This has been called the "putty-clay" concept, i.e. "putty" refers to the wide range of possibilities available before capital is installed, whereas "clay" refers to the greatly diminished possibilities afterwards.

This dichotomy allows a definition of the "short-run" to be that time period within which ε is determined by the capital in place, and the "long-run" where ε is determined by the spectrum of available technical knowledge (52). Thus, in the Marshallian tradition, the long-run elasticity of substitution (i.e. between labour and the entire spectrum of technical alternatives) is necessarily larger than the short-run elasticity of substitution (i.e. between labour and capital in place). This can then be extended to the concept of lon-run and short-run production functions. The long-run function allows a capital-labour mix to be chosen from a potentially large number of alternatives determined only by the technology of the moment. The short-run function describes the same relationship but under the additional constraint of the capital in place which may be so restrictive as to reduce substitution possibilities to zero, as in the diagram below. On relaxing the constraint of capital in place, the two functions merge into one.

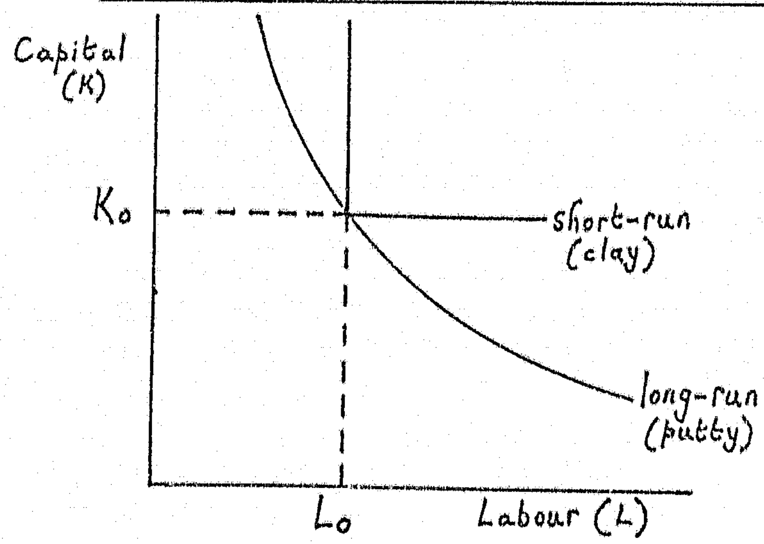
FIGURE 6.4 LONG-AND SHORT-RUN PRODUCTION FUNCTIONS



In defining the difference between the long-and short-run functions it is assumed that no technological change takes place in either period. Such a change alters the spectrum of possibilities which fixed plant, "gives hostage to fortune" as long as that plant is in existence. However, when plant comes to be renewed his substitution possibilities are once more opened up to him (51). This has been called the "putty-clay" concept, i.e. "putty" refers to the wide range of possibilities available before capital is installed, whereas "clay" refers to the greatly diminished possibilities afterwards.

This dichotomy allows a definition of the "short-run" to be that time period within which ε is determined by the capital in place, and the "long-run" where ε is determined by the spectrum of available technical knowledge (52). Thus, in the Marshallian tradition, the long-run elasticity of substitution (i.e. between labour and the entire spectrum of technical alternatives) is necessarily larger than the short-run elasticity of substitution (i.e. between labour and capital in place). This can then be extended to the concept of long-run and short-run production functions. The long-run function allows a capital-labour mix to be chosen from a potentially large number of alternatives determined only by the technology of the moment. The short-run function describes the same relationship but under the additional constraint of the capital in place which may be so restrictive as to reduce substitution possibilities to zero, as in the diagram below. On relaxing the constraint of capital in place, the two functions merge into one.

FIGURE 6.4 LONG-AND SHORT-RUN PRODUCTION FUNCTIONS



In defining the difference between the long-and short-run functions it is assumed that no technological change takes place in either period. Such a change alters the spectrum of possibilities which previously operated as a constraint on substitution possibilities in the long-run, thus changing the long-run function and its relationship with the short-run function. Going one step further, a secular production function can be defined as one where all restraints are suppressed, i.e. technological knowledge may vary. A technological change occurs when the long-run structural relationship between inputs and output changes in a significant way, thus bringing a new "technological epoch". Such an epoch is defined as a period of time within which the short-and long-run production functions are stable. The number of such epochs equals the number of signficant changes in the fund of technical knowledge. Such a discrete change in epochs can be called Schumpeterian (53). On the other hand, small, gradual and persistent changes can be called Usherian (54).

The transition period from one epoch to another becomes blurred as one moves from the micro to the macro level, i.e. from a section of a factory, to the whole factory, to the firm, to the industry, to the sector, to the economy. Although a micro unit (e.g. a factory department) may change technologies in a relatively short period of time at period t, another department may change only at t + 1, another firm may commence the process at t + 2, and so on, so that the larger the aggregate chosen, the longer is the change-over period, lending more credence to the Usherian approach. Thus transition periods between epochs can depend on the rate of diffusion of technological change.

The Diffusion of Technological Innovations

A technological innovation will not have its full impact upon productivity and output until it has been well diffused amongst firms and incorporated into their production process. The rate of imitation is, therefore, a crucial factor, and two questions immediately pose themselves - (i) how soon do other firms imitate an innovation introduced by a leader, and (ii) what factors determine how rapidly they follow? The pioneering work in this field was performed by Mansfield (55). He studied 12 innovations in four industries, examples being the shuttle car and continuous miner in coal mining; by-product coke oven and continuous wide strip mill in iron and steel; high-speed bottle filler in brewing; and the diesel locomotive in railroading.

Concentrating only on large firms he reached two conclusions on the rate of imitation. Firstly, it is generally a slow process. It took 20 years or more for all the major firms to install by-product coke ovens, for example. Only 3 innovations took 10 years or less. Secondly, there is a wide variation. For example, it took about 15 years for half the major pig-iron producers to use the by-product coke oven, but only about 3 years for half the major coal producers to use the continuous miner.

These results immediately raise the question of the factors which govern the rate of imitation from industry to industry. Mansfield constructed a deterministic model using as the dependent variable, λ , defined as the proportion of "hold-outs" (firms not using the innovation) at time t who decide to introduce it by time t + 1. Among the explanatory variables Mansfield tried the following;

- the proportion of firms already using the innovation at time t. The higher this proportion is, the more information and experience that accumulates, hence less risk is attached to the innovation. Competitive pressures will mount and bandwagon effects occur;
- the profitability of the innovation. The greater the profitability the greater the imitation;
- the size of the investment required. The greater the cost the slower the imitation;
- peculiar differences amongst industries. Examples include firms' aversion to risk, the competitiveness of the industry (the more the competition the greater the imitation), trade-union attitudes towards innovation, and financial healthiness of the industry;
- the durability of existing equipment. If capital in place is still performing well, and is not fully written off with a relatively long useful life ahead of it, the rate of imitation will be slower even though the innovation may be more productive than capital in place;
- the rate of expansion of firms. If firms are expanding rapidly the innovation can be incorporated into new plants built to meet the growing market. If there is little or no expansion, its introduction may have to await the wearing away of existing equipment, thus slowing down the rate of imitation;
- the passage of time. Over time the rate of imitation may increase due to better communication channels evolving, more-sophisticated techniques of evaluating equipment replacement, and more-favourable attitudes towards technological change;

the phase of the business cycle. It would be expected that the rate of imitation rises over the expansion phase and falls during the contraction phase.

Using the assumption that λij (t) can be adequately approximated by a Taylor's expansion, the model led to two predictions. Firstly, the number of firms having introduced an innovation, if plotted against time, should approximate a logistic function, an S-shaped growth curve frequently encountered in biology and the social sciences. Secondly, the rate of imitation should be higher for more-profitable innovations and those requiring relatively small investments.

An earlier study by Griliches (56) had concentrated on the factors causing wide regional differences in the use of hybrid corn seed in the United States. He fitted logistic growth functions to regional data and attempted to account for the differences by varying estimates of the three parameters of the logistic functions: origins, slopes, and ceilings. The development lag of hybrids for particular regions (origin differences) was explained by varying profitability as determined by market density, and innovation and marketing cost. Profitability differences of the shift from open pollinated to hybrid varieties in different regions, in part explained the ceilings (long-run equilibrium use) and the slopes (rate of approach to equilibrium).

Subsequent studies have tended to confirm that the central result of diffusion studies is the existence on an S-shaped diffusion curve. In addition to the logistic, the Gompertz and the log-normal curves have also merited attention. However, diffusion theory is still in its infancy (although there is no shortage of empirical studies) and the central approach of Mansfield, involving little more than a Taylor's expansion of an assumed general form, has not been materially improved upon. Recently, Stoneman and Ochoro (57) have attempted to tackle the problem of riskiness of new technology, previously largely ignored in the literature. They employ a means-variance approach in the modelling process and are also able to predict the existence of S-shaped diffusion curves, in so doing amending the theoretical and empirical "conventional wisdom" that the speed of diffusion is positively related to the level of profitability.

the phase of the business cycle. It would be expected that the rate of imitation rises over the expansion phase and falls during the contraction phase.

Using the assumption that λij (t) can be adequately approximated by a Taylor's expansion, the model led to two predictions. Firstly, the number of firms having introduced an innovation, if plotted against time, should approximate a logistic function, an S-shaped growth curve frequently encountered in biology and the social sciences. Secondly, the rate of imitation should be higher for more-profitable innovations and those requiring relatively small investments.

An earlier study by s (56) had concentrated on the factors causing wide regional dif ces in the use of hybrid corn seed in the United States. He fitted logistic growth functions to regional data and attempted to account for the differences by varying estimates of the three parameters of the logistic functions: origins, slopes, and ceilings. The development lag of hybrids for particular regions (origin differences) was explained by varying profitability as determined by market density, and innovation and marketing cost. Profitability differences of the shift from open pollinated to hybrid varieties in different regions, in part explained the ceilings (long-run equilibrium use) and the slopes (rate of approach to equilibrium).

Subsequent studies have tended to confirm that the central result of diffusion studies is the existence on an S-shaped diffusion curve. In addition to the logistic, the Gompertz and the log-normal curves have also merited attention. However, diffusion theory is still in its infancy (although there is no shortage of empirical studies) and the central approach of Mansfield, involving little more than a Taylor's expansion of an assumed general form, has not been materially improved upon. Recently, Stoneman and Ochoro (57) have attempted to tackle the problem of riskiness of new technology, previously largely ignored in the literature. They employ a means-variance approach in the modelling process and are also able to predict the existence of S-shaped diffusion curves, in so doing amending the theoretical and empirical "conventional wisdom" that the speed of diffusion is positively related to the level of profitability.

Further examination of diffusion theory is not warranted for the purposes of this thesis. The main conclusion is now obvious, namely that technological epochs for an aggregate (such as an industry) are more-sharply defined when the imitation rate of innovations is high than when it is low, for this results in more-distinct structural breaks.

REFERENCES

(1) International Labour Organisation, Coal Mines Committee, Fourth Session, Report III, Productivity in Coal Mines, Geneva, May 1951, page 8.

(2) International Labour Conference, 33rd Session, Geneva, 1950, Report I: Report of the Director-General, page 78.

(3) ILO, Productivity in Coal Mines, op cit., page 149.

(4) ibid, page 9.

(5) Encyclopaedia Britannica, 1974 edition.

(6) Salter, W.E.G. Productivity and Technical Change, Cambridge University Press, 1966, page 2.

(7) Oxford Illustrated Dictionary.

(8) Fabricant, S. "Basic Facts on Productivity Change. An Introduction", page xxxv in Kendrick, J.W. Productivity Trends in the United States, Princeton University Press, 1961, National Bureau of Economic Research, New York.

(9) ibid.

(10) Eilon, S. and Soesan, J. "Definitions and Prevailing Approaches", chapter 1 in Eilon, S. Gold, B. and Soesan, J. Applied Productivity Analysis for Industry, Pergamon International Library, 1976, page 3.

(11) Kendrick, J.W. Producti Trends in the United States, op cit., page 3.

(12) Salter, W.E.G. Productivity and Technical Change, op cit., pages 1-3.

(13) Gold, B. "A Framework for Productivity Analysis", chapter 2 in Eilon et al Applied Productivity Analysis for Industry, op cit.

(14) Solow, R.M. "Education and Economic Productivity", in Harris, S. and Levensohn, A. (eds), Educational and Public Policy, McCutcheon, Berkeley, California, 1965.

(15) Brown, M. On the Theory and Measurement of Technological Change, Cambridge University Press, 1966, chapter 2.

(16) Gehrig, W. "On Certain Concepts of Neutral Technical Progress: Definitions, Implications and Compatability", chapter 1 in Puu, T., and Wibe, S. (eds), The Economics of Technological Progress, Macmillan, 1980.

(17) Hicks, J.R. The Theory of Wages, Macmillan, 1932, pages 121-7.

(18) Salter, W.E.G. Productivity and Technical Change, op cit., pages 29-33.

(19) Harrod, R.F. Towards a Dynamic Economics, Macmillan, 1948, page 23.

(20) Robinson, Mrs. The Accumulation of Capital, Macmillan, 1956, pages 132-4.

(21) Hahn, F.H. and Matthews, R.C.O. "The Theory of Economic Growth: A Survey", in Surveys of Economic Theory: Growth and Development, St. Martin's Press, 1967, New York, page 50.

(22) Kennedy, C. "Technical Progress and Investment", Economic Journal, June 1961.

(23) Harrod, R.F. "The Neutrality of Improvements", Economic Journal, June 1961.

(24) Kennedy, C. "Harrod on Neutrality", Economic Journal, March 1962.

(25) Kennedy, C. "The Character of Improvements and of Technical Progress", Economic Journal, December 1962.

(26) Sato, R. and Beckmann, M. "Neutral Inventions and Production Functions", Review of Economic Studies, 1968.

(27) Sato, R. and Beckmann, M. "Aggregate Production Functions and Types of Technical Progress: A Statistical Analysis", American Economic Review, 1969.

(28) Gehrig, op cit., pages 3-6.

(29) Wallis, K.F. Topics in Applied Econometrics, Gray-Mills Publishing Ltd, 1973, page 48.

(30) Solow, R.M. "Investment and Technical Progress", chapter 7 in Arrow, K.J. Karlin, S. and Suppes, P. Mathematical Methods in the Social Sciences, Stanford University Press, 1960. page 90.

(31) Venieris, Y.P. and Sebold, F.D. Macroeconomic Models and Policy, Wiley/Hamilton, 1977, pages 594-9.

(32) Arrow, K.J. "The Economic Implications of Learning by Doing", Review of Economic Studies, June 1962.

(33) Brown, M. On the Theory and Measurement of Technological Change, op cit., page 78.

(34) Alchian, A. "Costs and Outputs", reprinted as chapter 12 in Breit, W. and Hochman, H.M. Readings in Microeconomics, Hold, Rinehart, and Winston, 1969, page 169.

(35) Wright, T.P. "Factors Affecting the Cost of Airplanes", Journal of the Aeronautical Sciences, 1936, volume 3.

(36) Asher, H. Cost-Quantity Relationship in the Airframe Industry, The Rand Corporation, Santa Monica, California, July 1956.

(37) Hirsch, W.Z. "Manufacturing Progress Functions", Review of Economics and Statistics, May 1952.

(38) Hirsch, W.Z. "Firm Progress Ratios", Econometrica, 1956.

(39) Lundberg, E. Produktiviteit och räntabilitet, P.A. Norstedt and Söner, Stockholm, 1961.

(40) Johansen, L. "Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis", Econometrica, April 1959.

(41) Solow, R.M. "Investment and Technical Progress", op cit.

(42) Salter, W.E.G. Productivity and Technical Change, op cit.

(43) Wonnacott, P. Macroeconomics, Irwin, 1978, pages 482-4.

(44) Solow, R.M. "Technical Progress, Capital Formation, and Economic Growth", American Economic Review, May 1962.

(45) Brown M. On the Theory and Measurement of Technological Change, op cit., pages 81-92.

(46) Cobb, C.W. and Douglas, P.H. "A Theory of Production", American Economic Review, Supplement, March 1928.

(47) Brown, M. On the Theory and Measurement of Technological Change, op cit., chapter 3.

(48) Arrow, K.J. Chenery, H.B. Minhas, B.S. and Solow, R.M. "Capital-Labour Substitution and Economic Efficiency", Review of Economics and Statistics, August 1961.

(49) Brown M. and De Cani, J.S. "Technical Change and the Distribution of Income", International Economic Review, September 1963.

(50) Wallis, K.F. Topics in Applied Econometrics, op cit., pages 52-55.

(51) Hicks, J.R. Theory of Wages, op cit., pages 182-3.

(52) Brown, M. On the Theory and Measurement of Technological Change, op cit., chapter 5.

(53) Schumpeter, J. Business Cycles, McGraw-Hill, 1939.

(54) Usher, A.P. History of Mechanical Inventions, McGraw-Hill, 1954.

(55) Mansfield, E. "Technical Change and the Rate of Imitation", Econometrica, October 1961.

(56) Griliches, Z. "Hybrid Corn: An Exploration in the Economics of Technological Change", Econometrica, October 1957.

(57) Stoneman, P. and Ochoro, W. "Means-Variance Approach to the Theory of Intrafirm Diffusion", chapter 2 in Puu, T. and Wibe, S. The Economics of Technological Progress, op cit.

CHAPTER SEVEN

METHODS OF PRODUCTIVITY MEASUREMENT : (i) UNREFINED STUDIES

Technical progress implies advances in knowledge, but advances in knowledge per se are not amenable to direct quantification. The best that can be done is to measure technical progress by its effects. This encompasses generally an improvement in human welfare, both quantitatively through increases in real income per capita and qualitatively through a wider range of goods and services and an extension of leisure. Thus technical progress may take different forms including new production processes, new goods and services, and new methods of industrial organisation. The effect of technological progress can be captured quantitatively by a shift in the production function enabling greater output to be produced with unchanged inputs or unchanged output to be produced with smaller levels of inputs. The most practical, and widely used, indicator of technical progress is a productivity index. Generally speaking, either an arithmetic index is obtained through dividing an output index by an input index, or a geometric index is obtained from the multiplicative form of the production function (1). These techniques form the subject matter of this chapter.

Ratio Method

Measurement of the volume (real value) of output can reveal characteristics of the growth or decline of an economy, or of segments of it, but it yields an inadequate picture of changes in the efficiency of the productive system from one time period to another. This is because it takes no account of the draft made upon the economic resources of the nation in producing goods and services. Any measure of productive efficiency, therefore, needs to relate the real output of the economy, or its segments, to the real inputs necessary to produce that output. The ratio method of measuring productivity can, therefore, be defined as the ratio between arithmetic indices of real output and real inputs. If output increases at a faster rate than inputs over a given time period this can only be because the economy is becoming more productive in transforming inputs into output. Productivity, therefore, is what is left after real input changes are purged from real output changes and for this reason can be referred to as a "residual". It is a "catch-all" element, a convenient name used to label all output changes which cannot obviously be ascribed to physical changes in input resources. It has even been referred to as "a measure of our ignorance" due to the fact that it is what is "left-over" at the end of the process and although it may be possible to statistically measure the size of the "left-over" there is little certainty what determines its size or what its major components (and their relative sizes) happen to be.

As recently as the late 1940's this ratio method concentrated virtually exclusively on measuring labour productivity, i.e. the ratio between real output and real labour input. Any of four measures were variously employed, namely output per man year, man day, man shift, or manhour. Other inputs, notably capital and raw materials, were ignored, although it was conceded that they had to be included to obtain meaningful results. In a major study of the American mining industry in 1944 Barger and Schurr (2) explained their decision to concentrate solely on labour productivity;

> "the input of resources - human and material - cannot be aggregated with the facility with which we can total the things emerging from a productive process. Partly for this reason, and partly because human resources are of special interest, we shall confine our measurements of input to labour, i.e. we shall measure only the volume of employment. Consequently our figures of productivity will be derived only in terms of labour, and we will neglect the input of capital and other factors".

They could also have mentioned that an additional factor in their decision concerned the myriad problems associated with measuring capital.

The problem with partial productivity studies such as that performed by Barger and Schurr is that it is not possible to tell whether an observed increase in labour productivity has been caused through a genuine increase in labour efficiency, or by the substitution of capital for labour. This is because partial productivity ratios reflect both inter-factor substitution as well as changes in overall productive efficiency. A potential solution is to analyse data on changes in capital-output ratios in order to throw light on the significance of changes in labour productivities. A more-efficient solution, however, is to relate output to all tangible inputs. This automatically eliminates the problem of inter-factor substitution making it immediately possible to see whether there has been a net saving in real costs per unit of output, i.e. a gain in productivity. This involves aggregating all the major tangible inputs into the ratio denominator and ascribing productivity or technological progress to the residual which is left-over only after the impact of all factor inputs has been purged from output changes.

Between 1948 and 1960 a series of articles appeared which attacked the problem of aggregating different factor inputs into a composite input index, thus solving Barger and Schurr's difficulty mentioned above. These articles, and the techniques they introduced, can now be examined.

In their study of United States' agriculture between 1910 and 1945, Barton and Cooper (3) defined productivity as the ratio of output to all inputs. Output was measured in terms of constant dollars. Inputs (land, labour, power and machinery) were measured in terms of their constant-dollar costs, and then combined to obtain an aggregative measure of "total physical inputs of operation". Either physical input quantities were weighted by average 1935-39 prices of input factors, or current-dollar costs were deflated by indices of prices.

BASE YEAR				CURRENT YEAR
ITEM	UNIT	q_o NO. OF UNITS	P AVERAGE 1935-39 PRICE PER UNIT	q_1 NO. OF UNITS
LABOUR (L)	MANHOURS	Lo	P_L	L_1
MACHINES (M)	NUMBER	Mo	P_M	M_1
LAND (LD)	ACRES	LDo	P_{LD}	LD_1
POWER (P)	K/WATTS	Po	P_P	P_1

The exercise involves a simple weighted aggregative index,

$$\frac{\Sigma\, q_1 p}{\Sigma\, q_0 p} = \frac{L_1 P_L + M_1 P_M + LD_1 P_{LD} + P_1 P_P}{L_0 P_L + M_0 P_M + LD_0 P_{LD} + P_0 P_P}\ .$$

The authors experimented with various years to act as the weighting period and found 1935-39 to be as good as any.

Their major finding was that output per unit of all inputs had shown an upward trend since the First World War. This had resulted from a "remarkable" stability of total inputs and a steady upward trend in the volume of output, implying that increases in physical efficiency were brought about by increasing output per unit of input rathe than by decreasing total inputs. Technological progress had thus been output-expanding in agriculture rather than input-contracting.

Schmookler (4) employed the similar concept of output per unit of total input measured in constant dollars to undertake a wider study of productivity in the American economy over the period 1869 to 1938. He called his ratio the "aggregate-efficiency index". Output was measured as Gross National Product in constant 1929 dollars. Inputs were categorised as land, labour, capital, and enterprise measured in constant 1929 dollars, and combined to form a single total input figure. Over the study period Schmookler found output per unit of input increased by 122 per cent, or at a compound rate of 1,36 per cent per annum, when labour is defined in manhours; and by 75 per cent, or at a compound rate of 0,92 per cent per annum, when labour is defined in manyears. The latter growth rate is lower because of the decrease in working hours.

The tradition of constant-dollar aggregation of inputs as formulated by Barton and Cooper and Schmookler was superseded from the mid-1950's onwards by what can be called the multi-factor productivity ratio. This was designed to measure the average product of all productive services, such services being combined by a certain weighting scheme. It is only movements in this index which are considered to be productivity advances, whereas increases in the average product of a particular service caused by a shifting input mixture in response to changing relative prices does not constitute productivity advances. Output is expressed as a percentage of a weighted sum of labour and capital, i.e.

$$\frac{X}{\alpha L + \beta K} \qquad \text{(i)}$$

where X = output, L = labour, K = capital, and α and β are the respective weights (usually the prices of the factors in certain selected periods). In a static sense this ratio is interpreted as a measure of the output per unit of resources foregone in its production.

The prototype of multi-factor productivity studies was provided by the work of Kendrick both in his 1956 article (5) but more especially in his monumental tome (6), commenced in 1953 and published in 1961. Kendrick called his index "total factor productivity" and used factor prices as weights, i.e. the market value of factor services. This is because in order to determine the changes in aggregate output and factor inputs (and hence productivity) it is necessary to combine unlike types of output and input units by weights that indicate their relative importance. Kendrick contended that for productivity analysis purposes, outputs should be weighted by product prices (theoretically at factor cost but in practice by the use of market prices) and inputs should be weighted by unit factor compensation (factor price). Accordingly, in the base period the values of output and input are equal. The unit values of the outputs are proportional to the values of the factor services required for their production, and the unit values of the inputs are proportional to the shares of the value of outputs which they obtain for their services. Under perfect competition the factor prices represent, in equilibrium, the relative values of their marginal contributions to output.

Considering the case of homogeneous output, the two crucial assumptions which are made are those of constant returns to scale and pure competition in factor markets. By Euler's Theorem, constant returns to scale implies that total output is equitable to the sum of the inputs multiplied by their respective marginal productivities. The assumption of pure competition in factor markets enables the use of market prices as weights in place of marginal productivities. By further assuming neutral technological progress the productivity-change index can be bracketed by the Laspeyres Index (employing base-period prices as weights) and by the Paasche Index (employing final-period prices as weights). The former gives the lower limit and the latter the upper limit.

Specifically the labour-input measure is based on estimates of manhours worked weighted by base-period average hourly earnings. Real capital input is measured by the constant-dollar value of the stock of real capital weighted by base-period rates of return. This can be expressed as a production equation for a fully integrated industry with a single product as,

$$X = P(w_0L + r_0K)$$

where X = output, L and K are labour and capital, as defined, in any given year, w_0 and r_0 are their respective weights, as defined, in the base year, and P is the arithmetic index of productivity which unites the two sides of the equation. Hence,

$$P = \frac{X}{w_0 L + r_0 K} \qquad \text{(ii)}$$

which correlates with expression (i).

An alternative, but identical, way of formulating the same expression is to reduce all variables to index numbers with a common base period with appropriate weights being the factor share of the relevant input, i.e,

$$\frac{X}{X_0} = P\left\{\gamma\frac{L}{L_0} + \sigma\frac{K}{K_0}\right\}$$

where X_0, L_0, and K_0 are output, labour, and capital respectively in the base year, and γ and σ are the shares of labour and capital respectively in the value of output in the base year. Hence,

$$P = \frac{X/X_0}{\gamma(L/L_0) + \sigma(K/K_0)} . \qquad \text{(iii)}$$

Expressions (ii) and (iii) are equivalent, and the choice between them is simply a matter of convenience. It is easily shown that expressing all variables as index numbers with a common base period and weighting inputs arithmetically by base-period factor shares is identical with the construction of a Laspeyres output per unit of input index. However, using final-period factor shares as weights is not equivalent to the Paasche Index.

Since the value of the whole product is absorbed by the inputs in the base year, the value of P starts from unity in the base year and either increases or retrogresses from unity with the passage of time.

The meaning and interpretation of "total factor productivity" as formulated in the above expressions is adequately described by

Kendrick. By weighting the real input units by their base-period prices (which approximate their marginal contribution to output in that period) measures are obtained of what the inputs of a given period would have produced had their productive efficiency remained at the base-year level. The ratio of these inputs of "standardised efficiency" to the actual output of the given period, at base-period prices, therefore yields the index of change in productive efficiency.

Over the study period 1899 to 1953, Kendrick found that total factor productivity in the United States' private domestic economy increased at an average annual rate of 1,75 per cent. Productivity gains thus accounted for more than half of the 3,3 per cent average rate of growth in real product.

It is a paradoxical situation that when Kendrick criticises partial productivity ratios and states that his objective is to relate output with the combined use of all resources he is simultaneously aware that in the context of his method such an achievement would be fruitless. Domar (7) notes, correctly, that "this scholarly triumph would almost obliterate the index", i.e., by definition, the index would never budge from unity. It effectively becomes dead the moment that success is achieved in decomposing output changes into all their component input parts. There becomes no residual which is left-over that can be labelled efficiency. What Kendrick really means is that output should be related to all "conventionally"-defined inputs. In practice this implied adding, at least, a "tangible" capital-input measure to the labour-input measure. Of course, there are "other forces" which should ideally be added to, or included in, these two conventionally-defined inputs but Domar adds the reminder that the index only retains "life and interest" because these other forces have not been counted amongst the inputs. It must be ensured that "something" is omitted from the inputs which can later be labelled "efficiency" or "technological change". In reality, it is stretching credibility too far to refer to P as productivity increases. In effect, P acts as a "sponge"; it absorbs everything left-over after changes in conventionally-defined inputs have been purged from output changes. Domar would prefer it to be called the "residual".

These arguments are, of course, well known to Kendrick. He acts as he does, not out of ignorance, but necessity. Domar pointedly adds the reminder that "like politics, empirical work is the art of the

Kendrick. By weighting the real input units by their base-period prices (which approximate their marginal contribution to output in that period) measures are obtained of what the inputs of a given period would have produced had their productive efficiency remained at the base-year level. The ratio of these inputs of "standardised efficiency" to the actual output of the given period, at base-period prices, therefore yields the index of change in productive efficiency.

Over the study period 1899 to 1953, Kendrick found that total factor productivity in the United States' private domestic economy increased at an average annual rate of 1,75 per cent. Productivity gains thus accounted for more than half of the 3,3 per cent average rate of growth in real product.

It is a paradoxical situation that when Kendrick criticises partial productivity ratios and states that his objective is to relate output with the combined use of all resources he is simultaneously aware that in the context of his method such an achievement would be fruitless. Domar (7) notes, correctly, that "this scholarly triumph would almost obliterate the index", i.e., by definition, the index would never budge from unity. It effectively becomes dead the moment that success is achieved in decomposing output changes into all their component input parts. There becomes no residual which is left-over that can be labelled efficiency. What Kendrick really means is that output should be related to all "conventionally"-defined inputs. In practice this implied adding, at least, a "tangible" capital-input measure to the labour-input measure. Of course, there are "other forces" which should ideally be added to, or included in, these two conventionally-defined inputs but Domar adds the reminder that the index only retains "life and interest" because these other forces have not been counted amongst the inputs. It must be ensured that "something" is omitted from the inputs which can later be labelled "efficiency" or "technological change". In reality, it is stretching credibility too far to refer to P as productivity increases. In effect, P acts as a "sponge"; it absorbs everything left-over after changes in conventionally-defined inputs have been purged from output changes. Domar would prefer it to be called the "residual".

These arguments are, of course, well known to Kendrick. He acts as he does, not out of ignorance, but necessity. Domar pointedly adds the reminder that "like politics, empirical work is the art of the possible". Fabricant (8), in the introduction to Kendrick's book, also discusses the definition of inputs mentioning particularly the exclusion of "intangible capital". This results in an understatement of combined inputs in the ratio denominator "for it is likely that intangible capital has risen in relation to the resources (inputs) he includes". Correspondingly, there is an overstatement of the rise in productivity. Intangible capital can be defined, in a broad sense, to include all the improvements in basic science, technology, business administration and education and training that aid in production. Whether these result from deliberate individual or collective investments for economic gain or are incidental by-products of efforts to reach other goals, is immaterial. Intangible capital, so defined, and included in the combined-input factor would probably eliminate much of the "residual". But since emp ... work is the art of the possible, an obvious difficulty of data measurement arises. In addition, it is also a matter of semantics whether certain aspects of intangible capital should be ascribed as input or whether they can genuinely be labelled efficiency increases. After all, is not an improvement in, say, administrative efficiency an increase in productivity? But Fabricant appears perfectly satisfied to lump together under the heading of productivity, and measure as a whole, both the indirect effects of the increases in conventional, tangible inputs and the effects of all other causes. Even though this method may do no more than suggest the high relative importance of the factors grouped under productivity, this in itself is significant. The productivity measure reflects, to a large extent, the excluded input of intangible capital accumulated in order to improve the efficiency (productive capabilities) of the tangible inputs, and this, and other qualitative elements, cannot be independently measured in a satisfactory way. For this reason alone Kendrick's index has a role to play.

Fabricant stresses that the residual contains not only the contributions from the various forms of intangible capital, but includes other elements also, namely,

- the economies resulting from increased specialisation within and between industries, made possible by growth in the nation's resources and in its scale of operations,

- the improvement or falling off of efficiency in the use of resources resulting from changes in the degree of competition, in the volume, direction or character of government subsidies, in the nature of the tax system, and in other government activities and regulations,
- the greater or smaller benefits resulting from changes in the volume, character, and freedom of commerce among nations.

Kendrick himself prefers to decompose his productive-efficiency residual into differently-labelled categories, namely,

- intangible capital,
- technological change or innovation, reflecting advanced know-how brought to bear by entrepreneurs on productive processes,
- changes in the rate or scale of output,
- changes in the rate of utilisation of capacity.

Of course, mere description of the components of changing productive efficiency does not explain the causes of the changes, but Kendrick is not trying to be all things to all men.

Kendrick attempts to minimise the effect of changing rates of utilisation of capacity by basing his analysis largely on "key years" of relatively high-level economic activity when it can be safely assumed that the major industries were operating near maximum capacity. By eliminating this variable in this manner he is able to simplify the decomposition of his productive-efficiency residual even more, namely, as reflecting the net effect of changes in scale as well as innovations that are not associated with changes in scale. The former refers to the net outcome of tendencies towards increasing and diminishing returns that arise because certain inputs need not be expanded proportionally with output, while others cannot be. The latter refers to both improvements in intangible capital and innovations in technological matters regardless of whether these have arisen autonomously or been induced by changes in scale. Invariably, inventions are induced when production is organised on a larger scale than hitherto. More opportunities are afforded for organisational innovations and the associated productivity increase depends on the alertness of management and their flexibility in adapting to the cost-reducing possibilities.

But without autonomous investment output growth would be slower with fewer attendant scale economies. In practice it is not feasible to decompose a given productivity change into a part resulting from autonomous innovations and a part resulting from induced scale innovations.

Some criticisms, clarifications and insights into Kendrick's index are worth mentioning (following Domar):

(a) The index is meant to measure "efficiency in the use of resources" not "economic efficiency" defined as the ratio of actual to potential output, or the proximity to some optimum. Even so between 1945 and 1960 the Soviet index grew faster than the United States' index but Domar was "not quite ready to award the Soviets the efficiency prize". A falling index is not necessarily a sign of inefficiency, e.g. poorer lands or ores may be being utilised or perhaps services are being expanded where "other forces" have less room to operate. On the other hand a rising index in a particular industry could have been caused by the peculiar behaviour of its inputs or outputs.

(b) The constant prices chosen as input weights represent, or approximate, the respective marginal products. This assumes pure competition, i.e. the firm is in short-run equilibrium with respect to labour (which can be accepted with reservations) and long-run equilibrium with respect to capital (a hazardous assumption in a study of economic growth). But since Kendrick's price of capital is the average rate of return and not a rental payment or market rate of interest there appears no reasonable alternative in the context of his method.

(c) The arithmetic combination of inputs used in expressions (ii) and (iii) is not a good production function since it assumes that the input's marginal products are changed only by the "other forces" and always in the same proportion so that their ratios remain constant and independent of the ratio of the quantities of the inputs, however fast capital grows relative to labour. This in spite of the fact that real income per unit of labour had risen many times faster than that of capital. To tackle this problem Kendrick changed his weights several times, generally using average prices in the terminal years of the various sub-periods as weights. Accordingly, by periodically changing the weights, productivity changes in each sub-period reflect the concurrent economic structure but simultaneously bias downwards the total factor productivity. But Domar suggests a better alternative would have been a geometric index with income shares, not prices, as weights, i.e

$$X = AL^{\gamma} K^{\sigma} \qquad \text{(iv)}$$

where A is the geometric index of productivity and the other variables are defined as previously. Correspondingly, the rate of growth of the geometric index becomes

$$\overline{A} = \overline{X} - \gamma\overline{L} - \sigma\overline{K} \qquad \text{(v)}$$

where the bar over a variable refers to the annual rate of growth of the variable. It must be admitted that labour's share has risen considerably over the century but the assumption of constant shares creates less trouble than constant relative prices.

(4) Since for the whole economy Kendrick expresses output as net national product at factor cost, consistency dictates that for individual industries output be expressed in a similar manner, i.e. net value-added. The output of the economy nets out intermediate goods and services (materials, fuel, services, semi-fabricated components). It is measured in terms of final products only which encompasses intermediate goods automatically. Thus, the inputs associated with national product reduce to the conventional labour and capital. Industry output to be consistent must be measured not on a gross basis but on a net basis where such intermediate goods are subtracted from gross output. What is left is value-added. Alternatively, output can be left gross and intermediate goods added as an additional input in the equation, i.e.,

$$X = P\,(w_oL + r_oK + h_oR) \quad , \qquad \text{(vi)}$$

Hence,

$$P = \frac{X}{w_o L + r_o K + h_o R} \qquad \text{(vii)}$$

where h_o is the real price of materials in the base year, and R is the input of materials in a given year.

(e) Kendrick's weighting system automatically gives a large relative weight to labour as compared to capital (about 8 to 2), so that although his index is a combined-input index and, therefore, theoretically better than a partial productivity index the movement of the index closely mirrors the movement of labour productivity because capital is assigned a minor role.

The discussion of Kendrick's method has been rather lengthy partly because several points are common, and need not be reiterated at length again.

Abramovitz (9) taking his inspiration from Kendrick, calculated index relatives, between two given dates, for the growth of manhours and the growth of capital and combined them using as weights their respective shares in the base period. This he called the "index of total input of resources", which when divided into the index relative for output (net national product) resulted in the "index of net national product per unit of total input". The excess of this ratio over 100 indicates the magnitude of productivity increase. The size of the excess is known as the "Abramovitz Residual", or, as he originally named it, "a measure of our ignorance" concerning the causes of economic growth since so little was known about the causes of productivity increase.

Abramovitz's conclusion was even more startling than that of Kendrick's in emphasizing the importance of the contribution made by technological progress. Over the study period 1869 to 1953 the average rate of growth of net national product had been 3,5 per cent per annum. The average rate of growth of net product per capita had been 1,9 per cent per annum of which almost the entire increase was found by Abramovitz to be associated with the rise in productivity.

Abramovitz merely isolated the residual and emphasised its size but could offer little explanation of why it was so significant, hence our "ignorance". Thus the residual still acts like a sponge and the same remarks are relevant here as were made in connection with Kendrick's method. The lopsided importance given to productivity increase, according to Abramovitz, should act as a sobering influence on students of economic growth simply because so little is known about the causes of productivity increase. Nevertheless, its significance lies by way of "some sort of indication of where we need to concentrate our attention".

Symbolically, if $\Delta X/X$ represents the percentage change in output over a period of time, and $\Delta L/L$ and $\Delta K/K$ the percentage changes in labour and capital respectively, the objective is to determine how much of $\Delta X/X$ can be attributed to something other (i.e. the productive-efficiency residual) than changes in the physical inputs. Since γ and σ represent labour and capital's shares respectively in the base period, therefore,

$$\frac{\Delta X}{X} - \gamma \frac{\Delta L}{L} - \sigma \frac{\Delta K}{K} = \text{residual} \qquad \text{(viii)}$$

which is exactly the same expression as in (v), which when integrated gives

$$\frac{X}{L^{\gamma} K^{\sigma}} = \int \text{residual } dt \qquad \text{(ix)}$$

which is exactly the same expression as in (iv), as recommended by Domar. This geometric index is, of course, of a form derived from the Cobb-Douglas production function, implying that the Abramovitz Residual is obtained on the basis of a production process which assumes a unitary elasticity of substitution.

Kendrick's index, it must be remembered, used a linear combination of labour and capital using factor prices as weights, employing Euler's Theorem on homogenous functions with an implicit modification for imperfect competition. Kendrick's formulation is more general than the Abramovitz-Domar one. The latter specializes the underlying production function to one homogeneous function (Cobb-Douglas) whereas the former's use of the Euler Theorem assumes a production function of any homogeneous form. In emphasis, however, Kendrick does not directly use a production function, nor specify a particular production function. His reliance on the Euler transformation is a very general formulation requiring only that the underlying production function be homogeneous.

Reddaway and Smith (10) introduced a slight re-formulation of the problem in their study of British manufacturing industry. They compared 1954 directly with 1948 to obtain a discrete measure of progress during this period, also using the concept of output per unit of all (labour and capital) inputs. They defined the simple productivity index as

$$\frac{X_2 \,/\, X_1}{I_2 \,/\, I_1} = \frac{\text{Net-Output Index}}{\text{Combined-Input Index}} . \qquad (x)$$

To obtain the weighted average of the labour and capital variables they used the conventional method of valuing the units at their respective prices in the base period and weighting labour by the base-year wage rate and capital by its unit return, i.e.,

$$I = \frac{w_1 L_2 + r_1 K_2}{w_1 L_1 + r_1 K_1} . \qquad (xi)$$

The productivity index can, therefore, be stated as

$$\frac{P_1 X_2 / P_1 X_1}{w_1 L_2 + r_1 K_2 / w_1 L_1 + r_1 K_1} \qquad (xii)$$

where P_1 is the net-output price in the base year.

Thus far, so much is obvious. The innovation of Reddaway and Smith stemmed from their dislike of using the base-year return to capital as a weight both because of its extreme volatility as a variable and the difficulty inherent in measuring the stock of capital. They therefore postulated on an a priori basis a capital charge to be applied to all industries (equal to 15 per cent) and developed a method using gross-capital formation instead of the stock of capital. They derived a formula which enabled the variables to be expressed as absolute changes thereby dispensing with the need for base-year total figures. Following Kendrick's utilisation of the Euler Theorem that inputs totally exhaust output, $w_1L_1 + r_1K_1 = P_1X_1$, the productivity index reduces to

$$P_1X_2/w_1L_2 + r_1K_2. \qquad \text{(xiii)}$$

Using the symbol Δ to refer to the extra amount of a variable produced or employed in 1954 as compared with 1948 they show that the productivity index reduces to

$$\frac{P_1\Delta X - (w_1\Delta L + r_1\Delta K)}{P_1X_1 + w_1\Delta L + r_1\Delta K} \qquad \text{(xiv)}$$

for which data is readily available. The symbols represented in (xiv) translate to

$$\frac{\text{Increase in Output minus Allowance for Extra Inputs}}{\text{Output which would have been attained with unchanged}}\ \text{productivity.}$$

Over the study period the authors found net output to have increased by 33,2 per cent (4,9 per cent per annum), but since labour input had increased by 12,6 per cent (2,0 per cent per annum), this implied an increase of labour productivity of 18,3 per cent (2,8 per cent per annum). The contribution of increased capital to this amount was only 0,7 per cent per annum leaving 2,1 per cent per annum to be explained by pure progress. This result once again confirmed the findings of previous studies placing the major emphasis on the contribution of technological progress.

Solow Method

The Solow Method of measuring technological change, using output per head, rather than output, as the dependent variable, was pioneered by Robert Solow (11), utilised by Massell (12), Lave (13), and Chandler

(14), and analysed and criticised by Hogan (15) and Domar (16).

Solow's justification for returning to what he called "this old-fashioned topic" was the novelty of a "new wrinkle" in segregating variations in output per head into changes in the availability of capital per head on the one hand, and technological change on the other. Operating on the assumptions that technological change is neutral and disembodied and that the economy is operating in the range of constant returns to scale, Solow makes use of marginal productivity conditions in an attempt to differentiate shifts of the production function from movements along it. The aggregate production function measured in physical units can be written as,

$$X = F(K,L;t)$$

where t stands for time, and allows for technological progress, thus representing any kind of shift in the production function. If technological change is neutral the expression can be written,

$$X = A(t)f(K,L) \qquad \text{(xv)}$$

where A(t) is a multiplicative factor measuring the accumulated effect of shifts over time. Differentiating totally with respect to time and dividing by X, the expression below can be obtained (where dots over variables indicate time derivatives),

$$\frac{\dot{X}}{X} = \frac{\dot{A}}{A} + A\frac{\partial f}{\partial K}\frac{\dot{K}}{X} + A\frac{\partial f}{\partial L}\frac{\dot{L}}{X} \qquad \text{(xvi)}$$

If the assumption is now added that factors are paid their marginal products, i.e. $\partial X/\partial K = r/p$ and $\partial X/\partial L = w/p$,therefore the relative shares of labour (γ) and capital (σ) can be defined as $\partial X/\partial L . L/X$ and $\partial X/\partial K . K/X$ respectively, and substituting into (xvi) it is possible to obtain,

$$\frac{\dot{X}}{X} = \frac{\dot{A}}{A} + \sigma\frac{\dot{K}}{K} + \gamma\frac{\dot{L}}{L} \qquad \text{(xvii)}$$

Data can be obtained for every term in expression (xvii) except

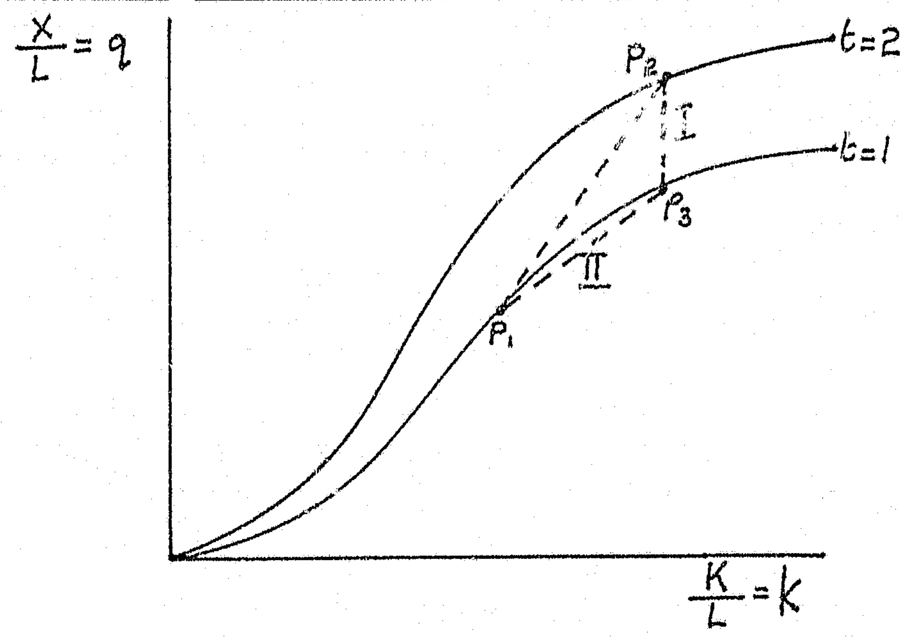
$\dot{A}/A$ which denotes technological change, and which can be derived as a residual once all the other terms are evaluated. It will be noted that expression (xvii) is identical to expressions (viii) and (ix), the "Abramovitz Residual". Thus the Solow measure employs a variant of the total factor productivity index, the Abramovitz-Domar method of isolating the residual, based on an underlying Cobb-Douglas production function. However, Solow simplifies the expression still further. Assuming constant returns to scale gives $\gamma + \sigma = 1$ by Euler's Theorem. Variables can be converted to per capita units. Letting $q = X/L$, and $k = K/L$, and $\gamma = 1-\sigma$,expression (xvii) becomes,

$$\frac{\dot{q}}{q} = \frac{\dot{A}}{A} + \frac{\sigma\dot{k}}{k} \quad . \qquad \text{(xviii)}$$

Therefore, to disentangle the technological-change index one needs only time-series data for output per manhour, capital per manhour, and the share of capital.

A simple answer can now be given to the question of how much of the increase in output per manhour is due to the increase in capital per manhour and how much to technological progress. The production function is completely represented by a graph of q against k (output per manhour against capital per manhour), but the function is shifting over time so that observed points in the plane are compounded out of movements along the curve (increase in k) and shifts of the curve (technological change), i.e. represented by II and I, respectively, in the diagram below.

FIGURE 7.1 SHIFTS OF, AND MOVEMENTS ALONG, THE PRODUCTION FUNCTION



Every estimate on the curve t = 1 has been multiplied by the same factor to give a neutral upward shift of the production function to t = 2. How to estimate this shift from knowledge of only two points, P_1 and P_2, is the problem, for obviously to fit a curve through P_1 and P_2 is misleading. Statistically the distinction is obscured by the fact that while a production function represents a range of hypothetical alternative factor combinations, at any one time only one combination is observed. If the function shifts over time due to technological progress only one point on each function is observable and the effects on output per head of technological progress and increased capital per head are compounded. What is needed is an estimate of the shift factor for each point of time to act as a correcting factor for technological progress. This can be obtained by dividing q(t), output per manhour, by A(t), the technological-change index. This estimates what would have occurred without technological change, i.e. it represents the increase in output per manhour attributable to the increase in capital per manhour, (i.e. distance II). Solow's finding are remarkable. For the period 1909 - 1949 he found that approximately 10 per cent of the increase in output per manhour was due to increases in capital per manhour (distance II) whilst the remaining 90 per cent was due to technological change (distance I).

Solow's analysis was concerned with the nonfarm private sector of the economy and Massell, Chandler and Lave subsequently took up Solow's suggestion that a more-appropriate study would be one having a narrowlydefined production function where inputs and outputs would be specifically enumerated.

Massell believes that he takes a step in this direction in considering only the manufacturing sector. Although inputs and outputs do not entirely comply with Solow's recommendation, nevertheless the study concentrates on a sector producing physical goods only so that increased output homogeneity is achieved. Massell studies the period 1919 - 1955 using Solow's model,

$$\frac{\dot{A}}{A} = \frac{\dot{q}}{q} - \alpha\frac{\dot{k}}{k} \quad . \qquad \text{(xix)}$$

His only point of departure lies in the definition and treatment of three time-series variables, and his results exactly parallel those of Solow, namely 90 per cent of the increase in output per manhour is attributable to technological change. Such a conclusion, according to Massell, should be a help to policy makers in determining what proportion of our investment resources should be devoted to improving the technology rather than to expanding existing types of capital equipment and structures.

Lave considers only the agricultural sector over the period 1850 - 1958. He too is concerned with dichotomising increases in output per head into increases in capital per head, and technological change. He also utilises Solow's formula although he derives it in a remarkably simple way using logic rather than mathematics. If the whole of an increase in output per head is attributed to technological change, i.e.,

$$\frac{\dot{A}}{A} = \frac{\dot{q}}{q}$$

then this is an overestimate by some percentage of the increase in capital per head. But what percentage? The most obvious answer is the ratio of capital's income to total income. If factors are paid their marginal product this equals capital's contribution and if the production function is linear and homogeneous this is the amount which should be subtracted to account for movements along the production function. In terms of the previous diagram, if $P_2 - P_1$ is used to approximate the shift of the function (technological progress), the overestimate is $P_3 - P_1$, or the increase in capital per head times the slope of the function, i.e., $\Delta k\, df/dk$, but $df/dk = \sigma$ the share of capital. Thus,

$$\Delta A(t) = (P_2 - P_1) - (P_3 - P_1) = (P_2 - P_1) - \sigma\Delta k \qquad \text{(xx)}$$

and the analysis returns to the Solow formulation,

$$\frac{\Delta A}{A} = \frac{\Delta P}{P} \left(\frac{=\Delta q}{q}\right) - \frac{\sigma\Delta k}{k} \quad . \qquad \text{(xxi)}$$

For the agricultural sector Lave confirmed the early Solow-Massell conclusions giving overriding importance to technological change, but this time capital was not so insignificant contributing between 27 and 40 per cent of the increase in output per head.

Chandler performed a three-way comparison between the overall economy, the farm sector, and the nonfarm sector for the period 1946 - 1958. Following the Solow method he calculated a technological-change index A(t). He dichotomised increased output per head into increased capital per head and technological change by an analogous method to Solow treating the study period as discrete. Hence $\Delta q = q_{58} - q_{46}$. Output per head in 1958 was deflated by A(t) to obtain this variable net of technological change and the excess of this over the 1946 level was imputed to increased capital per head. Hence,

$$\Delta k = \frac{q_{58}}{A(t)_{58}} - q_{46} \qquad \text{(xxii)}$$

and the remainder of the increase is imputed to technological factors, $\Delta q - \Delta k$.

He found that for the overall economy the contribution of technological change was 67 per cent whilst the figures for the farm and nonfarm sectors were 93 per cent and 54 per cent, respectively. Once again the overriding importance of technological progress had been confirmed.

Solow's innovation was an ingenious one, whilst the studies of Massell, Lave, Chandler (and others) merely represent attempts to climb on his methodological bandwagon by applying his technique to different sectors, using differently-defined variables, and introducing other minor refinements. Nevertheless, his methodology fell far short of general acceptance, and was strongly attacked from some quarters. Some of the more-important arguments can be discussed:

(i) Solow's assumptions received considerable criticism; constant returns to scale, factors being paid their marginal product, neutral and disembodied technological progress, and aggregate production functions being the major assumptions. Hogan is particularly severe. In reply (17), Solow is unrepentant about such liberties. Concerning the assumption of an aggregate production function he counters that he does not want to argue the case for and against the meaningfulness of such a concept (with or without marginal productivity), dismissing Hogan's criticism with an "art of the possible" defence,

> "...most economists have two compartments in their minds, one for rigorous economic theory, and one for empirical compromises. It is obvious in which compartment the notion of an aggregate production function belongs."

Defending his decision to work in terms of constant returns to scale, this again was determined simply because he "had to assume that factor returns measured marginal products and because the data always shows factor shares adding up to one". If returns to scale were non-constant there would have to be a positive or negative residual factor share each year. One could regard the factor shares as imperfect estimates of the "true" shares which need not add up to one. If so, the analysis would be unchanged except at the final stage: instead of assuming constant returns to scale and fitting $q/A = f(k)$, one could fit $X/A = F(K,L)$ and test for constant return to scale. Solow subsequently

He found that for the overall economy the contribution of technological change was 67 per cent whilst the figures for the farm and nonfarm sectors were 93 per cent and 54 per cent, respectively. Once again the overriding importance of technological progress had been confirmed.

Solow's innovation was an ingenious one, whilst the studies of Massell, Lave, Chandler (and others) merely represent attempts to climb on his methodological bandwagon by applying his technique to different sectors, using differently-defined variables, and introducing other minor refinements. Nevertheless, his methodology fell far short of general acceptance, and was strongly attacked from some quarters. Some of the more-important arguments can be discussed:

(i) Solow's assumptions received considerable criticism; constant returns to scale, factors being paid their marginal product, neutral and disembodied technological progress, and aggregate production functions being the major assumptions. Hogan is particularly severe. In reply (17), Solow is unrepentant about such liberties. Concerning the assumption of an aggregate production function he counters that he does not want to argue the case for and against the meaningfulness of such a concept (with or without marginal productivity), dismissing Hogan's criticism with an "art of the possible" defence,

> "...most economists have two compartments in their minds, one for rigorous economic theory, and one for empirical compromises. It is obvious in which compartment the notion of an aggregate production function belongs."

Defending his decision to work in terms of constant returns to scale, this again was determined simply because he "had to assume that factor returns measured marginal products and because the data always shows factor shares adding up to one". If returns to scale were non-constant there would have to be a positive or negative residual factor share each year. One could regard the factor shares as imperfect estimates of the "true" share, which need not add up to one. If so, the analysis would be unchanged except at the final stage: instead of assuming constant returns to scale and fitting $q/A = f(k)$, one could fit $X/A = F(K,L)$ and test for constant return to scale. Solow subsequently performed this test and found that "while increasing returns to scale are not ruled out, they are not especially indicated".

Finally, the assumption of neutral technological progress can be tackled by initially assuming away the possibility of non-neutral technological progress existing in the data and later testing the accuracy of this assumption. Solow's test for non-neutrality is to scatter the proportional changes in the measured function against the capital-labour ratio and if there is no relationship then he concludes technological progress to be neutral, on average, i.e. the shifts net out to be approximately neutral. However, Solow admits in his reply to Hogan that "it would be wrong to interpret this as an assertion that each and every shift was a neutral one". Even so, the capital-labour ratio could change in such a way as to allow the proportional changes in the function to be zero and still there might be non-neutrality. Correspondingly, Solow has to conduct his defence along the lines that he found "no evidence" of strong or persistent biases in the production function shifts and if there were it would take rather special behaviour of the capital-labour ratio to disguise it. Since there is no a priori reason to expect the ratio to behave in this way it seems justified to regard the shifts as, on balance, neutral, and to treat them as individually neutral.

(ii) In common with the Ratio Method, the magnitude of A is completely divorced from investment and capital accumulation. "Capital merely accumulates" criticises Domar, "it does not change its quality, form or composition; it does not serve as the instrument for the introduction of technological change into the productive process. It is this kind of capital accumulation (wooden ploughs piled up on top of existing wooden ploughs) that contributes so little to economic growth". Domar adds, of course, that a complete isolation of capital formation from technological progress is empirically impossible. As in the Ratio Method, also here, that because quality changes are not fully accounted for, both $\dot{K}$ and $\dot{L}$ are understated and hence $\dot{A}$ overstated. Solow is, of course, well aware of these points. Massell takes further the point of embodied progress. He argues that the large contribution of technological progress would have been smaller had capital formation not occurred, i.e. although much technological progress consists of organisational changes which require no new inputs performed this test and found that "while increasing returns to scale are not ruled out, they are not especially indicated".

Finally, the assumption of neutral technological progress can be tackled by initially assuming away the possibility of non-neutral technological progress existing in the data and later testing the accuracy of this assumption. Solow's test for non-neutrality is to scatter the proportional changes in the measured function against the capital-labour ratio and if there is no relationship then he concludes technological progress to be neutral, on average, i.e. the shifts net out to be approximately neutral. However, Solow admits in his reply to Hogan that "it would be wrong to interpret this as an assertion that each and every shift was a neutral one". Even so, the capital-labour ratio could change in such a way as to allow the proportional changes in the function to be zero and still there might be non-neutrality. Correspondingly, Solow has to conduct his defence along the lines that he found "no evidence" of strong or persistent biases in the production function shifts and if there were it would take rather special behaviour of the capital-labour ratio to disguise it. Since there is no a priori reason to expect the ratio to behave in this way it seems justified to regard the shifts as, on balance, neutral, and to treat them as individually neutral.

(ii) In common with the Ratio Method, the magnitude of A is completely divorced from investment and capital accumulation. "Capital merely accumulates" criticises Domar, "it does not change its quality, form or composition; it does not serve as the instrument for the introduction of technological change into the productive process. It is this kind of capital accumulation (wooden ploughs piled up on top of existing wooden ploughs) that contributes so little to economic growth". Domar adds, of course, that a complete isolation of capital formation from technological progress is empirically impossible. As in the Ratio Method, also here, that because quality changes are not fully accounted for, both $\dot{K}$ and $\dot{L}$ are understated and hence $\dot{A}$ overstated. Solow is, of course, well aware of these points. Massell takes further the point of embodied progress. He argues that the large contribution of technological progress would have been smaller had capital formation not occurred, i.e. although much technological progress consists of organisational changes which require no new inputs a much greater proportion is embodied in new capital goods. Thus technological progress is strongly influenced by the rate of capital formation both in the form of replacements of old machinery with better machinery, and in additions to the size of the capital stock.

(iii) Again in common with the Ratio Method, the technological-change variable A is a residual, as acknowledged by Solow. It is a "catch-all" incorporating any kind of shift in the production function - "slowdowns, speed-ups, improvem-nts in the education of the labour force", all appear as technological change. In the words of Domar "it absorbs like a sponge all increases in output not accounted for by the growth of explicitly-recognised inputs. It is not the input into technological progress even in the broadest sense; we do not as yet know the nature and the magnitude of inputs which would result in a given increment in A".

(iv) Also in common with the Ratio Method, the relative size of the weights crucially affects the final conclusion. Domar demonstrates why the contribution of technological progress is so large relative to capital. Re-writing expression (xvii) as

$$\bar{A} = \bar{X} - \gamma\bar{L} - \alpha\bar{K} \qquad \text{(xxiii)}$$

he takes $\bar{L}$ and $\bar{K}$ as 1,5 and 3,0 per cent per annum respectively and γ and α as 75 and 25 per cent respectively, closely paralleling United States' data. Without the residual, $\bar{X}$ is a weighted mean of $\bar{L}$ and $\bar{K}$ and due to labour's larger weight it is much closer to $\bar{L}$. In the example it equals 1,9 per cent and since actual $\bar{X}$ approximated 3,5 per cent, the remainder (1,6 per cent) is equal to $\bar{A}$. The growth of output per head being 2 per cent (3,5 - 1,5) the ratio of $\bar{A}$ to this figure is 80 per cent, closely approximating Solow's findings. With a weight of only 25 per cent there is little that capital can do. Even if $\bar{K}$ should double, ceteris paribus, $\bar{X}$ rises only from 3,5 to 4,2 per cent - modest reward for a major effort.

(v) A particularly disturbing factor is that the approach operates on such a high aggregation level that it confounds changes in the composition of output and changes in the production function. Only the latter should be measured, but if the economy utilised more intensively sectors with comparative advantages, but there is no technological progress, the overall productivity measure would increase despite unchanged production functions in each individual sector. But all aggregates bear this difficulty, including the Ratio Method.

(vi) Levine (18) has pointed out the arbitrariness of the method in that one measure is obtained if the technical-change contribution is calculated first leaving the remainder to capital and another different measure with the reverse procedure. Massell (19) points out that an assumption of exponential growth of technological progress and labour productivity nullifies this criticism.

Further Comments on the Ratio and Solow Methods

By way of conclusion, several similarities between the Ratio and Solow Methods are immediately noticeable, of which the most significant can be mentioned.

Firstly, they attained prominence mainly because very little prior knowledge of the production process was required. The functional form relating inputs to output is not known. If it was, it would be possible to exactly specify the production function, thus creating an exact and unambiguous method for measuring productivity by studying its shift pattern over time. However, in the absence of such perfect knowledge the Ratio and Solow Methods provide a means to measure productivity change whilst avoiding the problem of simultaneously deriving the production function. However, the argument that these methods need not make any assumptions about the production function is, strictly speaking, not true. The concept of an aggregate production function is specified implicitly rather than explicitly. Kendrick's arithmetic measure implies a linear production function in inputs. Solow's geometric index implies a Cobb-Douglas type (as also does the Abramovitz-Domar approach). However, what is meant by saying that they make no assumptions about the underlying production function is that the co-efficients of the implied function are allowed to vary. Solow employs a system of weights which changes every year. Kendrick also employs a shifting system but not on a yearly basis.

Secondly, Levhari, Kleiman and Halevi (20) have demonstrated the relationship between the two methods. The equivalence between Solow's expression (xvii) and the Abramovitz-Domar expressions (viii) and (ix) has already been shown. Additionally, Kendrick's expression (iii) is also equivalent for small changes in the quantities of inputs and outputs.

Thirdly, technological change is imputed as a residual after conventional measures of factor inputs have been purged from output increases and, hence, is necessarily an amalgamation of many unexplained elements, such as economies of scale, improvements in input qualities, improvements in the organisation of production, improvements in the efficiency of resource allocation, and "pure" technological progress.

Fourthly, all the studies reviewed are unanimous in their finding that the technological-change residual comprises the major component of observed increases in labour productivity.

Fifthly, both methods by their very nature are incapable of "decomposing" technological change. In particular, they cannot differentiate varying returns to scale (economies and diseconomies of scale), nor can they differentiate the relative contributions of neutral and non-neutral technological progress. The assumptions of constant returns to scale, perfectly-competitive product and factor markets, and neutral, disembodied technological progress are stipulated in the functioning of both methods.

Production Function Method

An alternative approach to the implicit production-function approaches of the Ratio and Solow Methods is to explicitly employ the concept of an aggregate production function in the analysis of technological progress. It was mentioned in chapter six that if one had complete knowledge of the production process it would be possible to specify exactly the functional form relating inputs to output. The production function could be fully specified in form (implicitly taking into account economies of scale) and in conventional inputs - labour, capital, and raw materials - and the entire problem of measuring productivity is solved by studying the shift pattern of such a function. However, in the real world of imperfect knowledge the only practical alternative is to employ the approximating functional forms represented by the Cobb-Douglas and the Constant-Elasticity-of-Substitution functions.

(I) Cobb-Douglas

In order to utilise the Cobb-Douglas function for the measurement of technological progress it is necessary to be familiar with its properties and these were summarized in chapter six. These properties assume a given and unchanging technology. One of the major advantages of the Cobb-Douglas function for measuring technological progress, in comparison with the Ratio and Solow Methods, is that it is capable of a "decomposition" of the technical-change residual, at least to take account of neutral and non-neutral technological progress and economies or diseconomies of scale. This is achieved by means of variations in the parameters of the function (A, σ and γ) to give expression to technological changes. There are three ways in which such changes are depicted in the Cobb-Douglas function:

- An all-round increase in efficiency which transforms similar physical amounts of inputs into a larger output is reflected by an increase in the scaling parameter A. Ceteris paribus, a proportional change in A produces a proportional change in output. The marginal rate of substitution between capital and labour is not affected so this represents a neutral technological change.

- The degree of returns to scale is indicated by $\sigma + \gamma$ and these can vary either as a result of changes in the scale of operations or changes in technology. Abstracting from the former and focusing only on the latter, an increase in the value of $\sigma + \gamma$ represents a neutral technological advance <u>if</u> the ratio of σ/γ remains unaltered (i.e. equal proportional algebraic changes in σ and γ) because the marginal rate of substitution remains unaltered.

- The only way in which a non-neutral technological change can be represented is by a change in the ratio of σ to γ . Obviously this alters the $MRTS_{LK}$ for each K:L ratio. The direction of change of the ratio σ/γ indicates either factor-saving or factor-using technological progress. If σ rises relative to γ then a labour-saving (capital-using) gain has occurred since capital's marginal product has risen relative to that of labour for each K:L combination. Conversely, a fall in σ relative to γ indicates that a less capital-intensive technology has been utilised.

It will be noted that of the four characteristics of technological change discussed in chapter six, only three can be analysed by a Cobb-Douglas function. This is because the value of the elasticity of substitution is always constrained to unity, and, hence, is invariant to technological progress.

However, despite the powers of the function to "break open" the technological-progress residual, early studies using the Cobb-Douglas function did not focus on this aspect of decomposition. Instead the authors seemed hypnotised into employing an "output-per-unit-of-input" approach in the best manner of the precedents set by the proponents of the Ratio and Solow Methods. Their objective seemed to be merely an attempt to isolate and measure the size of the technological-progress residual by a method of adding a trend term to an aggregate Cobb-Douglas function, and applying it to time-series data. Some of these studies can be examined. The values of the relevant parameters are, of course, obtained through econometric estimation.

In his numerous studies (invariably with co-authors) published during the 1920's, 1930's and 1940's, Douglas either specified or found constant returns to scale. He assumed an invariant technology, so that observed advances in labour productivity were caused, ceteris paribus, by the substitution of capital for labour. The first attempt to measure technological progress utilising a Cobb-Douglas function was made by Tinbergen in 1942 (21). He advanced the work of Douglas by introducing an additional source of productivity change - a rise in "efficiency". This he defined as the situation where the same physical quantities of labour and capital produce a higher output volume. By its nature, therefore, Tinbergen was restricting his measure to neutral technological progress. He attempted to capture the efficiency component by the addition of an exponential growth term $e^{\delta t}$ to the Cobb-Douglas function, i.e.,

$$X = AK^{\sigma}L^{\gamma}e^{\delta t} \quad . \tag{xxiv}$$

Tinbergen specified, a priori, constant returns to scale using the values $\gamma = 0,75$ and $\sigma = 0,25$ (roughly corresponding to Douglas' findings) and was able to estimate the productivity co-efficient by fitting time-series data to the logarithmic conversion of the function.

A similar approach was employed by Lomax (22). He attempted to determine a production function for the coal mining industry in Britain over the period 1927 - 1943 by applying multi-variate analysis to time-series data. He included a residual time trend to take account of productivity advances. However, by virtue of his definition of capital (the amount of coal cut by machinery taken together with that obtained independently by pneumatic picks) this implies that the residual trend is primarily connected with such factors as changes in skill, intensity of labour effort, and the state of seams being worked, rather than with technological progress. Lomax employed the Cobb-Douglas function because it provided an equation linear in time and in the logarithms of the variables. As already seen, input elasticities of output are then constant and the residual trend is a constant proportionate change per unit of time.

Lomax found $\gamma = 0,79$ and $\sigma = 0,29$. This indicates a tendency towards increasing returns to scale, but the sum of γ and σ was insignificantly different from unity at the 1-per-cent level, and only on the borderline of significance at the 5-per-cent level. This did not justify an assumption of a significant departure from constant returns to scale. The residual trend declined at an average 1,5 per cent per annum. This figure indicates the productivity fall independent of changes in capital and labour as defined by Lomax. Thus, it represents the output decline which would have occurred with an unchanged number of manshifts and a constant mechanically-obtained tonnage of coal.

Lomax also employed the identical technique in an earlier paper (23) for United Kingdom agriculture over the period 1924 - 1947. In addition to labour and capital he also experimented with land and fertilisers and feeding stuffs as explanatory inputs but these were finally excluded. He found $\gamma = 0,18$ and $\sigma = 0,37$, contradicting the Cobb-Douglas findings, and strongly suggesting decreasing returns to scale. The residual trend indicated an average annual rate of technological progress of 1,03 per cent.

The tradition of using an aggregate Cobb-Douglas production function with an added time trend to capture constant productivity advance was also employed by Aukrust (24) in his study of the Norwegian economy over the period 1900 to 1955. He pursued the theme of dispelling the myth that if one desired an increase in the economic growth rate then actions had to be concentrated on increasing the level of investments. He believed the relationship between investment and production increases was far more complicated than generally assumed. Specifically, his objective was to show that what he called the "human factor" (organisation, professional skills and technical knowledge) was at least as important to the rate of economic growth as the volume of physical capital. To show this he incorporated a factor he called "organisation" into the production function which included a conglomerate of elements such as the technical and commercial knowledge of managers, employees, and workers, their qualities as leaders, their will and ability to work, the whole social setting, and the international situation in which production takes place. In other words, "organisation" includes all elements which, together with labour and capital, determine what the results of the productive process will be. It is by implication, a residual. His method of incorporating this factor was not original and differed in no material way from the pioneering study of Tinbergen seventeen years previously. He employed the Tinbergen convention that "organisation" had increased at a constant rate and, hence, could be brought into the production function in the form of an exponential trend.

Whilst not restricting his model to constant returns to scale he found $\gamma = 0,76$ and $\sigma = 0,20$. The "organisation-improvement" trend he found to be at an average rate of 1,81 per cent per annum. He was then able to split total growth according to its causes. During the period 1948 - 1955 employment increased at an average rate of 0,6 per cent per annum whilst the figure for capital was 5,6 per cent. Hence,

growth caused by :	
increased employment :	0,76 x 0,6 = 0,46 per cent per annum
increased capital :	0,20 x 5,6 = 1,12 per cent per annum
improved organisation:	= 1,81 per cent per annum
Total growth rate	= 3,39 per cent per annum

which correlated exactly with actual growth during the period. Aukrust, therefore, claims to have shown that there is a trend factor which, together with increases in capital and labour, explains the increases in output which occurred during the period of study. Furthermore, this trend factor seems to be important compared with the effect of increases in real capital. He described it as the "driving force" in the process of economic growth to such an extent that growth could be considerably increased if new efforts were made in the fields of education and scientific research. Instead of concentrating on increasing real capital, the rate of progress could be more-meaningfully enhanced by a conscious effort to improve man himself.

Several points can now be made by way of an analysis of the Tinbergen-Lomax-Aukrust approach towards measuring technological progress using the Cobb-Douglas function. In essence, the Tinbergen convention of constraining the model to constant returns to scale and expressing all technological progress in an exponential form represents no real measurement improvement over the Ratio and Solow Methods. Even the unconstrained models of Lomax and Aukrust are similarly deficient. Some significant points are worth discussing. Firstly, technological change is still conceptually treated as a residual and is interpreted as an amalgamation of the same set of "unexplained" elements already mentioned under the Ratio and Solow Methods.

Secondly, these studies continue to ascribe the major component of labour-productivity advances to technological progress.

Thirdly, decomposition of technological progress is still ruled out. All productivity advances are of the neutral type. Although the unconstrained model allows a dichotomy of increasing and decreasing returns to scale there is a difficulty in distinguishing between types of scale economies, i.e. between economies that are inspired by technological progress and those that result simply from an increase in physical inputs.

Fourthly, disembodied technological progress is still stipulated.

Fifthly, when the model is applied to aggregate data it does not permit a distinction between technological progress and factor movements from a less-productive to a more-productive sector. Such aggregation bias suggests that only firm production functions, and perhaps industry production functions, may be the only feasible ones.

Sixthly, there are specific problems inherent in the Cobb-Douglas function itself. It is deficient in that it specifies a unitary elasticity of substitution which for certain purposes involves a serious specification error. In addition, there are problems of a statistical nature inherent in this function - the high correlation between capital and labour, identification of parameter estimates, derivation of consistent parameter esimates, and so on. These can only be mentioned in desultory fashion but are extensively analysed in good econometrics text-books, particularly Walters (25), Wallis (26), Klein (27), Cramer (28), and Leser (29). Econometric problems associated with the estimation of parameter co-efficients can be disturbing in the Cobb-Douglas context but they are common to a greater or lesser degree in much econometric time-series analysis. There are strong trends in the output and input series. Since there is a high correlation between capital and labour (multicollinearity) in the time-series data, this confuses the magnitude of the co-efficients on individual variables, and accordingly it is difficult to identify the co-efficient of the trend term as a measure of the rate of growth of total factor productivity. Ideally, estimation should be achieved in the context of a simultaneous-equation system (with labour and capital as separate dependent variables) employing a method such as two-stage least squares in order to rid the bias inherent in single-equation least-squares estimation. The lack of identifiability of the Cobb-Douglas function in the standard competitive model can be overcome in certain circumstances. If an industry has to supply, on demand, a product which cannot be stocked, then output, from the firm's point of view, is no longer an endogenous variable. It become pre-determined. This can happen in the case of public utilities. Klein (30) studied railroad transportation, whilst Nerlove (31) studied electricity supply. The former worked from a structural equation which is identified, namely the marginal-productivity conditions, whilst the latter employed the alternative approach of estimating the parameters by working through the reduced form, namely the cost-minimising input functions.

The question has certainly been raised of whether the Cobb-Douglas function contains too many specification errors to be useful for estimating technological progress. Whilst it cannot be denied that the function contains restrictive properties, whether these are prohibitively limiting for this specific purpose is debatable.

Of additional concern is a matter which has been touched on before, namely the whole concept of an aggregate production function. This is of relevance not only for the Cobb-Douglas approach, but the Ratio and Solow Methods as well. Such a concept, according to Mrs. Robinson (32), has acted as a "powerful instrument of miseducation". The production function is a micro-economic concept describing the technical substitutability between capital and labour in the production process, but it is debatable whether such technical relations can be measured by an aggregate relation between output, capital and labour in value terms. The extreme heterogeneity of inputs and outputs creates problems for aggregation. Conventionally, only two inputs, capital and labour, are recognised although in practice there are many varieties of capital and labour inputs. Ideally, all sub-categories of capital and labour should be introduced as arguments in the production function, but this is avoided, in practice, by combining them into single aggregate variables. Wallis (33) reminds us that this is only valid,

> "...provided that the marginal rate of substitution between any two kinds of one factor is independent of any variety of the other factor, and these aggregate variables can be treated as if they were actual individual inputs provided that they are linear homogeneous functions of the different varieties".

Accordingly, aggregation problems are avoided only at the extreme micro level of one man with one machine. They arise even at the level of the individual firm long before the question of aggregating equations in industry studies arises.

Although additive aggregation of outputs and inputs is practiced, the Cobb-Douglas function is multiplicative. This anomaly led Simkin (34) to point out that if aggregates are arithmetic sums a multiplicative aggregate function cannot be derived because logarithmic addition implies continued products of individual outputs and inputs. Domar's suggestion (35) was to take geometric averages of inputs and outputs which will generally give a different value for the technological-progress parameter. Massell (36), however, defends the measurement inconsistency on the grounds that it is necessary to discover to what extent inter-industry shifts explain the growth of total productivity, and whilst additive aggregation can isolate them, the multiplicative approach cannot.

(II) Constant-Elasticity-of-Substitution

The CES production function was presented in chapter six, and some of its properties were examined. It has recently become a popular tool in research work in preference to Cobb-Douglas. The latter's constraint of a unitary elasticity of substitution has been considered too restrictive, whilst the CES function avoids this by constraining the value to be a constant but not necessarily equal to unity. In common with Cobb-Douglas, but contrary to the Ratio and Solow Methods, the function is capable of "decomposing" the technical-change residual to take account of neutral and non-neutral technological progress, and economies or diseconomies of scale. This is achieved by means of variations in the parameters κ, v, δ and ε. There are four ways in which such changes are depicted in the CES function, following the analysis in chapter six:

- An all-round increase in efficiency is reflected by an increase in the scaling parameter κ. Since this is analagous to the parameter A in the Cobb-Douglas function it represents a neutral technological change by the same argument.
- The degree of returns to scale is indicated by the value of v and can vary either as a result of changes in technology or changes in the scale of operations. Since increases in v leave the $MRTS_{LK}$ unchanged, this also represents neutral technological progress.
- The parameter δ represents the degree to which technology is capital-intensive. If δ rises, then MP_K rises relative to MP_L for each K:L ratio. Thus, the change is capital-using. The opposite applies if δ falls. Hence, a change in δ represents non-neutral technological progress.
- The parameter ε represents the elasticity of substitution. If ε rises this means that capital and labour are becoming more similar and, hence, more substitutable for each other at each K:L ratio. Now if capital is growing more rapidly than labour, then capital will be substituted for labour - MP_K rises relative to MP_L - thus causing a capital-using change. However, if labour is growing more rapidly than capital, then labour will be substituted for capital - MP_L rises relative to

MP_K - thus causing a labour-using change. On the other hand, if ε falls this means that the factors are becoming more dissimilar and, hence, less substitutable. Because the value of ε is not constrained, a CES function can depict all four of the characteristics of technological change discussed in chapter six, whereas, as previously discussed, a Cobb-Douglas function can depict only three, since ε is constrained to unity.

At the outset the most important objective of CES studies was to test whether the elasticity of substitution differed significantly from the Cobb-Douglas value of unity. Unfortunately, the task of obtaining parameter estimates of the CES function is not straightforward because of the relatively unmanageable nature of the function. A simple logarithmic transform prior to confronting the function directly with data on output and inputs (as in the Cobb-Douglas case) is not possible, because of the non-linear way in which the parameter ρ enters. Accordingly, some econometric ingenuity is required before parameter estimates can be obtained. Three different methods are presented by Brown (37). The pioneering SMAC study of Arrow, Chenery, Minhas, and Solow (38) adopted the approach of estimating a side relation (in this case, labour's share) to induce an estimate of ε. Using least squares fitted to United States' non-farm production for the period 1909 - 1949, they estimated the elasticity of substitution to be 0,569, and concluded that there was strong evidence that the value of ε is between zero and unity.

Following SMAC's study there was a veritable plethora of reseach work published over the period 1961 - 1967, eagerly employing the new CES tool. The magnitude and diversity of these studies led Nerlove, in a review article in 1967 (39), to quote from Alice's Adventures in Wonderland to describe the haste of events,

> "this bottle was not marked 'poison', so Alice ventured to taste it, and finding it very nice, ..., she very soon finished it off".

It is not necessary to reference all these studies as they are adequately discussed by Nerlove. Of more importance for our purpose is the consensus reached regarding the value of the elasticity of substitution. A more-accurate statement, however, would be the lack of consensus. Wide and irreconcilable disparities of estimates of ε were observed taken from inter-country, inter-regional, inter-industry, cross-section, and time-series data, and between small differences in time periods. Results appeared to be extremely sensitive to variations in the specification of the fitted equation and to the data employed. One pattern which did emerge was that estimates of ε obtained from cross-section studies were generally larger than time-series estimates.

The meaningfulness of estimating the value of ε has, however, been questioned by Nelson (40). He observed that increases in the capital-labour ratio explained only a small fraction of productivity growth in the United States over the period 1945 - 1965. This low degree of explanation was not sensitive to the choice of any particular value of ε. Therefore, as long as ε did not differ markedly from unity, and the difference between the growth rates of capital and labour was not unrealistically wide, then there was little to choose between the CES and Cobb-Douglas functions. Since these conditions were more likely to be met in the short-and medium-term then the CES function had no advantage over the simpler Cobb-Douglas function over such time periods. However, over longer periods, or when capital grows more rapidly than labour, then deviations of ε from unity assume more significance.

Although estimates of the value of ε are illuminating it is our main objective in presenting the CES function to examine how it may be used for the measurement of technological progress. As with the Cobb-Douglas function a starting-point is provided by examining those studies which merely attempted to isolate and measure the size of the technological-progress residual by a method of adding a trend term to an aggregate CES function, and applying it to time-series data.

Diwan (41) follows the convention of adding a simple time trend to the CES function in order to capture neutral technological progress. He employs the general formulation of Brown and De Cani, which allows for varying degrees of returns to scale, and, hence, estimates the parameters of the function,

$$X = \kappa e^{\lambda t} \left[\delta K^{-\rho} + (1-\delta) L^{-\rho}\right]^{\frac{-v}{\rho}} \qquad \text{(xxv)}$$

for the manufacturing sector of the United States' economy over the period 1919 - 1958. Since the usual estimation problems arise, Diwan first estimated the parameters δ and ρ via the marginal-rate-of-substitution side relation, subsequently estimating λ and v through fitting equation (xxv) above. He found the elasticity of substitution to be significantly less than unity. There was evidence of increasing returns to scale. Neutral technological progress had proceeded at an average rate of 1,4 per cent per annum over the period. Diwan used several capital measures but no difference in results was observed. In an attempt to avoid the use of a capital measure he also estimated the elasticity of substitution by a labour productivity-wage relation, but did not attempt a measure of technological change by this technique.

In this regard, it can be shown (for instance by Leser (42)) that analysis of the CES function does not require capital data. Only data on output per manhour and the real wage rate are required. A regression in logarithmic terms of the former variable on the latter yields an estimate of the elasticity of substitution through the co-efficient on the wage variable, i.e.,

$$\log (X/L) = \text{constant} + \varepsilon \log \omega . \qquad \text{(xxvi)}$$

The wage rate is regarded as predetermined, and, hence, a simple least-squares regression can be employed to test whether the value of ε is close to the Cobb-Douglas assumption of unity.

Such a method was earlier followed by SMAC in their 1961 study (38), who employed international cross-section data for individual industries and found the value of ε to be significantly below unity.

In an attempt to measure the rate of technological progress, SMAC derived an estimable equation from the expression for labour's share, namely,

$$\omega L/PX = (1-\delta)^{\varepsilon} (\omega/\kappa)^{1-\varepsilon} = (1-\delta)^{\varepsilon} \kappa^{\varepsilon-1} \omega^{1-\varepsilon} \quad . \qquad \text{(xxvii)}$$

Under the assumption of neutral technological change (and not forgetting the SMAC formulation of $v = 1$, i.e constant returns to scale), the only parameter that varies is κ. If technological change proceeds at a constant geometric rate, so that $\kappa(t) = \kappa_0 10^{\lambda t}$, one can fit expression (xxvii) in logarithmic form, i.e.,

$$\log(\omega L/PX) = [\varepsilon \log(1-\delta) + (\varepsilon-1)\log \kappa_0] + (1-\varepsilon)\log \omega + \lambda(\varepsilon-1)t \qquad \text{(xxviii)}$$

which can be re-written

$$\log(\omega L/PX) = [\qquad a_0 \qquad] + a_1 \log \omega + a_2 t \ . \qquad \text{(xxix)}$$

Fitting equation (xxix) by least squares to data for United States' non-farm production for the period 1909 - 1949, estimates of a_1 and a_2 were obtained from which it is possible to solve for estimates of ε and λ from the relationships : $a_1 = 1-\varepsilon$, and $a_2 = -\lambda(1-\varepsilon)$. They estimated the elasticity of substitution to be 0,569 whilst the annual rate of growth of technological progress was calculated at 1,83 per cent.

An identical approach was adopted by Ferguson (43) in a later study. His proposed methodology was to distinguish between two CES functions, one representing Hicks-neutral technological progress, and the other being Harrod-neutral. However, this was not pursued after he discovered that they both resulted in identical regression equations, giving no grounds in statistical theory for choosing between the alternative hypotheses of Hicks-or Harrod-neutral technological progress. Working with the SMAC formulation of the CES function (constant returns to scale) with neutral technological progress captured by a time trend, namely,

$$X = \kappa e^{\lambda t} [\delta K^{-\rho} + (1-\delta) L^{-\rho}]^{-\frac{1}{\rho}} \qquad \text{(xxx)}$$

he obtains the familiar regression model already examined, namely,

$$\log (X/L) = b_0 + b_1 t + b_2 \log \omega \qquad \text{(xxxi)}$$

where ω is the wage rate. The b's can be estimated through least squares from which ε and λ can be obtained from the relations, $\varepsilon = b_2$ and $\lambda = b_1/(1-b_2)$. Employing United States' manufacturing data over the period 1929 - 1963, Ferguson found the elasticity of substitution to be less than unity (0,67), and technological progress to have proceeded at an average annual rate of 1,5 per cent. Over the shorter period 1948 - 1963, the value of ε rose to in excess of unity (1,16) with technological progress rising to 1,9 per cent per annum. Over both the long and the short periods technological progress accounted for more than 90 per cent of the increase in output per man.

Moving to the industry level from the aggregate level, two important early studies were performed by Ferguson (44) and McKinnon (45). Ferguson fitted the familiar regression equations (xxvi) and (xxxi) to time-series data covering two-digit United States' manufacturing industries over the period 1949 - 1961 in order to estimate the elasticity of substitution and the rate of neutral technological progress.

He found a wide diversity in the elasticity of substitution between industries, but on the whole his estimates were high. They varied from a low of 0,24 to a high of 1,3. Of the nineteen industries, nine had a value of ε below unity and the remaining ten had a value above unity. Statistically, however, the majority of industries had a value insignificantly different from unity. Three industries had a value insignificantly different from zero (indicating that the CES function reduces to a Leontief-type, fixed-proportions function); twelve industries had a value insignificantly different from unity (indicating that the CES function reduces to a Cobb-Douglas form); and four industries had a value statistically greater than unity (indicating that the CES general formulation is the only suitable model). Certainly Ferguson was able to conclude from these results that there were no grounds for rejecting the Cobb-Douglas hypothesis of unitary elasticity, thus contradicting the findings of several aggregate, time-series studies, which found it to be substantially less than unity.

In regard to neutral technological progress in these industries, Ferguson found it to be remarkably low. In only three cases did it exceed an average of one per cent per annum. In eleven of the industries the rate was 0,3 per cent per annum or less, and in seven of the cases was zero. These were materially below the rates for the aggregate economy as reported in previous studies. A reconciliation is attempted by Ferguson. He found some industries had experienced sustained capital-using progress, whilst others experienced capital-saving progress, and still others experienced both in roughly offsetting proportions. Accordingly, aggregation of these opposite tendencies for the manufacturing sector appears to result in neutral technological progress well in excess of his computed rates.

McKinnon also worked via the regression equations (xxvi) and (xxxi) but built a distributed lag into his analysis. The lagged effects of adjusting the actual (ex post) output-labour ratio to changes in wages depends on the speed with which it is profitable to substitute capital which is ex ante optimal for that already installed. Such adjustment effects are probably less regular and more noticeable if wage changes are sudden and discrete, due, perhaps, to union negotiations. Equation (xxxi) is an ex ante equilibrium relationship between X/L, ω and t, but the relationship is not directly observable because actual (ex post) X/L cannot adjust immediately to shifts of the other variables. To account for this, McKinnon incorporates a lagging procedure suggested by Nerlove (46) whereby in response to a sudden increase in wages, X/L moves to a new equilibrium position following an asymptotic path, rapidly at first and then more slowly. This implies that some of the firm's operations are immediately amenable to less labour intensity, whilst others are completely inflexible in the short-run and must wait for old equipment to depreciate. McKinnon derives the estimable equation,

$$\log (X/L)_t = \pi_0 + \pi_1 \log \omega_t + \pi_2 t + \pi_3 \log(X/L)_{t-1} \quad . \qquad \text{(xxxii)}$$

From the estimated values of the π's, b_1 and b_2 in expression (xxxi) can be solved from the relationships

$\pi_1/(1-\pi_3) = b_2$, and $\pi_2/(1-\pi_3) = b_1$.

The equation is fitted to time-series data covering eighteen two-digit United States' manufacturing industries over the period 1947 - 1958. From the solved values of b_1 and b_2, the elasticity of substitution and the rate of neutral technological progress can be estimated via the relationships presented earlier. His estimated values of ε were generally lower than those of Ferguson, and lying between zero and unity. His estimated rates of neutral technological progress were substantially higher than Ferguson's. These results, in contradicting Ferguson's, are more in line with the traditional findings of aggregate time-series studies.

Having examined these CES studies several points can now be made. In a survey of the attempts to estimate the parameters of the CES function, Nadiri (47) contended that with all its merits the function is found to be subject to certain important shortcomings. Some authors contend that the form of the function is not sufficiently flexible to adequately identify the sources of factor productivity. Others have emphasized that the fault lies in the inadequacy of estimation techniques. Empirical evidence has indicated that the CES parameters are highly sensitive to slight changes in the data, measurement of variables, and methods of estimation. The most crucial aspect concerns the value of the elasticity of substitution. If the value is insignificantly different from unity there appears to be no material advantage in employing the more-complex CES function in preference to the simpler, more-easily estimated, Cobb-Douglas function. However, evidence shows that the value of ε varies considerably for different sets of data, countries, industries, and levels of aggregation, as well as being sensitive to cyclical fluctuations of demand. The only tentative conclusion possible is that most time-series estimates of ε are below unity, whilst cross-section estimates are generally higher and close to unity.

Nadiri identifies three problems which seem to be responsible for the instability and inconsistency of the estimated parameters of the function:

(i) The basic difference between the time-series and cross-section input-output relations.

(ii) The parameters of the function often vary together and their separate effects cannot be identitied (the "identification problem") except under restrictive conditions and unless more information about the production process is available.

(iii) Estimation problems due to the simultaneity and non-linearities between the production function and the marginal productivity conditions.

These problems are extensively and excellently analysed by Nadiri and there is no need to repeat his arguments to understand the potential magnitude of the problem.

Turning now to another aspect of the function, an important dissimilarity between the Cobb-Douglas and CES function occurs in relation to the aggregation problem. The hope is that the aggregate production function can be interpreted in the same way as a micro production function, and this only occurs when it possesses the same broad form and properties of the micro functions from which it is derived. Walters (48) adds the reminder that an aggregate production function can be defined if, and only if, the micro functions are of a form that enables output to be broken down into two components, one due to labour and one to capital, with no interaction term reflecting, for example, the multiple of the two factors. The function, in terms of its factor components, is additively separable. Whilst Walters shows that the Cobb-Douglas function does not satisfy the aggregation conditions (no amount of juggling can put it into additively separable form), on the other hand, the CES function is additively separable. This enables a sensible aggregate function to be defined and hence it is clearly superior to the Cobb-Douglas on this score. However, whether the Cobb-Douglas is useless for fitting to aggregate data revolves on two points. Firstly, the size of the aggregation errors, which may or may not be significant. Secondly, employing an averaging or aggregating method (i.e. geometric rather than arithmetic means) that minimises these errors. In practice, geometric aggregates or means are rarely available, so that the argument revolves around the relative variation of the geometric and the arithmetic totals.

Conclusion on "Unrefined" Ratio, Solow, and Production Function Studies.

The studies examined so far in this chapter have in common a set of similarities which have been discussed several times on an on-going basis according to the context. Accordingly, these studies will be defined as "unrefined".

Many of the problematical points inherent in these unrefined studies were taken up and analysed in subsequent research. These later studies can be defined as "refined". In particular, it is worth mentioning that the overwhelming importance ascribed to technological progress and the demotion of the role of capital in the growth process was increasingly questioned in subsequent work. In order to better understand the process of economic growth the relative contributions of the unexplained elements of the "residual" and the "exponential trend" had to be studied. Subsequent research did attempt to refine the catch-all technological change category to distinguish among several component parts. Apart from these attempts at decomposition, further research on refining and improving the measurement of technological progress was undertaken. A school of thought began to originate that the excessive contribution of technical progress was due to the fact not only that it was a catch-all residual encompassing many unexplained factors (substitution of capital for labour, economies of scale, resource shifts, improvements in factor quality, and so on) but was also due to mistakes, errors of measurement, omission of some variables, use of wrong weights, use of wrong index-number formula, specification errors of the functional form, and so on. Attempts to place the contribution of technological progress in perspective led to subsequent studies which aimed at "correcting" for the above factors. These refined studies invariably led to a substantial reduction in the conventional indices of total factor productivity, and the more significant of these will be examined in chapter eight.

REFERENCES

(1) Kennedy, C., and Thirlwall, A.P., "Surveys in Applied Economics: Technical Progress", Economic Journal, March 1972.

(2) Barger, H., and Schurr, S.H., The Mining Industries, 1899 - 1939: A Study of Output, Employment and Productivity, National Bureau of Economic Research, New York, 1944.

(3) Barton, G.T., and Cooper, M.R., "Relation of Agricultural Production to Inputs", Review of Economics and Statistics, 1948.

(4) Schmookler, J., "The Changing Efficiency of the American Economy, 1869 - 1938", Review of Economics and Statistics, August 1952.

(5) Kendrick, J.W., "Productivity Trends : Capital and Labour", Review of Economics and Statistics, August 1956.

(6) Kendrick, J.W., Productivity Trends in the United States, N.B.E.R., Princeton University Press, 1961.

(7) Domar, E.D., "On Total Productivity and All That", Journal of Political Economy, December 1962.

(8) Fabricant, S., "Basic Facts on Productivity Change: An Introduction", in Kendrick, J.W., Productivity Trends in the United States, op cit.

(9) Abramovitz, M., "Resource and Output Trends in the United States since 1870", American Economic Review, Papers and Proceedings, May 1956.

(10) Reddaway, W.B., and Smith, A.D., "Progress in British Manufacturing Industries in the Period 1948 - 54", Economic Journal, March 1960.

(11) Solow, R.M., "Technical Change and the Aggregate Production Function", Review of Economics and Statistics, August 1957.

(12) Massell, B.F., "Capital Formation and Technological Change in United States' Manufacturing", Review of Economics and Statistics, May 1960.

(13) Lave, L.B., "Empirical Estimates of Technological Change in United States' Agriculture, 1850 - 1958", Journal of Farm Economics, November 1962.

(14) Chandler, C.A., "The Relative Contribution of Capital Intensity and Productivity to Changes in Output and Income in the U.S. Economy, Farm and Non-Farm Sectors, 1946 - 1958", Journal of Farm Economics, May 1962.

(15) Hogan, W.P., "Technical Progress and Production Functions", Review of Economics and Statistics, November 1958.

(16) Domar, E.D., "On the Measurement of Technological Change", Economic Journal, December 1961.

(17) Solow, R.M., "Reply to Hogan", Review of Economics and Statistics, November 1958.

(18) Levine, H.S., "A Small Problem in the Analysis of Growth", Review of Economics and Statistics, May 1960.

(19) Massell, B.F., "Another Small Problem in the Analysis of Growth", Review of Economics and Statistics, August 1962.

(20) Levhari, E., Kleiman, E., and Halevi, N., "The Relationship Between Two Measures of Total Productivity", Review of Economics and Statistics, August 1966.

(21) Tinbergen, J., "On the Theory of Trend Movement", Weltwirtschaftliches Archiv, May 1942.

(22) Lomax, K.S., "Coal Production Functions for Great Britain", Journal of the Royal Statistical Society, 1950.

(23) Lomax, K.S., "An Agricultural Production Function for the United Kingdom, 1924 to 1947", The Manchester School of Economic and Social Studies, May 1949.

(24) Aukrust, O., "Investment and Economic Growth", Productivity Measurement Review, February 1959.

(25) Walters, A.A., Introduction to Econometrics, Macmillan, London, 1970, 2nd Edition, chapter 10.

(26) Wallis, K.F., Topics in Applied Econometrics, Gray-Mills Publishing Ltd., 1st Edition, 1973, chapter 2.

(27) Klein, L.R., An Introduction to Econometrics, London, Prentice-Hall, 1962, pages 84-111.

(28) Cramer, J.S., Empirical Econometrics, Amsterdam, North-Holland Publishing, 1971, chapter 10.

(29) Leser, C.E.V., Econometric Techniques and Problems, 1974, 2nd edition, Griffin, London, chapter 5.

(30) Klein, L.R., A Textbook of Econometrics, Evanston: Row, Peterson, 1953, pages 226-236.

(31) Nerlove, M., "Returns to Scale in Electricity Supply", in Christ, C.F., et al, Measurement in Economics, Stanford University Press, 1963, pages 167-198.

(32) Robinson, J., "The Production Function and the Theory of Capital", Review of Economic Studies, Vol XXI, 1953-4.

(33) Wallis, K.F., Topics in Applied Econometrics, op cit., pages 25-6.

(34) Simkin, C.G.F.,"Aggregate Production Functions", Economic Record, May 1955.

(35) Domar, E.D., "On the Measurement of Technological Change", op cit.

(36) Massell, B.F., "Aggregative and Multiplicative Production Functions", Economic Journal, March 1964.

(37) Brown, M., On the Theory and Measurement of Technological Change, Cambridge University Press, 1966, chapter 9.

(38) Arrow, K.J., Chenery, H.B., Minhas, B.S., Solow, R.M., "Capital-Labour Substitution and Economic Efficiency", Review of Economics and Statistics, August 1961.

(39) Nerlove, M., "Recent Empirical Studies of the CES and Related Production Functions", in Brown, M., (ed), The Theory and Empirical Analysis of Production, N.B.E.R., Studies in Income and Wealth, vol 31, Columbia University Press, 1967.

(40) Nelson, R.R., "The CES Production Function and Economic Growth Projections", Review of Economics and Statistics, August 1965.

(41) Diwan, R.K., "An Empirical Estimate of the Constant Elasticity of Substitution Production Function", presented at the Econometrica Society, Copenhagen, July 1963, and abstracted in Econometrica, October 1964.

(42) Leser, C.E.V., Econometric Techniques and Problems, op cit., pages 83-4.

(43) Ferguson, C.E., "Substitution, Technical Progress, and Returns to Scale", American Economic Review, Papers and Proceedings, May 1965.

(44) Ferguson, C.E., "Time-Series Production Functions and Technological Progress in American Manufacturing Industry", Journal of Political Economy, April 1965.

(45) McKinnon, R.I., "Wages, Capital Costs, and Employment in Manufacturing: a Model Applied to 1947 - 58 U.S. Data", Econometrica, July 1962.

(46) Nerlove, M., Distributed Lags and Demand Analysis for Agricultural and Other Commodities, Department of Agriculture Handbook, 141, United States, 1958.

(47) Nadiri, M.I., "Some Approaches to the Theory and Measurement of Total Factor Productivity: A Survey". Journal of Economic Literature, December 1970, pages 1150-6.

(48) Walters, A.A., Introduction to Econometrics, op cit., pages 305-8.

CHAPTER EIGHT

METHODS OF PRODUCTIVITY MEASUREMENT : (ii) REFINED STUDIES

The discussion at the end of chapter 7 revealed the objective of more-refined studies of productivity measurement. Christ (1) expressed his dissatisfaction in 1961 with the conventional methodology and proceeded to indicate the new path which had to be pursued.

> "As a profession we have attempted to understand increases in gross output by measuring inputs in the form of labour and capital, and we have divided the real value-added output (or physical unit produced) by this real labour-and capital-input measure. The result is the familiar index of output over input.
>
> At first it seemed enough to compute such an index, to note that it appears to increase at about one per cent per year, and to attribute this growth to increases in technological knowledge. This is no longer sufficient. It is now necessary to try to get independent measurements of things that we believe are components of the index and see whether they account for observed rates of growth of that index. In other words, we should try to force to zero the residual or unexplained part of the increase in output."

This concept of "forcing the residual to zero" has two aspects - measurement and explanation - although the difference between the two may often be a matter of semantics. The "measurement aspect" concentrates on the approach that our prime objective remains to measure the size of technical progress as a catch-all residual. In effect, it still remains a "measure of our ignorance". However, it concentrates on removing many of the deficiencies of the "unrefined studies" of chapter 7 so that the measure of our ignorance is smaller. It will be recalled that the unrefined studies demoted the role of capital and emphasised the role of "unexplained" technological progress in explaining the growth of output. But such studies are so characterised by errors, deficiencies, inconsistencies, incorrect assumptions, wrong specifications, and so on, that their overall conclusion is open to suspicion. When these drawbacks are removed or amended the role of technological progress is found to be drastically reduced. On the other hand, the "explanation aspect" seeks to go one step further and remove even the drastically reduced "measure of our ignorance" to zero. The emphasis is on explaining fully the growth of output by decomposing it into all its component parts. This approach has been called "growth accounting". Obviously, both the measurement and explanation approaches are valid. Which is the most relevant depends upon the purpose of our study. An examination of both of these approaches forms the subject matter of this chapter.

A. THE MEASUREMENT APPROACH

There are many potential "mistakes" characteristic of the unrefined studies which could possibly account for the emphasis laid on the large role of technological progress. In the discussion below six of these have been identified and analysed, namely,

- the specification of inputs and output,
- the specification of the production function and its estimation,
- the assumption of neutral technological progress,
- the assumption of constant returns to scale,
- the assumption of disembodied technological progress,
- the distorting effect of resource reallocation.

Each of these will be examined in turn.

Specification of Inputs and Output

A criticism frequently levelled at the unrefined studies of chapter 7 is that they all but guarantee a major role for the technological-progress residual simply by virtue of the way in which inputs and output are specified. Specifically it is claimed that measures of inputs are seriously under-estimated. Two issues are involved in this whole aspect of whether factor inputs are appropriately measured, (i) the techniques for measuring inputs, and (ii) quality improvement in inputs. The discussion will not be exhaustive at this stage. Only a brief examination of certain points is undertaken as further analysis and elaboration of inputs and output occurs in Part III.

Conventional studies at the economy level relate output to two factors of production - labour and capital - whilst at the sector and industry levels, if output is defined in gross terms rather than as value-added, a third input must be added - raw materials. These input variables are composite measures. They must be regarded as aggregate concepts combining within themselves many sub-categories of inputs. Ideally, each of these input sub-categories should comprise a separate variable in the production function. Their numbers are so large, however, that for purposes of statistical estimation this is not feasible, and, accordingly, the many heterogeneous input sub-categories are combined into one overall input measure, be it called labour, capital, or raw materials. Exactly the same problem is experienced with output. The condensing of this multi-dimensioned structure, of numerous kinds of inputs and outputs, into a single measureable dimension of manageable proportions without any loss of essential information constitutes a major aggregation problem facing us with the familiar weighting and index-number problems.

This aggregation problem has already been introduced in chapter 7. Its importance lies in the fact that without proper aggregation one cannot interpret the properties of an aggregate production function, which governs the behaviour of total factor productivity. Neoclassical aggregation principles lay down that the necessary and sufficient conditions for grouping variables are:

(i) that the marginal rate of substitution between any two variables in a group shall be a function only of the variables in that group, and, therefore, independent of the value of any variables in any other group - often referred to as Leontief's functional-separability theorem;

(ii) that the marginal rate of substitution between any two types of a variable must be constant, i.e. the two types are perfect substitutes - this condition being required to ensure that the aggregate is a simple sum of different elements in the group.

These conditions are merely stated here without any additional discussion or elaboration. The literature on the aggregation problem is considerable and beyond the scope of this thesis. Extensive discussion of the major features is undertaken in Green (2) and Edwards and Orcutt

(3), amongst many others. One point immediately obvious is that technological facts dictate that many different types of machines and workers, for instance, are complementary and not perfect substitutes as required by the neoclassical aggregation principles. The one definite conclusion that can be drawn is that aggregation presents a serious problem affecting the magnitude, stability, and dynamic changes of total factor productivity.

Measurements of inputs and output must, of course, be made in volume terms since for productivity analysis the point of interest is the trend of the ratio of aggregate physical volume of output to physical volume of inputs. It must be remembered that the measurement of productivity is the sole purpose for which such input and output variables are being developed. The very concept of productivity implies that the contribution to output that a factor makes can differ for reasons other than differences in the quantity of that factor. The separation of such influences from quantity changes requires the development of indices of factor inputs in quantity terms, which can be studied in relation to changes in output also expressed in quantity terms. The concept of input measurement must always be considered in this context. This implies that if the efficiency concept of productivity is to mean anything, input and output must be so defined that they are not equal. In other words, no allowance must be made for quality changes. This "no-quality-change approach" is at the other extreme to the "explain-everything approach" which explicitly amends input measures to compensate for gradual quality improvement. This is an attempt to make total inputs add up to total output and in so doing eliminating changes in productivity by definition by constraining output per unit of input to unity. This surely violates the whole concept of what productivity indices are trying to measure.

(i) Capital

If this is the way in which input specification is to be regarded, then the ideal measurement of capital is in terms of directly-measured physical units. However, this is not possible since different kinds of capital are expressed in different physical units. The only alternative is to measure capital indirectly in some sort of comparable unit - namely value. This then poses the question of whether capital should be valued in terms of its cost or in terms of its contribution to production. The latter approach implies that technologically-induced quality improvements in capital are incorporated in the capital measure. This is the approach adopted in most studies of embodied technological progress but which violates, as previously argued, the concept of a productivity index. The former approach, however, reflects quality improvement in the technological-progress term and not in the capital measure. Denison (4) is often quoted as the standard reference for the concept of measuring capital by its cost. He defines gross stocks in the following way,

> "the value, in base-period prices, of the stock of durable capital goods measures the amount it would have cost in the base period to produce the actual stock of capital goods existing in the given year. Similarly, gross additions to the capital stock and capital consumption are valued in terms of base-year costs for the particular types of capital goods added or consumed."

Basic to this definition is that only quality change which is cost-associated is counted as quantity change. A quality improvement which leaves cost unchanged is not counted as a quantity improvement. However, a quality improvement accompanied by a rise or fall in cost is counted as a quantity improvement or deterioration, respectively. Measuring capital by cost implies that if the cost of two types of capital goods was the same (or would have been the same were both newly produced) in the year in whose prices the measures are expressed, they are considered to represent the same amount of capital regardless of differences in their ability to contribute to production. Thus old and new machines having identical deflated production costs are considered to be equal amounts of capital.

There is no obvious "price" of capital goods which can be used to deflate value figures to volume figures. Typically, an index of the cost of inputs making capital equipment is used but this tends to seriously over-deflate the capital measure and understate the increase in the quantity of capital. Kennedy and Thirwall (5) state, "it implies that an item which costs twice as much to produce as another item represents twice as much capital, which ignores scale economies and increases in the efficiency of factors in the capital-goods industries themselves are not reflected in the price index".

Ruggles and Ruggles (6) pursue the point by quoting the contrived case of two pipelines of a given diameter laid together over a desert. The cost would be less than twice that of installing a single pipeline due to economies achieved in putting them in simultaneously. In physical units the two pipes are, of course, twice as much capital as one pipeline; but in cost terms the two pipes are less than twice as much capital as one pipeline, thus under-estimating the actual physical units.

Gross capital stock is invariably reduced to net capital stock by the process of depreciating assets over their working lifespan, in order to take account of technical obsolescence and physical deterioration. The problem is, however, that obsolescent equipment can, and often does, continue contributing to production so that the flow of capital services does not decline with age at the rate suggested by depreciation measures. Additionally, Ruggles and Ruggles suggest that the concept of gradual depreciation of capital assets to take account of physical deterioration is internally inconsistent, in that if increases in efficiency are excluded from capital measures then decreases in efficiency due to ageing should be similarly excluded. Depreciation should not be deducted until the capital asset is retired. In short, measures of capital stock which are net through depreciation, and which exclude quality increases, are internally inconsistent and tend to seriously understate the contribution of capital to growth.

So far the discussion has concentrated on measuring the stock of capital assets, whereas for productivity analysis a flow measure is required. The simplest approach is to assume that the flow is always proportional to the stock by assuming a constant rate of capacity utilisation. This is not always a realistic assumption and more-serious studies attempt to refine the stock measure into a flow measure by compensating for measured variations in capacity utilisation. Various techniques are available. Invariably, studies which adjust the capital stock for changes in utilisation note a considerable increase in the sensitivity of output to capital.

(ii) Labour

The measurement of labour involves fewer conceptual problems than the measurement of capital. Although the aggregation problem is still present, the others are absent. Labour is already measured in physical units, e.g. total employment or manhours, so that the deflation problem from current to constant values does not exist. Depreciation is not deducted, consistent with our approach of not taking quality changes into account. Actually, an inconsistency can be noted here - traditional input measurements depreciate capital and not labour, on the assumption that human assets do not deteriorate with age in the same manner as capital assets. This may be a wholly unwarranted assumption as it could be argued that as workers age, their health deteriorates and motivation declines faster than their efficiency increases due to practice and experience, so that a case may be made out for depreciating labour. But this is not done and an inconsistency arises between the treatment of capital and labour. However, no inconsistency arises when all quality changes are excluded from all inputs, with depreciation only being deducted in full on the retirement of the asset. Finally, the stock-flow problem does not arise in labour measurement as long as labour is measured in manhours since this is already a flow variable. However, total employment is a stock concept which must be transformed to a flow concept via a measure of its utilisation.

(iii) Output

Where output consists of a completely homogeneous commodity, a measure of output in physical terms is merely a count of the number of units produced. Correspondingly, the aggregation problem is avoided. However, certain strict conditions have to be met for a commodity to be homogeneous. Invariably, plants produce heterogeneous products, and this problem of multi-dimensionality increases as one moves to higher aggregation levels - the firm, industry, sector, and economy. Measurement in terms of physical units now becomes impossible, and an overall output measure is only achieved through combination of all different goods and services on a common basis. Thus, the aggregation problem is present. Aggregation is achieved in value terms posing the problem of price deflation from current to constant terms. These problems are, therefore, similar to those experienced in capital measurement. However, depreciation problems are not encountered and stock-flow adjustments are similarly absent.

(iv) Raw Materials

"Raw materials" is the aggregate name given to an extremely heterogeneous collection of inputs encompassing anything which is completely used up in the production process. If output is defined in value-added terms then an adjustment for raw materials has implicitly been made, and hence, there is no need for a separate raw materials input variable in the production function. However, if output is defined gross then raw materials must be specifically included, but because of their extreme heterogeneity, measurement in standardised physical units is not possible. Thus, problems involving aggregation and price deflation are again evident. However, because raw materials are used up instantaneously and completely in the production process, depreciation and stock-flow problems are avoided.

(v) Removing Aggregation and Measurement Errors

Four problems have been identified above concerning the specification of inputs and output, namely: aggregation, price deflators, depreciation, and factor utilisation. Inappropriate specification procedures can, therefore, have a serious impact on the measurement of inputs and output and this will be reflected in the size of the total-factor-productivity residual. As far as is practically possible, therefore, these deficiencies should be avoided.

The potential magnitude of the problem is highlighted by Jorgenson and Griliches (7). They showed that the whole of the total-factor-productivity residual can be explained away in terms of errors made in aggregation and measurement in prices and quantities of the inputs and output. The authors identify, and remove, four main sources of error;

- aggregation errors in combining investment and consumption goods and labour and capital services,
- errors resulting from the aggregation of investment goods and capital services on the one hand and labour services on the other,
- measurement errors in the prices of investment goods resulting from the use of input prices into the investment goods sector rather than the use of output prices from this sector,

- errors arising from a failure to measure varying input utilisation.

Their exact methodology is too detailed to be examined at this stage, but further discussion is made later in this chapter and also in chapter 9. They showed that the removal of these errors from data on inputs and output for the U.S.A. private domestic economy over the period 1945 - 65 results in the virtual elimination of the productivity residual. The average rate of growth of total factor productivity of 1,6 per cent per annum before correction is reduced to 0,1 per cent per annum after correction. Put another way, the growth of total productivity explains only 3,3 per cent of the growth of output compared with 47,6 per cent before correction. This startling demotion of the role of technological progress is effectively criticised by Denison (8) who shows that Jorgenson's and Griliches' results are almost entirely due to an unwarranted adjustment in the capital-utilisation series, and that so-called "errors" removal does not have the dramatic impact the authors claim. This is partially admitted in an amending paper by Christensen and Jorgenson (9) where another approach to capital utilisation is adopted. They find a growth rate of total factor productivity over the period 1948 - 67 of 0,31 per cent per annum. This is a larger magnitude than discovered in the earlier paper but is still small enough to challenge Denison's assertion that "error removal" has very little impact.

(vi) Avoiding the Capital Problem

The conclusion of the "unrefined" studies that the major component of labour-productivity increases is due to technological progress with a minor role being attributed to increases in capital per head is challenged by a school of thought which believes such results are caused through deficiencies in the capital input. Jorgenson (10) argued that any index of total productivity growth can always be interpreted in terms of measurement errors in the capital series. Accordingly, it becomes impossible to distinguish what is an increase in productivity from what is a measurement error. Rather than trying to correct results by "removing" such errors in the Jorgenson-Griliches manner, the problem can be "avoided" altogether, by employing either of two approaches. Firstly, techniques can be devised of measuring productivity growth without the need for a capital-input series. Secondly, a proxy variable can be used for capital, measured in standardised physical units, thus avoiding the type of problems already discussed.

An example of the first type of approach is provided by Johansen. He contended (11) that results are often biased because capital data are generally of a lower reliability than data for labour and output and such a bias tends to understate the effect of capital on output and overstate the importance of factors represented by more-reliable series. It is always a problem in empirical studies of obtaining appropriate statistical information about capital accumulation. Johansen contends that usually the figures used do not correspond to the definition of capital most relevant for production analysis, and also the figures are unreliable as measurements of what they should measure according to their definitions. In view of such worries over the measurement of capital, Johansen derived a method (12) of separating the effects of capital accumulation and technological progress on the growth of labour productivity without using capital data. Instead he used only data on labour productivity and factor shares. His method required cross-section data for a set of several industries, and, hence, could not be applied to single industries. Writing the Cobb-Douglas production function for industry i at time t in the usual way as,

$$X_{it} = A_{it} L_{it}^{\gamma_i} K_{it}^{\sigma_i}$$

Johansen assumed γ_i and σ_i to be constant with $\gamma_i + \sigma_i = 1$, i.e. constant returns to scale. Shifts in the production function representing technological progress are captured by the constant term A, and, hence, neutral technological progress is implied. Of course, the use of Cobb-Douglas also implies a unitary elasticity of substitution.

Denoting labour productivity (output per unit of labour) by a_{it}, it can be shown that labour productivity in time period t=2 compared with t=1 (which need not be consecutive), can be written as,

$$\frac{a_{i2}}{a_{i1}} = \frac{A_{i2}}{A_{i1}} \left(\frac{K_{i2}/L_{i2}}{K_{i1}/L_{i1}}\right)^{\sigma_i}$$

which shows how the increase in labour productivity is compounded of shifts in the production function and the increase in capital per worker. Denoting the latter term by ω_i we have

$$\frac{a_{i}2}{a_{i}1} = \frac{A_{i}2}{A_{i}1} \cdot \omega^{\sigma_i} \quad .$$

Transforming to logarithms and denoting $\varepsilon_i = \log (A_{i}2/A_{i}1)$ gives

$$\log (a_{i}2/a_{i}1) = (\log \omega)\ \sigma_i + \varepsilon_i$$

which is used as the estimating equation. Accordingly, capital data is not required. What is needed is cross-section data on labour productivity and observed capital share for individual industries in order to enable an ordinary-least-squares regression of $\log (a_{i}2/a_{i}1)$ on σ_i to be run.

Johansen covered 28 industries and compared 1950 with 1924. He obtained the result,

$$\log (a_{i}2/a_{i}1) = 0{,}266\ \sigma_i + 0{,}080$$

the constant term of 0,080 implying an average production-function shift of approximately 20 per cent between 1924-1950, or approximately 0,7 per cent per annum. For an industry with $\sigma = 0{,}5$ it is possible to divide the labour-productivity increase between growth in capital per worker and production-function shifts. The former factor accounts for a growth in labour productivity of 36 per cent (antilog 0,133-1) and the latter for 20 per cent (antilog 0,080-1). Total growth would be 63 per cent (antilog 0,213-1) thus leaving 7 per cent for the interaction between the factors.

Johansen, therefore, shows a larger contribution for capital-per-worker increases than is indicated by traditional studies. However, the results obtained above are, of course, the average across industries regardless of the actual increases in output experienced. Johansen, therefore, subsequently divided the industries into 3 groups according to the size of increases in output experienced, and computed separate regressions for each group. The results were significant. The percentage of growth accounted for by production-function shifts was highest in those industries experiencing the largest output increases. However, increases in capital per worker usurped the major role in industries with the smallest output increases. This indicates how dependent the relative size of production-function shifts is on the output growth of the industry. Johansen's conclusion is that technological progress appears small relative to increases in capital per worker as an explanation of labour-productivity growth in slow-growing industries and the opposite in fast-growing industries.

Johansen's contention, therefore, that the contribution of capital per head rises when problems of capital measurement are avoided, is partially validated. But in devising such an ingenious method he has had to pay the cost of introducing some of the more-restrictive theoretical assumptions characteristic of the "unrefined" studies - neutral, disembodied technological change with constant returns to scale. Technological progress is still regarded as an amalgamation of many "unexplained" elements. Decomposition is not attempted.

An example of the second type of approach is provided by Maddala (13) who used the horsepower rating of power equipment as a proxy for capital input in his study of the bituminous coal industry in the U.S.A. over the period 1919-54. Like Johansen he also placed a question mark over the dominance of pure technological progress. Maddala employed a Cobb-Douglas production-function technique, but tried to encompass as much as possible within the function leaving as little as possible to be explained by shifts in the function. He used cross-section data at several points in time, thus avoiding the problem encountered in time-series studies, and tested for the stability of the function co-efficients. Shifts in the function were estimated by shift parameters rather than by time trends. The study made use of cross-section data on twenty states obtained from the Censuses of 1919, 1929, 1939, and 1954. His method of estimation was single-equation least squares, experimenting with various combinations of variables for significance.

The availability of several cross-sections enabled him to test for significant production-function changes. Such an estimation method does not require these changes to be smooth and uniform over time as in the case of fitting production functions with exponential trends to time-series data. The production-function shifts can be measured both by the shift parameters introduced into the function and by the conventional productivity indices. Such indices can be constructed in either of two ways: firstly, by using as weights the estimated parameters of the Cobb-Douglas function, or, secondly, by using as weights the observed factor shares. The former can allow for economies of scale and factor-market disequilibrium, whilst the latter cannot. The problem of which of these methods is the better is hardly settled. A general answer cannot be given since it depends on the whole mechanism generating the data. However, Maddala decided against using the weighting scheme based on factor shares and instead made use of the production-function co-efficients obtained as weights. The calculated co-efficients for labour and capital were 0,6 and 0,4 respectively - the labour co-efficient being lower than that obtained from data on factor shares. Over the period 1919-54 this resulted in an input index of 83,3 which when compared with the output index of 90,2 gave a productivity index of 108,3. This is far below the Kendrick index, using factor shares as weights, of 170,9 for the same period and industry.

This very small component of technological progress contradicts previous studies. It indicates that the increase in labour productivity over the period was almost totally explained by the increase in the horsepower rating of equipment per worker, the substitution between the two factors having occurred in response to changes in relative factor prices. Maddala believed conventional methods of productivity had not revealed this process of factor substitution adequately. Because of deficiencies in the measurement of inputs there was invariably a large unexplained residual going under the label of technological progress. But by changing the definition of capital, what was formerly considered technological progress (a shift in the function) is now revealed to be a process of factor substitution (a movement along the function). Of course, Maddala's findings depend crucially on the appropriateness of his capital proxy. He defends his measure on the grounds that there is no unique measure of capital suitable for all purposes. Most studies had used the conventional measure of capital standard in accounting and financial circles, but horsepower rating being based on physical productivity could be regarded as more appropriate for productivity analysis. Precedents in the use of such a capital proxy had already been set by Rostas (14), Melman (15), and Leser (16), so Maddala was hardly employing a new innovation.

Specification of the Production Function and Its Estimation

In the unrefined studies of chapter 7 it was implicitly assumed that there exists such a concept as an aggregate production function. Various aspects of this assumption were presented and analysed at the time, with an examination of the conditions under which it is possible to define an aggregate production function. Although economic production theory is strictly applicable only to micro units, economists have been unable to resist the temptation to apply similar concepts to firms as an aggregate. By adding together the inputs and outputs of each firm is it possible to interpret the resulting aggregate function for the group of firms in the same way as the production function of the firm? This aspect was analysed in chapter 7. Fisher (17) examined the problem of aggregating a number of technically-different microeconomic production functions. He showed that even with homogeneous capital goods and a neoclassical production function, labour, capital, and output aggregation over all production units required stringent conditions. With constant returns to scale and only two factors of production, the necessary aggregation condition is that all capital is perfectly substitutable and all technical changes are capital-augmenting. With non-constant returns to scale, capital aggregation is possible only under the restrictive assumption that the individual firm's production function can be made to yield constant returns after "stretching of the capital axis". Similar conditions have to be specified for labour aggregation.

Certainly caution is required in interpreting the results that depend on the existence and specification of an aggregate production function. Nadiri (18) reminds us that the aggregate production function "does not have a conceptual reality of its own" rather "it emerges as a consequence of the growth processes at various micro-economic levels and is not a causal determinant of the growth path of an economy". He considers that although reasonably-good estimates of factor productivity are obtained from use of an aggregate production function, this is due mainly to the narrow range of movement of aggregate data, rather than the solid foundation of the function.

Aside from the aggregation problem is the additional difficulty of specifying and estimating the form of the aggregate function to be employed. It was mentioned in chapter 6 that since the exact relationship between output and inputs is not known it is necessary to employ approximating functions - the two most popular having been Cobb-Douglas and CES. Our productivity indices are deduced either from explicit or implicitly-defined production functions, so that the accurate specification of the form and estimation of the parameters of the function are crucial to the measurement of these indices. Any errors or misspecifications will spill over to the measure of total factor productivity. At stake is the magnitude of the residual and its stability over time.

Chapter 7 analysed some of the more-important problems inherent in the specification and estimation of the Cobb-Douglas and CES functions. In an attempt to by-pass such problems two important outcomes have resulted from recent research work, namely, the development of more-generalised production functions, and, secondly, the development of indirect estimation techniques.

The formulation of new production functions has taken us increasingly into the realm of statistical ingenuity. Nadiri (19) contends that developments have taken place on three main fronts;

(i) amendment of the standard two-factor CES function,

(ii) indirect estimation of the parameters of multi-factor production functions by formulating first the relevant cost functions,

(iii) specification of inter-temporal production models which account explicitly for the costs of adjusting the level of inputs.

The main example of (i) is the development of the Variable Elasticity of Substitution production function (VES) which includes Cobb-Douglas and CES as special cases. The work of Lu and Fletcher (20) was important in this respect. Developments under (ii) are due to several authors (for instance, Diewert (21)) but invariably suffer from problems of econometric estimation due to the large number of parameters and the poor quality and collinear nature of the aggregate data. Accordingly, Hanoch (22) developed the Constant Difference Elasticities of Substitution function (CDE) to tackle this problem. Developments under (iii) are essentially an extension of the work of Eisner and Strotz (23). Important empirical studies are due to Schramm (24) and Nadiri and Rosen (25).

The development of more-appropriate and powerful estimation techniques has also increasingly taken us down the same path of statistical ingenuity. Remaining within the context of single-equation procedures, two developments can be noted, namely, the application of least squares to a linear approximation of a function, or, the application of non-linear least squares. In regard to the former, Kmenta (26) applied least squares to a linear approximation of the CES function. In regard to the latter, Bodkin and Klein (27) proposed a non-linear maximum-likelihood procedure. Assuming certain initial values for the parameters, the likelihood function is solved iteratively until the lowest sum of squared errors is obtained and the parameter estimates converge on a particular set of values. Also developed and employed has been a Bayesian estimation technique using a similar likelihood function but ignoring information about maximisation behaviour and market conditions. It directly estimates the parameters of an average production function and not the efficient combination of inputs suggested by economic theory. This approach has been formulated for the Cobb-Douglas function by Zellner, Kmenta, and Dreze (28), and for the CES function by Chetty (29).

Non-Neutral Technological Change and Scale Economies

The unrefined studies of chapter 7 all specified neutral technological change and (with a few exceptions) constrained the function to constant returns to scale. Any decomposition of the productivity residual would, therefore, have to tackle the problem of distinguishing between neutral and non-neutral technological progress and economies or diseconomies of scale. However, such attempts meet with serious identification problems. When testing for bias in technological progress it must be remembered that the tendency for capital to accumulate relative to labour is accompanied by a relative depression in the capital price compared with the labour price due to technical progress in the capital-goods sector. Increased capital usage could, therefore, have been caused by changes in relative prices initiating factor substitution, rather than a labour-saving bias in technological change. Similar difficulties arise in distinguishing between types of scale economies, that is, between those which are technologically determined and those induced from increases in physical inputs. Such a distinction is not achieved by simple production-function estimation by ordinary least squares.

(i) Non-Neutral Technological Change

In attempting to distinguish between neutral and non-neutral technological progress a problem arises in that it is possible to evaluate the latter if the former is not present at the same time. If, however, one attempts to estimate a production function with both kinds of progress being present, then identification problems are encountered and estimation is possible only under additional assumptions. The possibility of identifying both the production function and arbitrary forms of technical change is questioned in terms of the Diamond-McFadden "impossibility theorem" (see Nerlove (30)). Identification may be produced by making certain "smoothness" assumptions about the nature of technical change, namely, smooth exponential growth in the effectiveness of measured capital and labour inputs.

Bearing this in mind the discussion can now proceed to an examination of some of the more-important studies and techniques. Salter (31) developed a measure which attempted to break down the residual into neutral and non-neutral technological progress. His definitions of these terms have already been examined in chapter 6. Salter examined the influences leading to changes in "best-practice techniques" and labour productivity by analysing prices, costs, and output for various British and American industries over the period 1923 - 50. He identified three main influences on changes in best-practice techniques: neutral technological change, biased technological change, and factor substitution.

Neutral change (represented by Tr where r represents proportionate rates of change) bears equally on both factors of production and can be denoted by,

$$Tr = \frac{\omega\,(dL/dt) + q\,(dK/dt)}{\omega L + qK}$$

where ω and q are the prices of labour and capital respectively and t denotes time, and measures the extent to which unit production costs change while factor prices remain constant.

Biased change (represented by Dr) can be denoted by,

$$Dr = \frac{d\ (\tilde{K}/L)}{dt} \cdot \frac{L}{K}$$

where a positive Dr represents a labour-saving shift, and a negative Dr represents a capital-saving shift.

Substitution of one factor for another is generated by changes in the rate of growth of relative factor prices,

$$\frac{dq/\omega}{q/\omega} \quad .$$

Letting Lr and Kr be the proportionate rates of change of unit labour and capital requirements respectively, he adds up the separate effects and derives,

$$Lr = Tr - \pi Dr + \varepsilon\pi(q/\omega)r$$
$$Kr = Tr + (1-\pi)Dr + \varepsilon(1-\pi)\ (q/\omega)r$$

where ε is the elasticity of substitution and π the ratio of capital costs to total costs. The measures on the right-hand sides of the two above expressions represent independent phenomenon. Each should be held constant whilst measuring the others but this is an impossible goal since all the terms are inextricably linked. This implies that Salter's concept of focusing on the proportionate change in each factor resulting from neutral and biased technological progress and factor substitution and developing a measure of each component, breaks down at the point of trying to isolate the independent contribution of each component. In other words, it is one thing to formulate a set of definitions of forces affecting labour productivity, but another to translate this into measures of these forces. Accordingly, his stated measures can only be of limited applicability. His empirical findings concluded that differences in the rate of growth of labour productivity between industries were primarily attributable to technical progress and economies of scale, but with factor substitution making an important contribution. He reached the broad conclusion that neutrality of technical change was a reasonable hypothesis with unequal rates of neutral advance between industries being observed.

Ferguson (32) employed a side relation of the CES production function in order to estimate non-neutral technological progress. It can easily be shown that in such a function the expression for the marginal rate of technical substitution reduces to

$$\frac{\omega}{r} = \frac{1-\delta}{\delta}\left(\frac{K}{L}\right)^{1+\rho}$$

where the symbols are the same as those defined in chapters 6 and 7, and ω is the real wage rate and r the rate of return on capital. It is known that a rising value of δ represents capital-using progress, and, hence, the above expres ion is better written,

$$\log\left(\frac{\delta}{1-\delta}\right) = \log\left(\frac{r}{\omega}\right) + (1+\rho)\log\left(\frac{K}{L}\right).$$

To determine the nature of biased progress, Ferguson uses this expression to compute δ annually in a series of manufacturing industries over the period 1949 - 61. These δ values can then be plotted for each industry, and technological progress is judged to be capital-using (capital-saving) if the number of positive (negative) changes in δ predominates. Of the nineteen industries, eight displayed no predominance of either positive or negative changes, eight displayed a predominance of positive changes (capital-using), and three displayed a predominance of negative changes (capital-saving). Thus, on balance, technological progress appeared to be either neutral or capital- using in individual industries. However, the three capital-saving industries (chemicals, primary metals, and electrical machinery) tend to be large and when the manufacturing sector is aggregated they may offset the capital-using changes in the smaller industries.

Accordingly, it is not surprising that most time-series studies of the aggregate manufacturing sector show neutral progress to be a reasonable assumption.

Ferguson's approach, unfortunately, runs foul of the "impossibility theorem", as pointed out by Nerlove (33). He considers Ferguson's results to be "spurious" in the sense that "what he purports to have found can be shown (to be) meaningless". His approach of computing a non-smooth estimate of the bias in technological change is clearly nonsense on the basis of the "impossibility theorem" and makes his results arbitrary.

David and van de Klundert (34) also employ a CES production function in order to distinguish between neutral and biased technological progress over the period 1899 - 1960 for the U.S.A. private domestic economy. Unlike Ferguson, however, they are able to by-pass the "impossibility theorem" by assuming that all technical change is factor-augmenting and exponential. These assumptions are sufficient to identify the production function and technical change. They employ the constant returns to scale formulation,

$$X = [(E_L L)^{-\rho} + (E_K K)^{-\rho}]^{-\frac{1}{\rho}}$$

where the co-efficients E_L and E_K represent the levels of efficiency of the conventional inputs of labour and capital. Alterations in E_L and E_K through time are to be interpreted as labour-augmenting and capital-augmenting technical changes which can be related to the Hicksian concepts of neutral, labour-saving and capital-saving technological progress. A labour-augmenting bias is the same thing as a labour-saving innovation if, and only if, the elasticity of substitution is less than unity. Similarly, a capital-augmenting bias is the same thing as a capital-saving innovation on the same condition.

By assuming perfect competition, in place of the elasticity of output with respect to labour and capital respectively can be written the share of total output received by labour (γ) and capital (σ) respectively. Additionally, if it is assumed that any changes in the relative efficiency level of labour over time occur at a constant geometric rate ($\lambda_L - \lambda_K$) given by,

$$\left(\frac{E_L}{E_K}\right)_t = \frac{E_{L(o)}}{E_{K(o)}} e^{(\lambda_L - \lambda_K)t}$$

then the following expression can be derived from the constant-returns-to-scale formulation,

$$\log\left(\frac{K}{L}\right) = \left(\frac{\varepsilon}{1-\varepsilon}\right) \log\left(\frac{\gamma}{\sigma}\right) + (\lambda_L-\lambda_K)t + \log\left(\frac{E_{L(o)}}{E_{K(o)}}\right) .$$

The authors subsequently derive a regression equation from this expression permitting estimation of a constant rate of change in relative conventional input efficiency. In this derivation they introduce a number of modifications designed to cope with problems posed by cyclical variations in the utilisation rate of capital stock and lags in the response of the capital-labour ratio to changes in relative factor prices. The model is fitted to data allowing exponential bias in efficiency growth to be captured. Their equation is a complicated one and need not be reproduced here, but it does allow estimation of ε and $(\lambda_L - \lambda_K)$ which are the major variables of interest.

The elasticity of substitution is found to be 0,32, described by the authors as "arrestingly small" and far smaller than equivalent studies had found. The rate of bias in the growth of conventional-input efficiencies $(\lambda_L-\lambda_K)$ is found to be 0,0072 leading to the conclusion that over the period 1899 - 1960 technological progress had not been neutral but had instead increased conventional-labour-input efficiency more rapidly than the efficiency of conventional capital inputs. With $\varepsilon<1$, technological progress had been labour-saving. The magnitude of the differentially faster rate of labour-augmentation had been 0,72 per cent per annum. Over the sixty-year period the efficiency of labour increased by roughly 54 per cent more than the efficiency of capital.

Such an estimate of the long-run disparity between the rates of growth of labour and capital efficiency says nothing about the magnitudes of the actual rates of labour capital augmentation, or the importance of their respective contributions to the rate of growth of conventional total factor productivity. The authors find these values to be sensitive to the definition of labour's share employed. When defined as "employee compensation, exclusive of entrepreneurial income, as a proportion of gross private business product (=average 0,476)", the efficiency of labour had grown at an annual rate of 2,23 per cent; that of capital 1,51 per cent; the growth rate of weighted total-factor efficiency had been 1,85 per cent per annum; and 57 per cent of this had been accounted for by labour-augmenting technical changes. However, when labour's share is defined in relation to national income including an estimate of the wage component of entrepreneurial incomes (=average 0,751), the same values are calculated at 2,30 per cent per annum; 1,58 per cent per annum; 2,13 per cent per annum; and 81 per cent.

Van der Dussen (35) follows the lead of Ferguson in using the marginal-rate-of-technical-substitution side relation of the CES production function, namely,

$$\frac{\omega}{r} = \frac{1-\delta}{\delta}\left(\frac{K}{L}\right)^{1+\rho}$$

in order to estimate the rate of biased technological progress for ten South African industries over the period 1945 - 63. However, unlike Ferguson, he is able to by-pass the "impossibility theorem" through the nature of his smoothness assumptions.

The side relation can be re-arranged to render,

$$\frac{\omega L}{rK} = \frac{(1-\delta)}{\delta}\left(\frac{K}{L}\right)^{\rho}$$

which for time-series purposes can be written

$$\left(\frac{\omega L}{rK}\right)_t = \exp\,(\eta t).(1-\delta)/\delta.(K/L)_t^{\rho}$$

where η is a time-trend constant. Let the distributive-shares ratio above be D_t, and assume that both labour and capital improve technologically by different exponential rates M_L and M_K. Substituting into the above expression and re-arranging, we obtain,

$$D_t = \exp(\rho(M_L-M_K))t \;.(1-\delta)/\delta.(K/L)^{\rho}$$

so that η can be written as,

$$\eta = \rho\ (M_L-M_K) \quad .$$

Technological progress is neutral only if $M_L - M_K = 0$, labour-saving if $M_L > M_K$, and capital-saving if $M_K > M_L$.

Estimation is carried out by means of least-squares regression on a logarithmic first-difference equation, namely,

$$\log D_t - \log D_{t-1} = \eta + \rho(\log B_t - \log B_{t-1})$$

where $B_t = K_t/L_t$. (Since $1-\delta/\delta$ is a constant it disappears from the logarithmic first-difference specification).

This equation was applied to data on the ten industries in order to derive values in each case for η and (M_L-M_K). With only one insignificant exception (Food) the value of (M_L-M_K) was found to be positive indicating that labour-saving technological progress had taken place over the study period in the remaining nine industries.

So far the discussion has tended to concentrate on studies employing the CES production function to measure biased technological progress. However, it was seen in chapter 7, that the Cobb-Douglas function is equally capable of the same objective. In a series of articles, Murray Brown with two co-authors, (De Cani and Popkin) employed both CES and Cobb-Douglas functions to measure non-neutrality. Since his technique is of special interest, discussion of it is delayed until after the following section.

(ii) Scale Economies

A feature of the "unrefined" studies examined in chapter 7 was the constraining, either implicitly or explicitly, of the model to constant returns to scale. This means that the presence of increasing returns will be reflected in the size of the productivity residual, thus imparting an upward bias. It is, therefore, necessary to measure independently the magnitude of any scale economies and purge them from the residual as a vital step in forcing the size of the residual to zero. This element of the "decomposition" procedure was attempted by Walters (36). He noted that Solow in his pioneering 1957 article had estimated neutral technological progress at between 1,5 and 1,8 per cent per annum over the period 1909 - 1949, but since he had constrained his model to constant returns to scale his results confounded such progress with any scale economies present. To test for such economies, Walters fitted the Cobb-Douglas function with an exponential trend term

$$X = AK^{\alpha}L^{\gamma}e^{\delta t}$$

to Solow's data for the same period. Depending upon the definition of capital used, the sum of the co-efficients σ and γ was always significantly greater than unity, varying between 1,27 and 1,38. The effect of such economies was to reduce the rate of neutral technological progress (as reflected by the value of δ) from between 1,5 and 1,8 as estimated by Solow to between 1,0 and 1,25 per cent per annum. Walters estimated that between 27 and 35 per cent of the increase in output was due simply to scale economies.

Walters' approach is invaluable in drawing attention to the potential importance of scale economies, but, unfortunately, simply estimating the production function by ordinary least squares to obtain a measure of scale economies does not distinguish between those economies that are technologically determined and those which result from a sheer increase in physical inputs. The literature is rich with studies attempting to estimate the magnitude of scale economies and diseconomies for various countries and at different levels of aggregation, and over varying time periods. Results are extremely diverse but no attempt is made to reference or discuss these studies since the majority of them make no attempt to distinguish between scale economies which are technologically induced and those which are due to increases in factor inputs. In effect, a Walters'-type approach overcompensates for the impact of scale economies. It takes too much away from technological progress and ascribes it to scale economies, thus biasing downwards the rate of progress. To the extent that economies of scale are partly technologically determined a portion of such observed economies must be attributed back to technical progress.

The desired objective is that the Cobb-Douglas parameters σ and γ should capture only those economies which are induced by sheer increases in physical inputs (representing a movement along the production function), leaving those economies which are technologically determined to be captured within the technical-progress residual (representing a shift of the production function).

This corresponds to a mid-way position between the extremes of a function constrained to constant returns on the one hand, and the Walters'-type approach on the other.

Barzel (37) also demonstrates how the productivity residual is biased upwards if no allowance is made for economies of scale. He performs his analysis within the context of the Kendrick-type output-per-unit-of-input technique for the electric power industry in the U.S.A. over the period 1929 - 55. The productivity index is measured using the derived formula,

$$A_{12} = \frac{X_2 \Sigma I_1 P_1}{X_1 \Sigma I_2 P_1}$$

where A_{12} is the productivity index defined as productivity in year 2 relative to productivity in year 1. X, I, and P refer to output, inputs, and input prices respectively and the subscripts 1 and 2 refer to years 1 and 2 respectively. Barzel found that total factor productivity in 1955 was 2,83 times that in 1929, the geometric-mean annual increase being 4,1 per cent. However, this figure is likely to be exaggerated because the model is constrained to constant returns to scale and any economies which may be present due to a scale effect are incorrectly included in the index. Although the measure is valid from the viewpoint of output per unit of input, it cannot be considered solely as an estimate of shifts in the production function.

Barzel identifies three types of economies of scale which must be purged from the productivity index: (i) those due to the scale of the individual customer, i.e. supplying large quantities of electricity to the individual customer, (ii) those due to the scale of plant, (iii) those associated with the operation of plants at a higher load factor. The 1940 rate structure was used for estimating the effect of the first type of economies of scale. The other two were estimated on the basis of a cross-section of fifty plants constructed between 1953 and 1955. He found the combined influence of these economies to be substantial. Accordingly, he was able to decompose the total-factor-productivity index of 2,83 between 1929 and 1955 into a product of 1,48 due to economy (i), 1,23 due to economy (ii), and 1,18 due to economy (iii). This reduced the index to 1,32 (i.e. 2,83/1,48 ÷ 1,23 ÷ 1,18) which could now be regarded as a measure of pure technological progress, i.e. due to shifts of the production function only.

However, Barzel cannot escape the same criticism aimed at Walters, namely that some of the economies he has purged from the productivity residual must have been technologically determined, in addition to those which are scale inspired, and, rightfully, must be ascribed back to technological progress. Although the figure of 2,83 is acknowledged as an over-estimate of such progress, the final figure of 1,32 is an under-estimate due to overcompensation for scale economies.

The task of achieving a proper dichotomy between scale-induced economies and technologically-determined economies becomes progressively more complex at higher levels of data aggregation. Detailed, specially-collected, data at the micro level is necessary to achieve such an objective. The abundance of published data at plant level for the steam power industry in the United States has led to a plethora of studies concerning the production operations of this industry. Discussion here will concentrate on those which have attempted the task of separating and quantifying the effects of pure technological progress and scale-induced economies. Important studies include Barzel (38), Komiya (39), Dhrymes and Kurz (40), and Galatin (41).