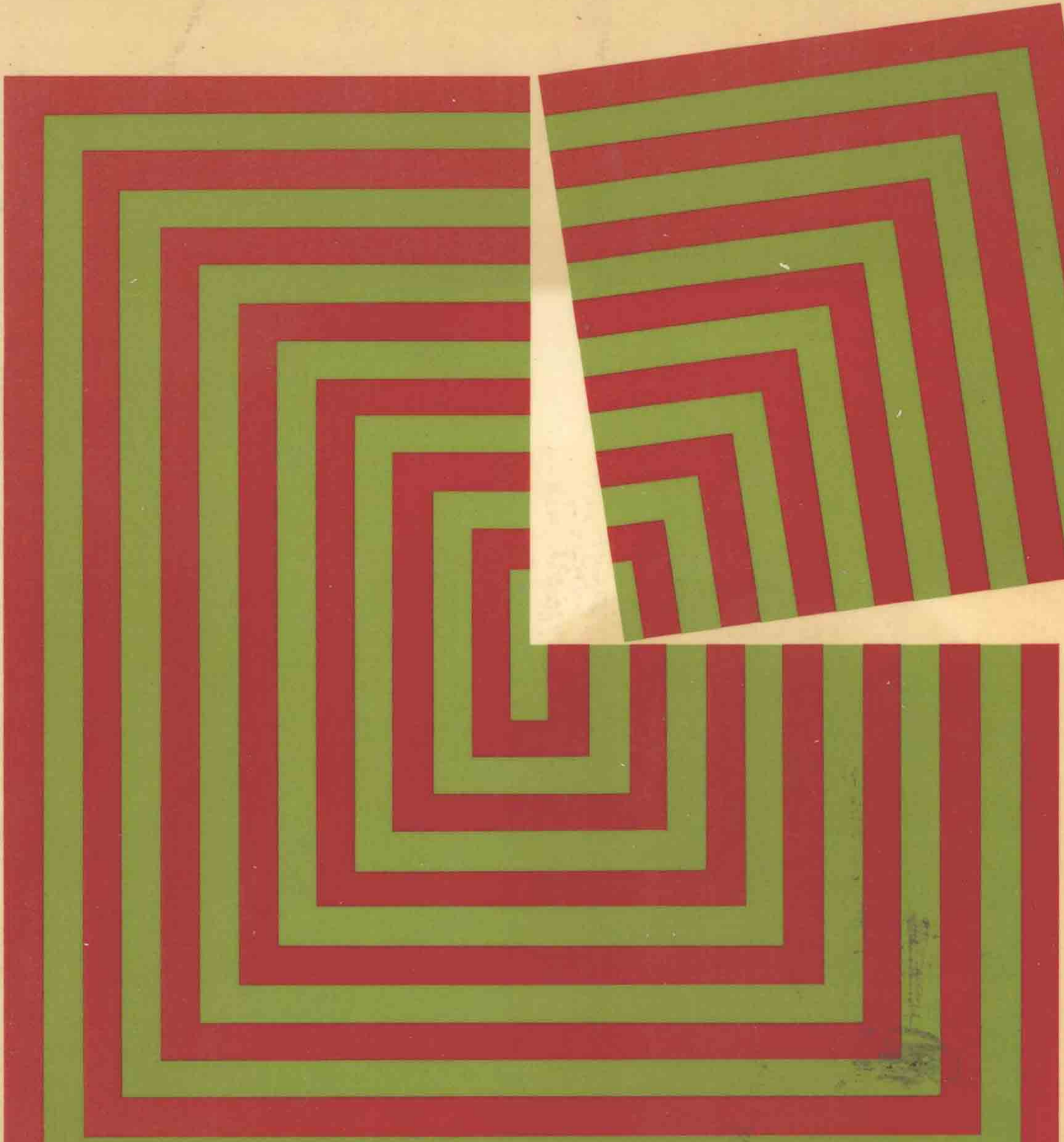
THE ECONOMIC ANALYSIS OF TECHNOLOGICAL CHANGE

PAUL STONEMAN



The Economic Analysis of Technological Change

PAUL STONEMAN

OXFORD UNIVERSITY PRESS
1983

Oxford University Press, Walton Street, Oxford OX2 6DP

London Glasgow New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Beirut Berlin Ibadan Mexico City Nicosia

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data
Stoneman, Paul
The economic analysis of technological change.
1. Technological innovation – Economic aspects
I. Title
338'.06 HC79.T4
ISBN 0-19-877194-0
ISBN 0-19-877193-2 Pbk

Library of Congress Cataloging in Publication Data
Stoneman, Paul.
The economic analysis of technological change.
Bibliography: p.
Includes indexes.
1. Technological innovations – Economic aspects.
2. Diffusion of innovations. I. Title.
HD45.S84 1983 338'.06 83-8311
ISBN 0-19-877194-0 (U.S.)
ISBN 0-19-877193-2 (U.S. : pbk.)

Set by Hope Services, Abingdon
Printed in Great Britain
at the University Press, Oxford
by Eric Buckley
Printer to the University

Preface

The economics of technological change has, for a topic in a young science, a long history. Despite this, recent advances in the subject are not readily available in one source for the interested student or researcher. One aim of this book is to make them available.

My biases should be recognized. As an economist I consider technological change from the economic point of view. The perspectives provided by sociologists, technologists, industrial relations researchers and others are not considered in this volume. This is not to say that such literature should be ignored, but rather that I claim no expertise in expositing the material. Even as an economist, my bias is towards the analytical aspects of the subject rather than its empirical manifestations. For this reason the book reflects a view that empirics must be preceded by theory.

The material in the book is related to material used in lectures given at Warwick University to undergraduate and postgraduate students of industrial economics and macroeconomics. It also reflects material used in a course, taught jointly with Maxine Berg, on the Technology and Industrial Development of the UK Economy since 1800. The degree of mathematical sophistication reflects the form in which this material has been presented to the students on these courses. The mathematical content of the volume is reasonably high, but in general the mathematics is simple. For the most part, it should not really be beyond any undergraduate who has followed a standard course in mathematical analysis. In addition to assuming such a knowledge, much of the book assumes a basic knowledge of macro- and microeconomic theory. This I hope suggests that much of the material will be comprehensible to many second- and third-year UK undergraduates. Other, more sophisticated, material assumes a background appropriate to a graduate student. However, the book is not written exclusively as a text for students; it is also aimed at the researcher in the field, and thus only on occasion is simplification introduced solely to ease exposition.

An attempt has been made to make the book reasonably comprehensive as to the current state of the art. This does not imply however that it even pretends to be a survey of all the literature in the area. The selection is governed by objectives, space and limitations on my own knowledge of the literature. I have attempted to present what I see as the main analytical material, but I have not made a conscious effort to be exhaustive in the empirical area. The book should thus be considered complementary to more descriptive views of the process of technological change. It is also orientated more towards the spread of new technology and the impact of new technology than to the sources of new technology. This reflects my own conception of the relative importance of the different topics and the extent to which material on spread and impact are available in the existing literature. We have not ignored the sources; one cannot do that. We have, however, devoted more space to spread and impact.

The biases I have as an author will mean that others, also economists, will find that certain topics are not covered in this book. Thus, for example, the social consequences of technological change, defined as effects on social structures, the environment, conditions of work, etc., have not really been treated. The economics of education have also been largely ignored. Again, the excuses are a lack of expertise and a shortage of space. It is however hoped that the material that is presented will provide a useful core for any study of technological change considered more widely than we have chosen to consider it here.

In writing this book I have been helped and encouraged by a number of people, not all of whom I can mention here. I owe particular thanks however to my colleagues at Warwick, especially Keith Cowling, who acted as a most useful guinea-pig on whom I could test an earlier version. I am also grateful to Paul Geroski and Chris Freeman for their comments on an earlier draft. I believe that by reacting to their comments the book has been much improved. I must also thank the various typists who have struggled with my handwriting, but especially Liz Cross, who has borne the largest part of this burden. In the usual way, however, any errors or omissions that still remain are my sole responsibility.

University of Warwick
December 1982

Contents

Chapter 1

A Preamble and Some Basic Concepts

1.1 The importance of technological change

Technological advance has probably been the major influence on the nature of the lives that we lead relative to the lives that our forebears had and our children and grandchildren will have. In a world that has over the last half-century acquired the ability to undertake space travel, destroy itself through nuclear explosions, transmit messages around the globe by satellite, and travel at twice the speed of sound, it is difficult to conceive of any other factor that has so dramatically shaped our lives. However, we also live in a world where the benefits of new technology do not extend to all, where many of the world's children are starving, and where people die through a lack of appropriate medical care. We live in a world where technological advance introduces fear as well as improved living standards – the fear of nuclear holocaust or the fear of mistakes in genetic engineering may dominate the pleasure we derive from cheaper power or new improved strains of wheat. We see, by looking at our own lives, or by comparing the lives of our children with our own childhood, how much technology has influenced our lives over a relatively short period. Our children take for granted the power of micro-computers and the novelty of video games. They regard the modern motor car as a basic component of life and yawn at the latest space shot. Our children view powerful modern drugs as commonplace. By comparing our own childhood with those of our own children, we can attempt to bring into focus the power of technology. By watching a child and its grandparent together, the strength of technological advance is perhaps even more vividly illustrated. Watching a three-year-old show a seventy-year-old how to operate a video recorder is an illuminating experience.

Do we, however, feel that our own children will have a better life than ourselves? As we observe the environmental pollution that new technology has caused and the changes in social habits that television and other forms of communication have introduced, can the new really be considered to be better than the old? As one hears of the latest advances in weapon warfare can one really believe that advances in peaceful technology are sufficent to compensate for this? Moreover, do we really see that the technological advances of the developed world will improve, or even lengthen, the lives of many children in the poorer nations of the world?

Technological change is often two-edged: it changes many things, often for the better, but these improvements are often achieved at a cost. Why is it that technology keeps advancing, and what is it that leads it to advance in the directions that it does advance in? As it advances, what are the benefits and costs of this advance? These are the questions with which this book is concerned. It would be misleading to even pretend, however, that we have provided answers to the global questions that condition our view towards new technology. Our aims are much more modest. First, we have restricted ourselves largely to the analysis of modern capitalist economies. Second, we have constrained ourselves largely to that part of those economies that might be called the commercial sector. Finally, we have constrained ourselves to the consideration of the economist's contribution to the analysis of the relevant issues. We thus do not squarely face why the world has developed nuclear weapons or undertaken space travel – the decision to proceed with such projects was probably not based on economic reasoning at all. At the same time, we do not discuss the very wide issues of how new technology has impacted on social structures, social mores of behaviour, or society in general. This is not to dispute the importance of such issues, for they are important. My own limitations and the limitations of time and space preclude them. Thus to a large extent the perspectives provided by sociologists, technologists, industrial relations researchers and others are not considered in this volume.

The economist's viewpoint is of course concerned with social welfare and the impact of technology on this, but practicalities largely limit the concepts of social welfare discussed. Our discussions of the impact of technological change are largely centred on output, employment, growth, and income distribution. This is justified on the grounds that these factors impinge on social welfare, and although they may not give us the whole picture, economic wellbeing for which they are a proxy is a major component of social welfare.

The justification for presenting the economic point of view is two-fold.

1 Unlike many social sciences, economics has a number of well defined conceptual and theoretical frameworks that enable one to structure questions and provide answers to relevant questions. These answers may not be precise and they may not be definitive, but they represent an important contribution to an important debate.
2 Over the last few years the economic contribution to the analysis of technological change has grown apace, and it appears to be a particularly appropriate time to bring together these contributions.

The time is appropriate on a number of grounds. From a European perspective, we are at present going through one of the periodic explosions of interest in the impact of particular new technologies. For the last five years a major topic of concern within much of Europe has been the impact that new micro-electronic or information technology will have on national economies. In fact, 1982, the year in which most of this volume was written, was Information Technology Year in the UK. Such periodic bouts of interest at the national level with regard to particular new technologies seem to appear every twenty years or so. The last such occasion concerned the then new computer technology. In the present discussion three main issues have been at the centre of the debate:

1 What impact will the new technology have on employment?

2 What impact will the new technology have on the international division of labour and wealth?
3 What opportunities does the new technology offer to improve upon or maintain previous levels of economic performance?

The first of these topics is particularly relevant when unemployment (in the UK) is at record high levels. In the UK also, after having undergone a long period of relative industrial decline, the fear that the new technology will accelerate this decline, plus the belief that new technology could reverse it, means that a particularly heavy emphasis is being placed on the importance of new technology in economic development.

In such discussions the majority of the emphasis is naturally on the impacts of new technology. However, the impact will be at the end of a long chain of economic decision-making involving the generation and implementation of new technology. To consider only the impacts would be misleading. To consider only the impacts would suggest that new technology arrives *deus ex machina*, whereas the processes of research and development and selection and rejection of technologies will shape the nature of new technologies coming forth, the rate at which they arise, and the origin of the new technologies. Moreover, the time it takes for a new technology to have an impact may be long: it can take up to twenty years for it to be used extensively, and if one is to gain any insight into what impacts a technology is going to have, one must investigate the process by which it spreads. A concern about impacts thus necessitates an investigation of the whole process of technological change.

It is the grand vision of this whole process that we try to represent in this volume. In Part I we look at the generation of new technology and in Part II we consider its spread. In Part III we discuss the impact of new technology. To fill in this division with greater detail, it is useful first to explore some of the basic representations of technological change that have been used in the literature, which will also provide building blocks for later work. Having done so, we can then proceed to explain more fully the structure of the book.

1.2 Some basic concepts

We do not have a neat, completely general definition of the concept of technological change; thus we turn to some of the representations of technological change that have been used in the literature. In the neoclassical literature one defines a production function, $Q = F(K;L;t)$ where K and L are capital and labour inputs, Q is output, and t is time; technological change is the process justifying t in this function. On a more specific level, technological change is a change in the economy's information set detailing the relationships between inputs and outputs in the economy. To be even more concrete, technological change is the process by which economies change over time in respect of the products they produce and the processes used to produce them. To continue such representations is not, however, particularly instructive in itself. What is more interesting is to consider some of the distinctions between types of technological change that have been advanced, and thereby to illuminate the whole issue.

Consider first the neoclassical concepts as summarized by t in the production function. If technology is so represented, then a technological advance (as opposed to merely a change) enables the economy to obtain greater outputs from the same inputs as time proceeds. The major distinctions that have been derived from this approach are two fold: the first is the distinction between embodied and disembodied technical change; the second concerns the bias or direction of technological change.

Technological advance is disembodied if, 'independent of any changes in the factor inputs, the isoquant contours of the production function shift towards the origin as time passes' (Burmeister and Dobell, 1970, p. 66). This is often referred to as technological change of the 'manna from heaven' type. Embodied technological change comprises improvements that can be introduced only by investment in new equipment and skills, with old equipment left unenhanced. The new technology 'is built into or embodied in new capital equipment or newly trained or retrained labour' (Burmeister and Dobell, 1970, p. 66). Disembodied change is automatic; embodied change is not. The rate of embodied change and its direction will be the result of numerous forces in the economy, to the analysis of which a large part of this book is directed.

To illustrate the concept of bias, consider disembodied change. Write the production function as

$$Q = F(K, L; t). \tag{1.1}$$

Allow that the production function has positive first- and negative second-order derivatives with respect to K and L for a given t, and also has the property that it can be rewritten in per capita form (i.e., is homogeneous of degree one) as

$$q = f(k; t) \tag{1.2}$$

where $q = Q/L$ and $k = K/L$.

We now consider that (1.1) can be written as

$$Q = F(K, L; t) = G\{b(t)K, a(t)L\}, \tag{1.3}$$

in which case technological change is known as the factor-augmenting type. Using the dot convention for representation of a derivative with respect to time, if $\dot{b}(t) > 0$ and $\dot{a}(t) = 0$, the change is purely capital-augmenting, and if $\dot{a}(t) > 0$, $\dot{b}(t) = 0$, the change is purely labour-augmenting. If $\dot{a}(t) = \dot{b}(t)$, then the change is equally labour- and capital-augmenting.

To define the bias of technological change we must first define neutrality. A neutral technical change is one that shifts the production function in such a way as to leave undisturbed the balance between capital and labour in current production. The most common definition of this balance refers to factor shares in national income. A change is neutral if factor shares are not affected. However, the development of factor shares will also depend on how the economy is developing over time. We thus have different definitions of neutrality depending on how the economy is developing.

1. If we consider an economy developing such that the capital–output ratio remains constant, then 'Harrod neutrality' exists if factor shares are not affected by technological change.
2. If we consider an economy developing such that the labour–output ratio is constant, then 'Solow neutrality' exists if factor shares do not change with technological change.
3. If we consider an economy developing such that the capital–labour ratio is constant, then 'Hicks neutrality' exists if factor shares do not change with technological change.

Thus, Harrod neutrality requires that, for a given capital–output ratio, the shift in the production function is such as to keep the marginal product of capital constant. To show this we define capital's share in national income as Π. Then, assuming factors are paid their marginal products,

$$\Pi = \frac{\partial f}{\partial k} \cdot \frac{k}{q}$$

If $d(k/q)/dt = 0$ and Harrod neutrality holds, then $\partial f/dk$ *is constant and* $\dot{\Pi} = 0$. It can be shown (Burmeister and Dobell, 1969) that Harrod neutrality requires the production function to be of the form of (1.4); i.e., $b(t) = 1, \dot{b}(t) = 0$:

$$Q = G\{K, a(t)L\}. \tag{1.4}$$

For obvious reasons Harrod-neutral progress is called labour-augmenting.

By similar arguments we may show that Solow neutrality will occur if, for a given output–labour ratio, the marginal product of labour does not change with a change in technology. In this case the production function can be written as

$$Q = G\{b(t)K, L\} \tag{1.5}$$

and technical change is purely capital-augmenting.

In the Hicksian case, technical change is Hicks-neutral if, for a given capital–labour ratio, the marginal rate of substitution between capital and labour does not change with a change in technology. In this case the production function can be written as (1.6), where $c(t) = a(t) = b(t)$:

$$Q = c(t)\,G(K, L). \tag{1.6}$$

Expression (1.6) represents equal capital and labour augmentation and is often called product-augmenting technical change.

These three types of neutrality thus refer to pure labour-augmenting, pure capital-augmenting, and product-augmenting technical change. One should not confuse these concepts with technological change brought about by using better labour, better machines, or by making better products. A change could be pure labour-augmenting although it is realized by introducing better machines. All one is doing is characterizing the shift in the isoquant or production function.

With these definitions of neutrality we can now consider bias. A technological change is labour-saving if the relative share of labour falls along the path with respect to which neutrality is defined. A technological change is capital-saving if the relative share of capital falls. The logic of this classification is that a labour-saving change with a falling labour share implies that a given labour force can be employed only at lower wages after the change and thus must be in less demand. Now, although these definitions are essentially macroeconomic-based, shifts in micro-production functions can be classified using the same definitions of neutrality. Moreover, one should note that what is labour- or capital-saving under one definition of neutrality may not be under another.

We have then three definitions of neutrality. We now state the following two properties.

1 Only under Harrod neutrality can a steady-state growth path in an economy be maintained.
2 Two types of technical progress are compatible if there exists a production function that represents both concepts.

The three concepts above are compatible and the production function within which they are compatible is the Cobb-Douglas. Let technical progress proceed at a given proportional rate λ, and write the Cobb-Douglas as

$$Q = e^{\lambda t} K^{\alpha} L^{1-\alpha}. \tag{1.7}$$

If technical progress is Harrod-neutral at rate m, then

$$Q = K^{\alpha} (Le^{mt})^{1-\alpha} = e^{m(1-\alpha)t} K^{\alpha} L^{1-\alpha} \tag{1.8}$$

and $\lambda = m(1-\alpha)$.

If technical progress is Solow-neutral at rate m, then

$$Q = (Ke^{mt})^{\alpha} L^{1-\alpha} = e^{m\alpha t} K^{\alpha} L^{1-\alpha} \tag{1.9}$$

and $\lambda = m\alpha$.

If technical progress is Hicks-neutral at rate m, then

$$Q = (Ke^{mt})^{\alpha} (Le^{mt})^{1-\alpha} = e^{mt} K^{\alpha} L^{1-\alpha} \tag{1.10}$$

and $\lambda = m$.

Thus (1.7) is at the same time Harrod-, Hicks-, and Solow-neutral. Only the Cobb-Douglas has this property (Uzawa, 1961).

We therefore have three concepts of neutrality. There are however others in the literature. To the three above we can add the following (Gehrig, 1980).

1 For a constant capital-output ratio the marginal product of labour is constant. This requires a production function of the form

$$F(K, L; t) = G\{K, L + b(t)K\} \tag{1.11}$$

and is known as labour-combining.

2 For a constant output-labour ratio the marginal product of capital is constant. This requires

$$F(K, L; t) = G\{K + a(t)L, L\} \tag{1.12}$$

and is known as capital combining.

3 At a constant capital-labour ratio the marginal product of labour is constant, yielding

$$F(K, L; t) = b(t)K + G(K, L), \tag{1.13}$$

known as capital-additive.

4 At a constant capital-labour ratio the marginal product of capital is constant, yielding

$$F(K, L; t) = a(t)L + G(K, L) \tag{1.14}$$

and known as labour-additive.

These variants have not however received the attention in the literature given to the original three.

We have defined these concepts of bias for *disembodied* technical change. Consider now that technical change is *embodied*. We identify a machine by a number ν, denoting its date of manufacture or vintage. Machines of different vintages are assumed to be operated independently. We then can define a vintage production function that relates output to inputs for machines of vintage ν. Thus, if we let $Q(\nu, t)$ be output on vintage ν at time t, $K_\nu(t)$ be capital stock of vintage ν in time t, and $L_\nu(t)$ be labour employed on vintage ν at time t, we write the production function as

$$Q(\nu, t) = F\{K_\nu(t), L_\nu(t), \nu\}. \tag{1.15}$$

This relation stresses the point that it is date of manufacture that affects productivity and not time. Definitions of neutrality follow as above, but now are taken with respect to different vintages rather than times.

This vintage approach with its concept of embodiment does take us away from the automaticity of the disembodied approach; however, it is still open to two basic objections.

1 No distinction is made between new technology that changes processes and that which changes products. The product/process distinction can be an important one. Of course, in an interrelated economy one industry's new products can be another's new process, but especially when we consider technological change at the micro-level the product/process distinction can be important.

2 The production function approach does not tell us anything of the sources of technological change. Why, for example, is vintage ν equipment better than that of vintage $\nu - 1$? To investigate this further we turn to the Schumpeter trilogy.

Schumpeter defined three phases in the process of technological change: invention, innovation, and diffusion. Using Freeman's (1974) definitions, invention is 'an idea, a sketch or a model for a new improved device, product process or system. Such inventions may often (not always) be patented but they do not necessarily lead to technical innovation. . . . An innovation, in the economic sense, is accomplished only with the first commercial transaction involving the new product, process, system or device.' Unfortunately, innovation is often used to describe either the whole process of technological change or an act that is original to the decision-maker under discussion but not to the economy as a whole. In general, we shall use the term as defined above. The third part of the trilogy is diffusion, which occurs after the invention and innovation stages and refers to the process by which the innovation spreads across the market.

In this trilogy, patenting comes in at the invention stage and inventions will be produced partly through expenditure on research and development (R & D). However, the step from invention to innovation also depends on R & D. In the UK, R & D expenditures have split on average in the postwar period in the following proportions: basic research, 3 per cent; applied research, 22 per cent; development, 75 per cent. Although it is not a completely accurate distinction, development expenditures could be associated with innovation, suggesting that most R & D is an input to the innovation process. However, unless the user of the innovation is the same firm (individual) as the producer, we would expect there be two parties to an innovation – a commercial transaction has two sides. Only one of these sides – the producer – may be a major spender of R & D funds.

The innovation process is sometimes referred to as additions to the economy's book of blueprints; however to confuse matters further, this blueprint terminology is sometimes used to refer to invention, and then only certain blueprints are selected for development to the innovation stage. However, once we have innovations, these start to have an impact on the economy as they are used or produced. This is the process of diffusion. As diffusion proceeds we may think of the disembodied production function shifting. Alternatively, we may think of the embodied production function being different for different vintages because of innovation, and the process of diffusion concerns how machines of vintage v take over from those of other vintages. We do not wish to carry this parallel too far, however, for diffusion has also been analysed outside the context of the vintage model.

1.3 An outline

We can now proceed to outline the structure of the book. Part I, on the generation of new technology, is divided into three chapters. The first looks at the patent system, patenting activity as a proxy for invention, and innovation. The second considers the determination of R & D expenditures, and the third is concerned with the determination of the capital or labour saving bias in technological change. The three chapters as a whole represent an attempt to look at