Review of Economic Studies (2002) **69**, 781–809
0034-6527/02/00310781$02.00

Directed Technical Change

DARON ACEMOGLU
MIT

First version received May 2001; *final version accepted February* 2002 (*Eds.*)

For many problems in macroeconomics, development economics, labour economics, and international trade, whether technical change is biased towards particular factors is of central importance. This paper develops a simple framework to analyse the forces that shape these biases. There are two major forces affecting equilibrium bias: the price effect and the market size effect. While the former encourages innovations directed at scarce factors, the latter leads to technical change favouring abundant factors. The elasticity of substitution between different factors regulates how powerful these effects are, determining how technical change and factor prices respond to changes in relative supplies. If the elasticity of substitution is sufficiently large, the long run relative demand for a factor can slope up.

I apply this framework to develop possible explanations to the following questions: why technical change over the past 60 years was skill biased, and why the skill bias may have accelerated over the past 25 years? Why new technologies introduced during the late eighteenth and early nineteenth centuries were unskill biased? What is the effect of biased technical change on the income gap between rich and poor countries? Does international trade affect the skill bias of technical change? What are the implications of wage push for technical change? Why is technical change generally labour augmenting rather than capital augmenting?

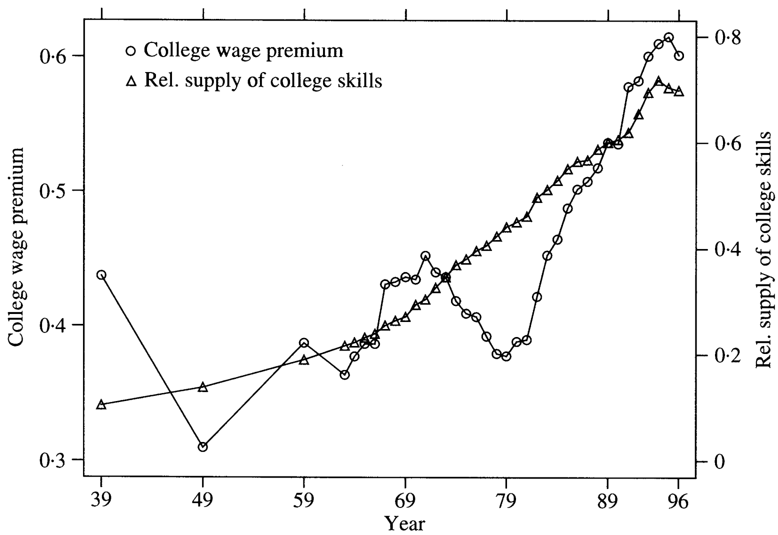
1. INTRODUCTION

There is now a large and influential literature on the determinants of the aggregate technical progress (see, among others, Romer (1990), Segerstrom, Anant and Dinopoulos (1990), Grossman and Helpman (1991), Aghion and Howitt (1992), Young (1993)). This literature does not address questions related to the *direction* and *bias* of technical change. In most situations, however, technical change is not neutral: it benefits some factors of production more than others. In this paper, I develop a simple framework of directed technical change to study these biases. In this framework, profit incentives determine the amount of research and development directed at different factors and sectors.

To see the potential importance of the biases, consider a number of examples:

(1) Figure 1 plots a measure of the relative supply of skills and a measure of the return to skills, the college premium. It shows that over the past 60 years, the U.S. relative supply of skills has increased rapidly, but there has been no tendency for the returns to college to fall in the face of this large increase in supply—on the contrary, there has been an increase in the college premium over this time period. The standard explanation for this pattern is that new technologies over the post-war period have been *skill biased*. The figure also shows that beginning in the late 1960's, the relative supply of skills increased more rapidly than before, and the skill premium increased sharply beginning in the late 1970's. The standard explanation for this pattern is an acceleration in the skill bias of technical change (*e.g.* Autor, Krueger and Katz (1998)).

(2) In contrast, technical change during the late eighteenth and early nineteenth centuries appears to have been *unskill biased* (skill replacing). The artisan shop was replaced by the factory and later by interchangeable parts and the assembly line (*e.g.* James and Skinner (1985), Goldin and Katz (1998)). Products previously manufactured by skilled



Relative Supply of College Skills and College Premium

FIGURE 1

The behaviour of the (log) college premium and relative supply of college skills (weeks worked by college equivalents divided by weeks worked by noncollege equivalents) between 1939 and 1996. Data from March CPSs and 1940, 1950 and 1960 censuses

artisans started to be produced in factories by workers with relatively few skills, and many previously complex tasks were simplified, reducing the demand for skilled workers (*e.g.* Mokyr (1990, p. 137)).

(3) Over the past 150 years of growth, the prices of the two key factors, capital and labour, have behaved very differently. While both in the U.S. and in other Western economies, the wage rate has increased steadily, the rental rate of capital has been approximately constant. This pattern indicates that most of the new technologies are *labour augmenting*.

(4) Beginning in the late 1960's and the early 1970's, both unemployment and the share of labour in national income increased rapidly in a number of continental European countries. During the 1980's, unemployment continued to increase, but the labour share started a steep decline, and in many countries, ended up below its initial level. Blanchard (1997) interprets the first phase as the response of these economies to a wage push, and the second phase as a possible consequence of *capital-biased* technical change.

These examples document a variety of important macroeconomic issues where biased technical change plays a key role. They also pose a number of questions: why has technical change been skill biased over the past 60 years? Why was technical change biased in favour of unskilled labour and against skilled artisans during the nineteenth century? Why has there been an acceleration in the skill bias of technical change during the past 25 years? Why is much of technological progress labour augmenting rather than capital augmenting? Why was there rapid capital-biased technical change in continental Europe following the wage push by workers during the 1970's?

These questions require a framework where the equilibrium bias of technical change can be studied. The framework I present for this purpose generalizes the existing endogenous technical change models to allow for technical change to be directed towards different factors: firms can invest resources to develop technologies that complement a particular factor. The relative profitabilities of the different types of technologies determine the direction of technical change.

I show that there are two competing forces determining the relative profitability of different types of innovation: (i) the price effect, which creates incentives to develop technologies used in the production of more expensive goods (or equivalently, technologies using more expensive factors); (ii) the market size effect, which encourages the development of technologies that have a larger market, more specifically, technologies that use the more abundant factor. These two effects are competing because, while the price effect implies that there will be more rapid technological improvements favouring scarce factors, the market size effect creates a force towards innovations complementing the abundant factor.[1] I will show that the elasticity of substitution between the factors determines the relative strengths of these two effects. When the elasticity of substitution is low, scarce factors command higher prices, and the price effect is relatively more powerful.

The first major result of this framework is a "weak induced-bias hypothesis": irrespective of the elasticity of substitution between factors (as long as it is not equal to 1), an increase in the relative abundance of a factor creates some amount of technical change biased towards that factor. The second major result is a "strong induced-bias hypothesis", and states that if the elasticity of substitution is sufficiently large (in particular, greater than a certain threshold between 1 and 2), the induced bias in technology can overcome the usual substitution effect and increase the relative reward to the factor that has become more abundant. That is, directed technical change can make the long-run relative demand curve slope up. The long run relative demand curve may be upward sloping in this set-up because of the underlying "increasing returns to scale" in the R&D process: a new machine, once invented, can be used by many workers.[2]

Figure 2 illustrates these results diagrammatically. The relatively steep downward-sloping lines show the constant-technology relative demand curves. The economy starts at point A. In the absence of endogenous technical change, the increase in the supply shown in the figure moves the economy along the constant-technology demand to point B. The first result of this framework, the weak induced-bias hypothesis, implies that, as long as the elasticity of substitution between factors is different from 1, the increase in the supply will induce biased technical change and shift the constant-technology demand curve out. The economy will therefore settle to a point like C. In other words, the (long-run) endogenous-technology demand curve will be flatter than the constant-technology curve. The second result, the strong induced-bias hypothesis, implies that the induced bias in technology can be powerful enough to create a sufficiently large shift in the constant-technology demand curve and take the economy to a point like D. In this case, the endogenous-technology demand curve of the economy is upward sloping and the relative reward of the factor that has become more abundant increases.

After outlining the general forces shaping the direction of technical change and the main results, I return to a number of applications of this framework. I discuss: (1) why technical change over the past 60 years was skill-biased, and why skill-biased technical change may have accelerated over the past 25 years. Also why new technologies introduced during the late eighteenth and early nineteenth centuries were labour biased. (2) Why biased technical change is likely to increase the income gap between rich and poor countries. (3) Why international trade may induce skill-biased technical change. (4) Under what circumstances labour scarcity will spur faster technological progress as suggested by Rothbarth (1946) and Habakkuk (1962). (5) Why technical change tends to be generally labour augmenting rather than capital augmenting.

1. Another important determinant of the direction of technical change is the form of the "innovation possibilities frontier"—*i.e.* how the relative costs of innovation are affected as technologies change. I discuss the impact of the innovation possibilities frontier on the direction of technical change in Section 4.

2. This is related to the nonrivalry in the use of ideas emphasized by Romer (1990).

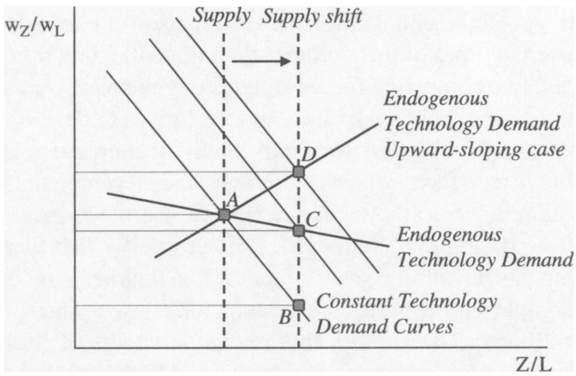


FIGURE 2

Constant technology and endogenous technology relative demand curves. Constant technology: A → B. Endogenous technology: A → C. Endogenous technology with powerful market size effect: A → D

(6) Why a large wage push, as in continental Europe during the 1970's, may cause capital-biased technical change and affect the factor distribution of income.

This list is by no means exhaustive, and there is much research to be done to understand the implications of biased technical change and the determinants of equilibrium bias of new technologies. It is part of my aim in this paper to stress the importance of thinking about biased technical change, and to provide a set of tools that are likely to be useful for future research on these biases.[3]

Although there is relatively little current research on biased technical change, an earlier literature was devoted to studying related issues. It was probably Hicks in *The Theory of Wages* (1932) who first discussed the determinants of equilibrium bias.[4] He wrote: "A change in the relative prices of the factors of production is itself a spur to invention, and to invention of a particular kind—directed to economizing the use of a factor which has become relatively expensive." (pp. 124–125). Hicks' reasoning, that technical change would attempt to economize on the more expensive factor, was criticized by Salter (1966) who pointed out that there was no particular reason for saving on the more expensive factor—firms would welcome all cost reductions. Moreover, the concept of "more expensive factor" did not make much sense, since all factors were supposed to be paid their marginal product.

These questions were revived by the "induced innovation" literature. An important paper by Kennedy (1964) introduced the concept of "innovation possibilities frontier", capturing the trade-off between different types of innovations, and argued that it is the form of this frontier—rather than the shape of a given neoclassical production function—that determines the factor distribution of income. Kennedy, furthermore, argued that induced innovations would push the economy to an equilibrium with a constant relative factor share (see also Samuelson (1965), Drandakis and Phelps (1965)). Around the same time, Habakkuk (1962) put forth the thesis that labour scarcity, by increasing wages, induced firms to search for labour saving inventions and

3. The framework here focuses on one type of biased technical change: that which increases the relative productivity of one factor permanently. Alternatively, as emphasized by Nelson and Phelps (1966), Schultz (1975) and especially Galor and Maov (2000), technological progress may be biased towards one of the factors, skilled labour, in the short run, because higher ability and skilled workers may be better at adapting to a changing environment.

4. Marx also touched on these issues. He argued that labour scarcity—the exhaustion of the reserve army of labour—would induce the capitalist to substitute machinery for labour and spur growth. See for example the discussion in Habakkuk (1962, p. 44).

spurred technological progress (see also Rothbarth (1946)). This literature was also criticized for lack of micro-foundations, however. First, with specifications as in Kennedy, the production function at the firm level exhibited increasing returns to scale because, in addition to factor quantities, firms could choose "technology quantities". Second, as pointed out by Nordhaus (1973), it was not clear who undertook the R&D activities and how they were financed and priced. These shortcomings reduced the interest in this literature, and there was little research for almost 30 years.

The analysis here, instead, starts from the explicit micro-foundations laid out by the endogenous technical change models. In addition to providing an equilibrium framework for analysing these issues, I demonstrate the presence of the market size effect, which did not feature in the earlier literature (but see Schmookler (1966)). More explicitly, the framework I present here synthesizes my previous work in Acemoglu (1998, 1999*a*,*b*) and Acemoglu and Zilibotti (2001), as well as work by Kiley (1999) (see also Lloyd-Ellis (1999) and Galor and Maov (2000), for different perspectives). The results in these papers show that whether technical change results from quality improvements, expanding variety of products, or expanding variety of machine types is not essential. For this reason, I choose one of the specifications and highlight the modelling choice that turns out to be more important—the form of the innovation possibilities frontier.

The rest of the paper is organized as follows. In the next section, I define some of the terms that will be used throughout the paper and clarify the distinction between factor augmenting and factor-biased technical change. In this section, I also give a brief overview of the main results. In Section 3, I introduce the basic framework that determines the demand for innovation and I highlight the price and market size effects on the direction of technical change. Section 4 introduces the innovation possibilities frontier and shows how different forms of this frontier affect the equilibrium bias of technology. Sections 5 and 6 apply the framework developed in Sections 3 and 4 to a variety of situations where biased technical change appears to be important. Section 7 concludes.

2. FACTOR-AUGMENTING, FACTOR-BIASED TECHNICAL CHANGE AND AN OVERVIEW

Consider an aggregate production function, $F(L, Z, A)$, with two inputs, L, labour, and Z, which could be capital, skilled labour or land. A is a technology index. Without loss of generality imagine that $\partial F/\partial A > 0$, so a greater level of A corresponds to "better technology" or to "technological progress". Technical change is *L-augmenting* if the production function takes the more special form $F(AL, Z)$. *Z-augmenting* technical change is defined similarly. Technical change is *L-biased*, on the other hand, if

$$\frac{\partial \frac{\partial F/\partial L}{\partial F/\partial Z}}{\partial A} > 0,$$

that is, if technical change increases the marginal product of L more than that of Z.

To clarify the difference between these two concepts, consider the more specialized constant elasticity of substitution (CES) production function

$$y = \left[\gamma(A_L L)^{\frac{\sigma-1}{\sigma}} + (1-\gamma)(A_Z Z)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}},$$

where A_L and A_Z are two separate technology terms, $\gamma \in (0, 1)$ is a distribution parameter which determines how important the two factors are, and $\sigma \in (0, \infty)$ is the elasticity of substitution between the two factors. When $\sigma = \infty$, the two factors are perfect substitutes, and the production function is linear. When $\sigma = 1$, the production function is Cobb–Douglas, and when $\sigma = 0$, there is no substitution between the two factors, and the production function is Leontieff. When $\sigma > 1$, I refer to the factors as gross substitutes, and when $\sigma < 1$, I refer to them as gross complements.[5] By construction, A_L is L-augmenting and A_Z is Z-augmenting. I will also sometimes refer to A_L as labour complementary, and to A_Z as Z-complementary.

Whether technical change is labour biased or Z-biased, on the other hand, depends on the elasticity of substitution. To see this, calculate the relative marginal product of the two factors:

$$\frac{MP_Z}{MP_L} = \frac{1-\gamma}{\gamma}\left(\frac{A_Z}{A_L}\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{Z}{L}\right)^{-\frac{1}{\sigma}}. \tag{1}$$

The relative marginal product of Z is decreasing in the relative abundance of Z, Z/L. This is the usual *substitution effect*, leading to a downward sloping relative demand curve. The effect of A_Z on this relative marginal product depends on σ, however. If $\sigma > 1$, an increase in A_Z (relative to A_L) increases the relative marginal product of Z. When $\sigma < 1$, an increase in A_Z reduces the relative marginal product of Z. Therefore, when the two factors are gross substitutes, Z-augmenting (Z-complementary) technical change is also Z-biased. In contrast, when the two factors are gross complements, then Z-augmenting technical change is L-biased. Naturally, when $\sigma = 1$, we are in the Cobb–Douglas case, and neither a change in A_Z nor in A_L is biased towards any of the factors.

The intuition for why, when $\sigma < 1$, Z-augmenting technical change is L-biased is simple: with gross complementarity, an increase in the productivity of Z increases the demand for the other factor, labour, by more than the demand for Z, effectively creating "excess demand" for labour. As a result, the marginal product of labour increases by more than the marginal product of Z.

Now to obtain an overview of the results that will follow, imagine a set-up where A_L and A_Z are determined endogenously from the type and quality of machines supplied by "technology monopolists". One of the major results of the more detailed analysis below will be that the profitability of developing new Z-complementary machines relative to the profitability of labour-complementary machines will be proportional to (see equation (17))

$$\left(\frac{A_Z}{A_L}\right)^{-\frac{1}{\sigma}}\left(\frac{Z}{L}\right)^{\frac{\sigma-1}{\sigma}}. \tag{2}$$

The basic premise of the approach here is that profit incentives determine what types of innovations will be developed. So when (2) is high, A_Z will increase relative to A_L. Inspection of (2) shows that when the two factors are gross substitutes ($\sigma > 1$), an increase in Z/L will increase the relative profitability of inventing Z-complementary technologies. To equilibrate innovation incentives, A_Z/A_L has to rise, reducing (2) back to its original level. Intuitively, in this case, of the two forces discussed in the introduction, the market size effect is more powerful than the price effect, so technical change is directed towards the more abundant factor. In contrast, when the two factors are gross complements ($\sigma < 1$), an increase in Z/L will lead to a fall in A_Z/A_L. However, recall that when $\sigma < 1$, a lower A_Z/A_L corresponds to Z-biased technical change. So in this case an increase in Z/L reduces the relative *physical* productivity of factor Z, but increases its relative *value* of marginal product, because of relative price changes. Therefore, in both cases, an increase in the relative abundance of Z causes Z-biased technical change. We will also see that if σ is sufficiently large, this induced biased technical change can be so powerful that the increase in the relative abundance of a factor may in fact increase its relative reward—*i.e.* the long run relative demand curve for a factor may be upward sloping.

5. I use this terminology because the demand for Z increases in response to an increase in the price of L, w_L, holding its price, w_Z, and the quantity of L constant if and only if $\sigma > 1$, and vice versa.

3. THE DEMAND SIDE

I now develop the basic framework for analysing the determinants of the factor bias of technical change, first focusing on the demand for new technology (innovation). The next section then introduces "the innovation possibilities frontier" and discusses the supply side of innovations.

3.1. *The environment*

Consider an economy that admits a representative consumer with the usual constant relative risk aversion (CRRA) preferences

$$\int_0^\infty \frac{C(t)^{1-\theta}-1}{1-\theta}e^{-\rho t}dt, \tag{3}$$

where ρ is the rate of time preference and θ is the coefficient of relative risk aversion (or the intertemporal elasticity of substitution). I suppress the time arguments to simplify the notation, and I will do so throughout as long as this causes no confusion. The budget constraint of the consumer is

$$C+I+R \le Y \equiv \left[\gamma Y_L^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma)Y_Z^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{4}$$

where I denotes investment, and R is total R&D expenditure. I also impose the usual no-Ponzi game condition, requiring the lifetime budget constraint of the representative consumer to be satisfied. The production function in (4) implies that consumption, investment and R&D expenditure come out of an output aggregate produced from two goods, Y_L and Y_Z, with elasticity of substitution ε, and γ is a distribution parameter which determines how important the two goods are in aggregate production. Of these two goods, Y_L is (unskilled) labour intensive, while Y_Z uses another factor, Z, intensively. In this section and the next, I will not be specific about what this factor is, but the reader may want to think of it as skilled labour for concreteness.

These two goods have the following production functions[6]

$$Y_L = \frac{1}{1-\beta}\left(\int_0^{N_L} x_L(j)^{1-\beta}dj\right)L^\beta, \tag{5}$$

and

$$Y_Z = \frac{1}{1-\beta}\left(\int_0^{N_Z} x_Z(j)^{1-\beta}dj\right)Z^\beta, \tag{6}$$

where $\beta \in (0,1)$, and L and Z are the total quantities of the two factors, assumed to be supplied inelastically for now. The labour-intensive good is therefore produced from labour and a range of labour-complementary intermediates or machines, the x_L's. For simplicity, I will refer to the x's as "machines". The range of machines that can be used with labour is denoted by N_L. The production function for the other good, (6), uses Z-complementary machines and is explained similarly. Notice that given N_L and N_Z, the production functions (5) and (6) exhibit constant returns to scale. There will be aggregate increasing returns, however, when N_L and N_Z are endogenized.

I assume that machines in both sectors are supplied by "technology monopolists". In this section, I take N_L and N_Z as given, and in the next section, I analyse the innovation decisions of these monopolists (the supply of innovations) to determine N_L and N_Z. Each monopolist sets a rental price $\chi_L(j)$ or $\chi_Z(j)$ for the machine it supplies to the market. For simplicity, I assume

6. The firm level production functions are also assumed to exhibit constant returns to scale, so there is no loss of generality in focusing on the aggregate production functions.

that all machines depreciate fully after use, and that the marginal cost of production is the same for all machines and equal to ψ in terms of the final good.[7]

The important point to bear in mind is that the set of machines used in the production of the two goods are different, allowing technical change to be biased. The range of machines, N_L and N_Z, determine aggregate productivity, while N_Z/N_L determines the relative productivity of factor Z.

3.2. *Equilibrium*

An equilibrium (given N_L and N_Z) is a set of prices for machines, $\chi_L(j)$ or $\chi_Z(j)$, that maximize the profits of technology monopolists, machine demands from the two sectors, $x_L(j)$ or $x_Z(j)$, that maximize producers' profits, and factor and product prices, w_L, w_Z, p_L, and p_Z, that clear markets. I now characterize this equilibrium and show that it is unique.[8] The levels of N_L and N_Z will be determined in the next section once I introduce the innovation possibilities frontier of this economy.

The product markets for the two goods are competitive, so market clearing implies that their relative price, p, has to satisfy

$$p \equiv \frac{p_Z}{p_L} = \frac{1-\gamma}{\gamma}\left(\frac{Y_Z}{Y_L}\right)^{-\frac{1}{\varepsilon}}. \tag{7}$$

The greater the supply of Y_Z relative to Y_L, the lower is its relative price, p. The response of the relative price to the relative supply depends on the elasticity of substitution, ε.

I choose the price of the final good as the numeraire, so

$$\left[\gamma^{\varepsilon} p_L^{1-\varepsilon} + (1-\gamma)^{\varepsilon} p_Z^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} = 1. \tag{8}$$

Since product markets are competitive, firms in the labour intensive sector solve the following maximization problem

$$\max_{L,\{x_L(j)\}} p_L Y_L - w_L L - \int_0^{N_L} \chi_L(j) x_L(j) dj, \tag{9}$$

taking the price of their product, p_L, and the rental prices of the machines, denoted by $\chi_L(j)$, as well as the range of machines, N_L, as given. The maximization problem facing firms in the Z-intensive sector is similar. The first-order conditions for these problems give machine demands as

$$x_L(j) = \left(\frac{p_L}{\chi_L(j)}\right)^{1/\beta} L \quad \text{and} \quad x_Z(j) = \left(\frac{p_Z}{\chi_Z(j)}\right)^{1/\beta} Z. \tag{10}$$

These equations imply that the desired amount of machine use is increasing in the price of the product, p_L or p_Z, and in the firm's employment, L or Z, and is decreasing in the price of the machine, $\chi_L(j)$ or $\chi_Z(j)$. Intuitively, a greater price for the product increases the value of the marginal product of all factors, including that of machines, encouraging firms to rent more

7. Slow depreciation of machines does not change the BGP equilibrium, and only affects the speed of transitional dynamics. For example, if machines depreciate at some exponential rate δ, monopolists will produce the required stock of machines after the discovery of the new variety, and then will replace the machines that depreciate. The rental price will then be a mark-up over the opportunity cost of machines rather than over the production cost. To keep notation to a minimum, I do not consider the case of slow depreciation.

8. In this paper, I only characterize equilibrium allocations. The social planner's solution differs from equilibrium allocations because of monopoly distortions. However, exactly the same equations, in particular, equations (21) and (26), determine the bias of technology in the social planner's allocation. Details available upon request.

machines. A greater level of employment, on the other hand, implies more workers to use the machines, raising demand. Finally, because the demand curve for machines is downward sloping, a higher cost implies lower demand.

Next, the first-order condition with respect to L and Z gives the factor rewards as

$$w_L = \frac{\beta}{1-\beta} p_L \left(\int_0^{N_L} x_L(j)^{1-\beta} dj \right) L^{\beta-1} \qquad \text{and}$$
$$w_Z = \frac{\beta}{1-\beta} p_Z \left(\int_0^{N_Z} x_Z(j)^{1-\beta} dj \right) Z^{\beta-1}. \tag{11}$$

My interest is with the determinants of the direction of technical change. As discussed above, profit-maximizing firms will generate more innovations in response to greater profits, so the first step is to look at the profits of the technology monopolists. Recall that each monopolist faces a marginal cost of producing machines equal to ψ. Therefore, the profits of a monopolist supplying labour-intensive machine j can be written as $\pi_L(j) = (\chi_L(j) - \psi)x_L(j)$. Since the demand curve for machines facing the monopolist, (10), is iso-elastic, the profit-maximizing price will be a constant mark-up over marginal cost: $\chi_L(j) = \frac{\psi}{1-\beta}$. To simplify the algebra, I normalize the marginal cost to $\psi \equiv 1-\beta$.[9] This implies that in equilibrium all machine prices will be given by $\chi_L(j) = \chi_Z(j) = 1$. Using these prices and the machine demands above, profits of technology monopolists are obtained as

$$\pi_L = \beta p_L^{1/\beta} L \qquad \text{and} \qquad \pi_Z = \beta p_Z^{1/\beta} Z. \tag{12}$$

What is relevant for the monopolists is not the instantaneous profits, but the net present discounted value of profits. These net present discounted values can be expressed using standard dynamic programming equations:

$$rV_L - \dot{V}_L = \pi_L \qquad \text{and} \qquad rV_Z - \dot{V}_Z = \pi_Z, \tag{13}$$

where r is the interest rate, which is potentially time varying. The equations relate the present discounted value of future profits, V, to the flow of profits, π. The $\dot{V}$ term takes care of the fact that future profits may not equal current profits, for example because prices are changing.

To gain intuition, let us start with a steady state where the $\dot{V}$ terms are 0 (*i.e.* profits and the interest rate are constant in the future). Then,

$$V_L = \frac{\beta p_L^{1/\beta} L}{r} \qquad \text{and} \qquad V_Z = \frac{\beta p_Z^{1/\beta} Z}{r}. \tag{14}$$

The greater V_Z is relative to V_L, the greater are the incentives to develop Z-complementary machines, N_Z, rather than labour-complementary machines, N_L. Inspection of (14) reveals two forces determining the direction of technical change:

(1) The price effect: there will be a greater incentive to invent technologies producing more expensive goods, as shown by the fact that V_Z and V_L are increasing in p_Z and p_L.
(2) The market size effect: a larger market for the technology leads to more innovation. Since the market for a technology consists of the workers who use it, the market size effect encourages innovation for the more abundant factor. This can be seen from the fact that V_Z and V_L are increasing in Z and L.

Notice that an increase in the relative factor supply, Z/L, will create both a market size effect and a price effect. The latter simply follows from the fact that an increase in Z/L will

9. This is without loss of any generality, since I am not interested in comparative statics with respect to β.

reduce the relative price $p \equiv p_Z/p_L$. Equilibrium bias in technical change—whether technical change will favour relatively scarce or abundant factors—is determined by these two opposing forces.[10] An additional determinant of equilibrium bias is the form of the innovation possibilities frontier, which will be introduced in the next section.

Next, it is useful to investigate the strength of the price and market size effects in more detail. To do this, let us substitute (10) into the production functions, (5) and (6). This gives

$$Y_L = \frac{1}{1-\beta} p_L^{(1-\beta)/\beta} N_L L \qquad \text{and} \qquad Y_Z = \frac{1}{1-\beta} p_Z^{(1-\beta)/\beta} N_Z Z. \tag{15}$$

Substituting these into (7) and using some algebra, we obtain the relative price of the two goods as a function of relative technology, N_Z/N_L, and the relative factor supply, Z/L:

$$p = \left(\frac{1-\gamma}{\gamma}\right)^{\beta\varepsilon/\sigma} \left(\frac{N_Z Z}{N_L L}\right)^{-\beta/\sigma}, \tag{16}$$

where ε is the elasticity of substitution between the two goods, Y_L and Y_Z, while σ is the (derived) elasticity of substitution between the two factors, Z and L, defined as

$$\sigma \equiv \varepsilon - (\varepsilon - 1)(1-\beta).$$

Note that $\sigma > 1$ if and only if $\varepsilon > 1$—that is, the two factors are gross substitutes only if the two goods are gross substitutes.

Now using (14) and (16) and still assuming steady state, we can write the relative profitability of creating new Z-complementary machines as

$$\frac{V_Z}{V_L} = \underbrace{p^{1/\beta}}_{\text{price effect}} \cdot \underbrace{\frac{Z}{L}}_{\text{market size effect}} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\sigma}} \left(\frac{N_Z}{N_L}\right)^{-\frac{1}{\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-1}{\sigma}}. \tag{17}$$

This expression shows that the relative profitability of the two types of innovation are determined by the price and market size effects. An increase in the relative factor supply, Z/L, will increase V_Z/V_L as long as the elasticity of substitution between factors, σ, is greater than 1 and it will reduce V_Z/V_L if $\sigma < 1$. Therefore, the elasticity of substitution between the two factors (and indirectly between the two goods) regulates whether the price effect dominates the market size effect so that there are greater incentives to improve the (physical) productivity of scarce factors, or whether the market size effect dominates, creating greater incentives to improve the productivity of abundant factors. When the factors are gross substitutes, the market size effect dominates. And when they are gross complements, the price effect dominates.

Finally, to see another important role of the elasticity of substitution, consider the relative factor rewards, w_Z/w_L. Using equations (10) and (11), and then substituting for (16), we have

$$\frac{w_Z}{w_L} = p^{1/\beta}\frac{N_Z}{N_L} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\sigma}} \left(\frac{N_Z}{N_L}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{Z}{L}\right)^{-\frac{1}{\sigma}}. \tag{18}$$

First, note that the relative factor reward, w_Z/w_L, is decreasing in the relative factor supply, Z/L. This is simply the usual substitution effect: the more abundant factor is substituted for the less abundant one, and has a lower marginal product.

10. This discussion emphasizes the role of price and market size effects, while factor prices do not feature in (14). The early induced innovations literature, instead, argued that innovations would be directed at "more expensive" factors (*e.g.* Hicks (1932), Fellner (1961), Habakkuk (1962)). Using equations (10) and (11), we can re-express (14) as $V_L = (1-\beta)w_L L/rN_L$ and $V_Z = (1-\beta)w_Z Z/rN_Z$, which show that an equivalent way of looking at the incentives to develop new technologies is to emphasize factor costs, w_L and w_Z (as well as market sizes, L and Z). As conjectured by the induced innovations literature, there will be more innovation directed at factors that are more expensive.

We also see from equation (18) that the same combination of parameters, $(\sigma - 1)/\sigma$, which determines whether innovation for more abundant factors is more profitable, also determines whether a greater N_Z/N_L increases w_Z/w_L: when $\sigma > 1$, it does, but when $\sigma < 1$, it has the opposite effect and reduces the reward to Z relative to labour. This is because N_Z/N_L is the relative *physical* productivity of the two factors, not their relative *value* of marginal product. The latter also depends on the elasticity of substitution between the two factors (recall equation (1) in Section 2). Therefore, as in Section 2, Z-biased technical change corresponds to an increase in $(N_Z/N_L)^{(\sigma-1)/\sigma}$, not simply to an increase in N_Z/N_L. Consequently, when $\sigma < 1$, a decrease in N_Z/N_L corresponds to Z-biased technical change. This feature, that the relationship between relative physical productivity and the value of marginal product is reversed when the two goods (factors) are gross complements, will play an important role in the discussion below.

4. THE SUPPLY OF INNOVATIONS: THE INNOVATION POSSIBILITIES FRONTIER

The previous section outlined how the production side of the economy determines the return to different types of innovation—the demand for innovation. The other side of this equation is the cost of different innovations, or using the term introduced by Kennedy (1964), the "innovation possibilities frontier". The analysis in this section will reveal that in addition to the elasticity of substitution, the degree of *state dependence* in the innovation possibilities frontier will have an important effect on the direction of technical change. The degree of state dependence relates to how future relative costs of innovation are affected by the current composition of R&D (current "state" of R&D). I refer to the innovation possibilities frontier as "state dependent" when current R&D directed at factor Z reduces the relative cost of Z-complementary R&D in the future.

I follow the endogenous growth literature in limiting attention to innovation possibilities frontiers that allow steady growth in the long run. Sustained growth requires that the innovation possibilities frontier takes one of two forms. The first, which Rivera-Batiz and Romer (1991) refer to as the lab equipment specification, involves only the final good being used in generating new innovations. With this specification, steady-state growth is generated with an intuition similar to the growth model of Rebelo (1991) whereby the key accumulation equation is linear and does not employ the scarce (non-accumulated) factors, such as labour. The second possibility is the knowledge-based R&D specification of Rivera-Batiz and Romer (1991) where spillovers from past research to current productivity are necessary to sustain growth. It is useful to distinguish between these two formulations because they naturally correspond to different degrees of *state dependence* in R&D.

4.1. *The direction of technical change with the lab equipment model*

Consider the following production function for new machine varieties

$$\dot{N}_L = \eta_L R_L \qquad \text{and} \qquad \dot{N}_Z = \eta_Z R_Z, \tag{19}$$

where R_L is spending on R&D for the labour-intensive good (in terms of final good), and R_Z is R&D spending for the Z-intensive good. The parameters η_L and η_Z allow the costs of these two types of innovations to differ. The innovation production functions in (19) imply that one unit of final good spent for R&D directed at labour-complementary machines will generate η_L new varieties of labour-complementary machines (and similarly, one unit of final good directed at Z-complementary machines will generate η_Z new machine types). A firm that discovers a new machine variety receives a perfectly enforced patent on this machine and becomes its sole supplier—the technology monopolist. Notice that with the specification in (19), there is no state dependence: $(\partial \dot{N}_Z/\partial R_Z)/(\partial \dot{N}_L/\partial R_L) = \eta_Z/\eta_L$ is constant irrespective of the levels of N_L and N_Z.

I start with a balanced growth path (BGP)—or steady-state equilibrium—where the prices p_L and p_Z are constant, and N_L and N_Z grow at the same rate. This implies that the $\dot{V}$ terms in equation (13) are 0, and V_Z/V_L is constant. Moreover, this ratio has to be equal to the (inverse) ratio of η's from (19) so that technology monopolists are willing to innovate for both sectors. This immediately implies the following "technology market clearing" condition:[11]

$$\eta_L \pi_L = \eta_Z \pi_Z. \tag{20}$$

This condition requires that it is equally profitable to invest money to invent labour- and Z-complementary machines, so that along the BGP N_L and N_Z can both grow. Defining $\eta \equiv \eta_Z/\eta_L$ to simplify notation and using (12) and (16), this technology market clearing condition can be solved for

$$\frac{N_Z}{N_L} = \eta^{\sigma}\left(\frac{1-\gamma}{\gamma}\right)^{\varepsilon}\left(\frac{Z}{L}\right)^{\sigma-1}, \tag{21}$$

where recall that ε is the elasticity of substitution between the goods and $\sigma \equiv \varepsilon - (\varepsilon - 1)(1-\beta)$ is the elasticity of substitution between the factors. Equation (21) shows that with the direction of technical change endogenized, the relative bias of technology, N_Z/N_L, is determined by the relative factor supply and the elasticity of substitution between the two factors.

If $\sigma < 1$, *i.e.* if the two factors are gross substitutes, an increase in Z/L will raise N_Z/N_L, hence the physical productivity of the abundant factor tends to be higher. Moreover, since $\sigma > 1$, a higher level of N_Z/N_L corresponds to Z-biased technical change. Therefore, technology will be endogenously biased in favour of the more abundant factor.

When $\sigma < 1$, *i.e.* when the two factors are gross complements, equation (21) implies that N_Z/N_L is lower when Z/L is higher. Nevertheless, because $\sigma < 1$, this lower relative physical productivity translates into a higher *value* of marginal product. Therefore, even when $\sigma < 1$, technology will be endogenously biased in favour of the more abundant factor, giving us the "weak induced-bias hypothesis".[12]

Next consider factor prices. Using equations (18) and (21), we obtain

$$\frac{w_Z}{w_L} = \eta^{\sigma-1}\left(\frac{1-\gamma}{\gamma}\right)^{\varepsilon}\left(\frac{Z}{L}\right)^{\sigma-2}. \tag{22}$$

Comparing this equation to (18), which specified the relative factor prices as a function of relative supplies and technology, we see that the response of relative factor rewards to changes in relative supply is always more elastic in (22): $\sigma - 2 > -1/\sigma$. This is not surprising given the LeChatelier principle, which states that demand curves become more elastic when other factors adjust. Here, the "other factors" correspond to the machine varieties N_L and N_Z.

The more surprising result is the "strong induced-bias hypothesis": that if σ is sufficiently large, in particular if $\sigma > 2$, the relationship between relative factor supplies and relative factor rewards can be upward sloping. Intuitively, with exogenous technology when a factor becomes more abundant, its relative reward falls: *i.e.* given N_Z/N_L, w_Z/w_L is unambiguously decreasing in Z/L due to the usual substitution effect (see equation (18)). Yet, because technology is endogenously biased towards more abundant factors, the overall effect of factor abundance on factor rewards is ambiguous.

11. This condition assumes no corner solution (where R&D would be directed only towards one of the factors), but does not rule out the economy converging to a long-run equilibrium with only one type of innovation.

12. The exception to this result is when $\sigma = 1$, *i.e.* when the production function is Cobb–Douglas. In this case, the elasticity of substitution is equal to 1, and technical change is never biased towards one of the factors (unless different types of technical change affect γ differently).

Next define the relative factor shares as

$$\frac{s_Z}{s_L} \equiv \frac{w_Z Z}{w_L L} = \eta^{\sigma-1}\left(\frac{1-\gamma}{\gamma}\right)^{\varepsilon}\left(\frac{Z}{L}\right)^{\sigma-1}, \tag{23}$$

which implies that, with endogenous technology, the share of a factor in GDP will increase in the relative abundance of that factor as long as $\sigma > 1$.

For completeness, it is also useful to determine the long-run growth rate of this economy. To do this, note that the maximization of (3) implies $g_c = \theta^{-1}(r - \rho)$, where g_c is the growth rate of consumption and recall that r is the interest rate (though there is no capital, agents are delaying consumption by investing in R&D as a function of the interest rate). In BGP, this growth rate will also be equal to the growth rate of output, g. So $r = \theta g + \rho$. Next, using (14), the free-entry condition for the technology monopolists working to invent labour-complementary machines is obtained as $\eta_L V_L = 1$. In steady state, this condition implies $\eta_L \beta p_L^{1/\beta} L/r = 1$. Now using (8), (16), and (21) to substitute for p_L, we obtain[13]

$$g = \theta^{-1}\left(\beta\left[(1-\gamma)^{\varepsilon}(\eta_Z Z)^{\sigma-1} + \gamma^{\varepsilon}(\eta_L L)^{\sigma-1}\right]^{\frac{1}{\sigma-1}} - \rho\right).$$

Finally, it is useful to briefly look at the behaviour of the economy outside the BGP. It is straightforward to verify that outside the BGP, there will only be one type of innovation.[14] If $\eta V_Z/V_L > 1$, there will only be firms creating new Z-complementary machines, and if $\eta V_Z/V_L < 1$, technology monopolists will only undertake R&D for labour-complementary machines. Moreover, V_Z/V_L is decreasing in N_Z/N_L (recall equation (17)). This implies that the transitional dynamics of the system are stable and will return to the BGP. When N_Z/N_L is higher than in (21), there will only be labour-augmenting technical change until the system returns back to balanced growth, and vice versa when N_Z/N_L is too low.

4.2. *The direction of technical change with knowledge-based R&D*

With the lab equipment model of the previous subsection, there is no state dependence. I now discuss an alternative specification which allows for potential state dependence. In the lab equipment specification there are no scarce factors that enter the accumulation equation of the economy. If there are scarce factors used for R&D, then growth cannot be maintained by increasing the amount of these factors used for R&D. So for sustained growth, these factors need to become more and more productive over time. This is the essence of the knowledge-based R&D specification, whereby spillovers imply that current researchers "stand on the shoulder of giants", ensuring that the marginal productivity of research does not decline. Here for simplicity, I assume that R&D is carried out by scientists, and there is a constant supply of scientists equal to S.[15] If there were only one sector, the knowledge-based R&D specification would require that $\dot{N}/N \propto S$ (proportional to S). With two sectors, instead, there is a variety of specifications with different degrees of state dependence, because productivity in each sector can depend on the state of knowledge in both sectors. A flexible formulation is

$$\dot{N}_L = \eta_L N_L^{(1+\delta)/2} N_Z^{(1-\delta)/2} S_L \quad \text{and} \quad \dot{N}_Z = \eta_Z N_L^{(1-\delta)/2} N_Z^{(1+\delta)/2} S_Z, \tag{24}$$

where $\delta \leq 1$ measures the degree of state dependence. When $\delta = 0$, there is no state dependence—$(\partial \dot{N}_Z/\partial S_Z)/(\partial \dot{N}_L/\partial S_L) = \eta$ irrespective of the levels of N_L and N_Z—because

13. The no-Ponzi game condition requires that $(1-\theta)g < \rho$.

14. See Proposition 1 in Acemoglu and Zilibotti (2001) for a formal proof that only one type of innovation will take place outside the BGP and for a proof of global stability in a related model. The proof here is identical.

15. The results generalize to the case where the R&D sector uses labour (or when Z is taken to be skilled labour, a combination of skilled and unskilled labour), but the analysis of the dynamics becomes substantially more complicated.

both N_L and N_Z create spillovers for current research in both sectors. In this case, the results are very similar to those in the previous subsection. In contrast, when $\delta = 1$, there is an extreme amount of state dependence: $(\partial \dot{N}_Z/\partial R_Z)/(\partial \dot{N}_L/\partial S_L) = \eta N_Z/N_L$, so an increase in N_L today makes future labour-complementary innovations cheaper, but has no effect on the cost of Z-complementary innovations.

The condition for technology market clearing in BGP now changes to

$$\eta_L N_L^{\delta} \pi_L = \eta_Z N_Z^{\delta} \pi_Z. \tag{25}$$

When $\delta = 0$, this condition is identical to (20) from the previous subsection. Solving condition (25) together with equations (12) and (16), we obtain the equilibrium relative technology as

$$\frac{N_Z}{N_L} = \eta^{\frac{\sigma}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{1-\delta\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-1}{1-\delta\sigma}}. \tag{26}$$

Now the relationship between the relative factor supplies and relative physical productivities depends on δ. This is intuitive: as long as $\delta > 0$, an increase in N_Z reduces the relative cost of Z-complementary innovations, so to restore technology market equilibrium, π_Z needs to fall relative to π_L, and N_Z/N_L needs to increase more. Substituting (26) into (18) gives

$$\frac{w_Z}{w_L} = \eta^{\frac{\sigma-1}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{(1-\delta)\varepsilon}{1-\delta\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-2+\delta}{1-\delta\sigma}}. \tag{27}$$

Relative factor shares are then obtained as

$$\frac{s_Z}{s_L} \equiv \frac{w_Z Z}{w_L L} = \eta^{\frac{\sigma-1}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{(1-\delta)\varepsilon}{1-\delta\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-1+\delta-\delta\sigma}{1-\delta\sigma}}. \tag{28}$$

It can be verified that when $\delta = 0$, when there is no state dependence in R&D, these equations are identical to their counterparts in the previous subsection.

The growth rate of this economy is determined by the number of scientists. In BGP, both sectors grow at the same rate, so we need $\dot{N}_L/N_L = \dot{N}_Z/N_Z$, or $\eta_Z N_Z^{\delta-1} S_Z = \eta_L N_L^{\delta-1} S_L$, which gives $S_L = (\eta_Z (N_Z/N_L)^{\delta-1} S)/(\eta_Z + \eta_L (N_Z/N_L)^{\delta-1})$, and $g = (\eta_L \eta_Z S)/(\eta_Z (N_Z/N_L)^{(1-\delta)/2} + \eta_L (N_Z/N_L)^{(\delta-1)/2})$ with N_Z/N_L given by (26).

State dependence in this knowledge-based R&D specification implies that the dynamics of the system can be unstable. In particular, there will now only be labour-augmenting technical change if $\eta N_Z^{\delta} V_Z/N_L^{\delta} V_L < 1$, and only Z-augmenting technical change if $\eta N_Z^{\delta} V_Z/N_L^{\delta} V_L > 1$. However, $\eta N_Z^{\delta} V_Z/N_L^{\delta} V_L$ is not necessarily decreasing in N_Z/N_L. Inspection of (17) shows that this depends on whether $\sigma < 1/\delta$ or not. When $\sigma < 1/\delta$, $\partial(N_Z^{\delta} V_Z/N_L^{\delta} V_L)/\partial(N_Z/N_L) < 0$ and transitional dynamics will take us back to the BGP. In contrast, when $\sigma > 1/\delta$, equilibrium dynamics are unstable and will take us to a corner where only one type of R&D is undertaken. Intuitively, a greater N_Z/N_L creates the usual price and market size effects, but also affects the relative costs of future R&D. If δ is sufficiently high, this latter effect becomes more powerful and creates a destabilizing influence. For example, in the extreme state dependence case where $\delta = 1$, the system is stable only when $\sigma < 1$, *i.e.* when the two factors are gross complements.

In what follows, I restrict my attention to cases where the stability condition is satisfied, so we have $\sigma < 1/\delta$. Next, note from equation (27) that the relationship between relative supplies and factor prices depends on whether

$$\sigma > 2 - \delta. \tag{29}$$

If (29) is satisfied, an increase in the relative abundance of a factor raises its relative marginal product. When $\delta = 0$, this condition is equivalent to $\sigma > 2$, which was obtained in the previous subsection. It is also clear that when $\delta = 1$, (29) implies that $\sigma > 1/\delta$, so the long-run relative demand curve cannot be upward sloping in a stable equilibrium. However, for all values of δ less than 1, the conditions $\sigma < 1/\delta$ and $\sigma > 2 - \delta$ can be satisfied simultaneously. So the general conclusions reached in the previous subsection continue to apply as long as $\delta < 1$. In fact, some degree of state dependence makes it more likely that long-run demand curves slope up.

The case of extreme state dependence, $\delta = 1$, on the other hand, leads to an interesting special result. In this case, to have innovations for both sectors, we need equation (28) to be satisfied, which implies $s_Z/s_L \equiv w_Z Z/w_L L = \eta^{-1}$ (and in addition, we need $\sigma < 1$ for stability). Technical change is therefore acting to equate relative factor shares, as conjectured by Kennedy (1964), who suggested that induced innovations will push the economy towards constant factor shares (see also Samuelson (1965), Drandakis and Phelps (1965)). Here we find that this result is obtained when the innovation possibilities frontier takes the form given in (24) with an extreme amount of state dependence, $\delta = 1$.

4.3. *Discussion*

The analysis so far has highlighted two important determinants of the direction of technical change. The first is the degree of substitution between the two factors. When the two factors are more substitutable, the market size effect is stronger, and endogenous technical change is more likely to favour the more abundant factor. The second determinant of the direction of technical change is the degree of state dependence in the innovation possibilities frontier. To clarify the main issues, I kept the analysis as simple as possible. I now briefly discuss a number of possible generalizations, and some issues that arise in using this model to think about the real world.

The first issue to note is that the model here exhibits a "scale effect" in the sense that as population increases, the growth rate of the economy also increases. Jones (1995) convincingly shows that there is little support for such a scale effect, and instead proposes a model where steady growth in income *per capita* is driven by population growth. Since the scale effect is related to the market size effect I emphasize here, one might wonder whether, once we remove the scale effect, the market size effect will also disappear. The answer is that the scale effect and the market size effect emphasized here are distinct, and a very natural formulation removes the scale effect while maintaining the market size effect. Briefly, consider the following form of the innovation possibilities frontier: $\dot{N}_L = \eta_L N_L^{\lambda} S_L$ and $\dot{N}_L = \eta_Z N_Z^{\lambda} S_Z$, where $\lambda \in (0, 1]$. As long as $\lambda < 1$, as in Jones (1995), there is no scale effect, and the long-run rate of growth of the economy is $n/(1 - \lambda)$, where n is the rate of population growth. However, all the results on the direction of technical change obtained above continue to apply, with the only difference that in all the equations from Section 4.2, λ replaces δ.

Second, the analysis here emphasizes profit-motivated R&D. In practice, there are important advances, perhaps including some of the most major scientific discoveries, that are not purely driven by profit motives. Introducing such non-profit-motivated innovations into this framework is straightforward. One possibility would be that non-profit motives drive what Mokyr (1990) refers to as "macro-innovations", or what Bresnahan and Trajtenberg (1995) call "general-purpose technologies", while the profit motives determine how these macro-innovations are developed for commercial use. Via this channel, it will be the profit motives that shape the direction of technical change. Another possibility is to assume that as well as profit-driven innovations, there is some "technology drift". For example, the innovation possibilities frontier could take the form $\dot{N}_L = \xi_L N_L + \eta_L R_L$ and $\dot{N}_Z = \xi_Z N_Z + \eta_Z R_Z$, where the ξ_L and ξ_Z terms capture the technology drift, that is, improvements in technology that are unrelated to R&D directed at different types of innovation. The presence of the ξ terms does not change the marginal return to different types of innovation, and for an equilibrium in which there are both types of R&D, we still need the innovation equilibrium condition, (20), to hold.

Third, the analysis so far treated factor supplies, L and Z, as given. Clearly, these supplies can, and typically will, respond to relative prices. I briefly discuss the case where Z corresponds to capital and accumulates in response to the interest rate in Section 6.1. When Z corresponds to skilled labour, the relative price of Z, w_Z/w_L, will determine the relative supply, Z/L. I analysed this case in Acemoglu (1998).[16] Here, it suffices to note that standard arguments imply an upward sloping relative supply of skills, *e.g.* $Z/L = f(w_Z/w_L)$, so when the relative demand for skills is upward sloping, an exogenous increase in Z/L will increase w_Z/w_L, encouraging further accumulation of skills, and yet more skill-biased technical change. Note also that when the demand for skills is sufficiently upward sloping, the relative supply and the relative demand curves could intersect more than once, and multiple long-run equilibria are possible (see Acemoglu (1998)).

Finally, it is useful to briefly discuss empirical evidence on the degree of state dependence. To the best of my knowledge, there has been no direct investigation of this issue. The data on patent citations analysed by, among others, Jaffe, Trajtenberg and Henderson (1993), Trajtenberg, Henderson and Jaffe (1992) and Caballero and Jaffe (1993) may be useful in this regard. These papers study the extent of subsequent citations of patents and interpret a citation of a previous patent as evidence that a current invention is exploiting information generated by this previous invention. This corresponds to the presence of N_L and N_Z terms in equation (24) in the model here. Patent citations data can be used to investigate whether there is state dependence at the industry level. Unfortunately, it is currently impossible to investigate state dependence at the factor level, which is more relevant for the discussion here. This is because, although we have information about the industry for which the patent was developed, we do not know which factor the innovation was directed at. In any case, the results reported in Table 1 of Trajtenberg *et al.* (1992) suggest that there is some limited amount of industry state dependence. For example, patents are likely to be cited in the same three-digit industry from which they originated; but on average, they are more often cited in a different three-digit industry (see below).

5. APPLICATIONS (WITH LIMITED STATE DEPENDENCE)

In this section, I discuss some of the applications of directed technical change. I will emphasize both why models with endogenously biased technical change are useful in thinking about a range of problems and how the results depend on the elasticity of substitution. I start with a range of applications where a formulation with limited state dependence, *i.e.* the case with $\delta < 1$, appears to be more appropriate.

5.1. *Endogenous skill-biased technical change*

Figure 1 plots a measure of the relative supply of skills (the number of college equivalent workers divided by noncollege equivalents) and a measure of the return to skills (the college premium).[17] It shows that over the past 60 years, the U.S. relative supply of skills has increased rapidly, and in the meantime, the college premium has also increased. The figure also shows that beginning in the late 1960's, the relative supply of skills increased more rapidly than before. In response,

16. See Heckman, Lochner and Taber (1998) for a more detailed analysis of the response of the supply of skills to an increase in the returns to skills.

17. The samples are constructed as in Katz and Autor (2000). See Katz and Autor (2000) or Acemoglu (2002) for details.

the skill premium fell during the 1970's, and then increased sharply during the 1980's and the early 1990's.

Existing explanations for these facts emphasize exogenous skill-biased technical change.[18] Technical change is assumed to naturally favour skilled workers, perhaps because new tasks are more complex and generate a greater demand for skills. This skill bias explains the secular behaviour of the relative supply and returns to skills. Moreover, the conventional wisdom is that there has been an acceleration in the skill bias of technology precisely around the same time as the relative supply of skills started increasing much more rapidly. This acceleration explains the increase in the skill premium and wage inequality during the 1980's.

A model based on directed technical change suggests an alternative explanation. Suppose that the second factor in the model above, Z, is skilled labour. Also assume that the innovation possibilities frontier is given by (24), so that we are in the knowledge-based R&D model. This is without any loss of generality since for the focus here, the lab equipment specification corresponds to the case with $\delta = 0$ in terms of (24). Then, equation (26) from Section 4 gives the equilibrium skill bias and equation (27) gives the long-run skill premium.

The results above, in particular equation (26) and the weak induced-bias hypothesis, imply that the increase in the supply of skills creates a tendency for new technologies to be skill biased. This offers a possible explanation for the secular skill-biased technical change of the twentieth century irrespective of the value of the elasticity of substitution, σ (as long as it is not equal to 1), since the predictions regarding induced skill-biased technical change do not depend on whether the long-run relative demand curve for skill is upward sloping.

Equally interesting, if $\sigma > 2 - \delta$, then we have the strong induced-bias hypothesis: now the long-run relationship between the relative supply of skills and the skill premium is positive, and greater skill abundance leads to a greater skill premium. Intuitively, with σ large enough, the market size effect is so powerful that it not only dominates the price effect on the direction of technical change, but also dominates the usual substitution effect between skilled and unskilled workers at a given technology (as captured by equation (18) above). As a result, an increase in the relative supply of skills makes technology sufficiently skill biased to raise the skill premium. In this case, the framework predicts sufficiently biased technical change over the past 60 years to actually increase the skill premium, consistent with the pattern depicted in Figure 1.

The strong induced-bias hypothesis, *i.e.* the case with $\sigma > 2 - \delta$, also offers an explanation for the behaviour of the college premium during the 1970's and the 1980's shown in Figure 1. Suppose that the economy is hit by a large increase in H/L. If this increase is not anticipated sufficiently in advance, technology will not have adjusted by the time the supply of skills increases. The initial response of the skill premium will be given by equation (18) which takes technology as given. Therefore, the skill premium will first decline, but then as technology starts adjusting, it will rise rapidly. Figure 3 draws this case diagramatically.[19]

This framework also provides a possible explanation for why technical change during the late eighteenth and early nineteenth centuries may have been biased towards unskilled labour. The emergence of the most "skill-replacing" technologies of the past 200 years, the factory

18. For example, Autor *et al.* (1998), Galor and Maov (2000) or Krusell, Ohanian, Rios-Rull and Violante (2000).

19. There was also a very large increase in educational attainment in the U.S. during the first half of the century. High school enrolment and graduation rates doubled in the 1910's (*e.g.* Goldin and Katz (1995)). The skill premium fell sharply in the 1910's. But, despite the even faster increase in the supply of high school skills during the 1920's, the skill premium levelled off and started a mild increase. Goldin and Katz (1995) conclude that the demand for high school graduates must have expanded sharply starting in the 1920's, presumably due to changes in office technology and higher demand from new industries such as electrical machinery, transport and chemicals (see also Goldin and Katz (1998)). Therefore, this historical episode also suggests that the demand for skills expanded sharply in response to the increase in the supply of skills, giving empirical support to the weak induced-bias hypothesis. However, it does not support the strong induced-bias hypothesis.

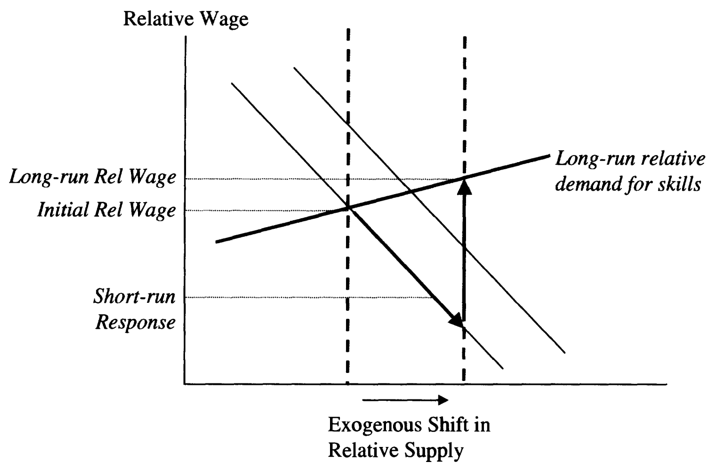


FIGURE 3

Short-run and long-run responses of the skill premium to an increase in the relative supply of skills when condition (29) is satisfied

system, coincided with a substantial change in relative supplies. This time, there was a large migration of unskilled workers from villages and Ireland to English cities and a large increase in population (see, for example, Habakkuk (1962), Bairoch (1988) or Williamson (1990)). This increase created profit opportunities for firms in introducing technologies that could be used with unskilled workers. In fact, contemporary historians considered the incentive to replace skilled artisans by unskilled labourers as a major objective of technological improvements of the period. Ure, a historian in the first half of the nineteenth century, describes these incentives as follows: "It is, in fact, the constant aim and tendency of every improvement in machinery to supersede human labour altogether, or to diminish its costs, by substituting the industry of women and children for that of men; of that of ordinary labourers, for trained artisans." (quoted in Habakkuk (1962, p. 154)). The framework developed here is consistent with this notion that the incentives for skill-replacing technologies were shaped by the large increase in the supply of unskilled workers.

Therefore, this model provides an attractive explanation for the skill-replacing technical change of the nineteenth century, the secular skill bias of technology throughout the twentieth century, and the recent acceleration in this skill bias. In addition if $\sigma > 2 - \delta$, it explains both the increase in the skill premium over the past 60 years and its surge during the past 25 years without invoking any exogenous change in technology.

What is the empirically plausible range for $\sigma - 2 + \delta$? First note that Figure 1 suggests the demand for skills over the past 50 years has increased somewhat faster than the supply of skills. In the context of the model here, this requires $\sigma - 2 + \delta$ and $(\sigma - 2 + \delta)/(1 - \delta\sigma)$ to be positive, but not too large. The elasticity of substitution, σ, is generally difficult to estimate, but there is a relatively widespread consensus that the elasticity between skilled and unskilled workers is greater than 1, most likely, greater than 1.4, and perhaps as large as 2 (see, for example, Freeman (1986)). An interesting study by Angrist (1995) uses a "natural experiment" arising from the large increase in university enrolments for Palestinian labour. His estimates imply an elasticity of substitution between workers with 16 years of schooling and those with less than 12 of schooling of approximately $\sigma = 2$.

What about δ? Unfortunately, the magnitude of this parameter is much harder to ascertain. The best we can do is to make a very rough guess based on the evidence presented by Trajtenberg *et al.* (1992). Looking at patent citations, they construct an index of whether a patent from a given industry is citing patents from the same industry. Their index, TECHF, takes the value 0 when the citing and originating patents are in the same three-digit industry, the value 0·33 when they are in the same two-digit industry, the value 0·66 when they are in the same one-digit industry, and the value 1 if they are not even in the same one-digit industry. The average of TECHF in their sample is approximately 0·31, implying that there is some limited degree of state dependence at the industry level. On the basis of this, a plausible value for δ would be less than 0·31, perhaps 0·2 (though, of course, it is possible that state dependence at the factor level is greater than that at the industry level). Combining $\delta = 0{\cdot}2$ with $\sigma = 1{\cdot}4$, we obtain $(\sigma - 2 + \delta)/(1 - \delta\sigma) = -0{\cdot}55$, while combining it with $\sigma = 2$, we obtain $(\sigma - 2 + \delta)/(1 - \delta\sigma) = 0{\cdot}33$. So with relatively limited state dependence, the plausible range of the long-run elasticity of the skill premium with respect to the relative supply of skills would lie between $-0{\cdot}55$ and $0{\cdot}33$, compared to the range of the short-run elasticity of $[-0{\cdot}7, -0{\cdot}5]$. As we consider greater values of δ, the range of the long-run elasticity shifts to the right.[20]

5.2. *Directed technical change and cross-country income differences*

Many less developed countries (LDCs) use technologies developed in the U.S. and other OECD economies (the North). A number of economists, including Atkinson and Stiglitz (1969), David (1975), Stewart (1977), Basu and Weil (1998), have pointed out that imported technologies may not be "appropriate" to LDCs' needs. Directed technical change increases these concerns: it implies that technologies will be designed to make optimal use of the conditions and factor supplies in the North. Therefore, they will be highly inappropriate to the LDCs' needs (Acemoglu and Zilibotti, 2001). Because it is still often profitable for the LDCs to use these technologies rather than develop their own, the extent of directed technical change will determine how inappropriate technologies used by the LDCs' are to their needs. Via this channel, directed technical change will influence the income gap between the North and the LDCs. I now use the above framework to discuss this issue.

Suppose that the model outlined above applies to a country I refer to as "the North" (either the U.S. or all the OECD countries as a whole). Also to simplify the algebra, I now focus on the case with no state dependence $\delta = 0$, though all the results generalize to the case of limited state dependence immediately. Suppose also that there is a set of LDCs in this world economy that will use the technologies developed in the North. Take Z to be skilled labour, though for the results in this subsection, it could also stand for physical capital. For simplicity, I take all of these countries to be identical, with L' unskilled workers and H' skilled workers.

A key characteristic of the LDCs is that they are less abundant in skilled workers than the North, so $H'/L' < H/L$. I assume that all LDCs can copy new machine varieties invented in the North, but firms face a price of $\kappa^{-\beta/(1-\beta)}$ rather than 1 as in the North. This cost differential may result from the fact that firms in the LDCs do not have access to the same knowledge base as the technology monopolists in the North. Also, there is no international trade between the North and the LDCs.

An analysis similar to above immediately gives the LDC output levels (recall equation (15)): $Y'_L = (p'_L)^{(1-\beta)/\beta}\kappa N_L L'/(1-\beta)$ and $Y'_H = (p'_H)^{(1-\beta)/\beta}\kappa N_H H'/(1-\beta)$, where p''s denote prices in the LDCs, which differ from those in the North because factor proportions are different and there is no international trade. The parameter κ features in these equations since machine

20. Finally, to see another reason why for this application the model with of limited state dependence is more relevant, recall from equation (28) that if $\delta = 1$, in the long run we would have $s_H/s_L = \eta^{-1}$. Therefore, the relative shares of skilled and unskilled labour in GDP whould be constant. This is clearly not consistent with the pattern depicted in Figure 1, where both the number of skilled workers and their relative renumeration have been increasing over time.

costs are different in the LDCs. It is natural to think of κ as less than 1, so that machine prices are higher and fewer machines are used in the LDCs than in the North. Notice also that the technology terms, N_L and N_H, are the same as in the North, since these technologies are copied from the North. Using this equation and (4) and (15), the ratio of aggregate income in the LDCs to that in the North can be written as

$$\frac{Y'}{Y} = \frac{\kappa\left[\gamma\left((p_L')^{(1-\beta)/\beta}N_L L'\right)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma)\left((p_H')^{(1-\beta)/\beta}N_H H'\right)^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}}{\left[\gamma\left(p_L^{(1-\beta)/\beta}N_L L\right)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma)\left(p_H^{(1-\beta)/\beta}N_H H\right)^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}}. \tag{30}$$

Some tedious algebra (see Acemoglu (2001)) shows that

$$\frac{\partial Y'/Y}{\partial N_H/N_L} \propto \frac{1-\beta}{\sigma}\left(\frac{N_H}{N_L}\right)^{-\frac{1}{\sigma}}\left(\left(\frac{H'}{L'}\right)^{\frac{\sigma-1}{\sigma}} - \left(\frac{H}{L}\right)^{\frac{\sigma-1}{\sigma}}\right). \tag{31}$$

Since $H'/L' < H/L$, this expression implies that when $\sigma > 1$, *i.e.* when the two factors are gross substitutes, an increase in N_H/N_L raises the income gap between the LDCs and the North (*i.e.* reduces Y'/Y). In contrast, when $\sigma < 1$, so that the two factors are gross complements, an increase in N_H/N_L narrows the income gap. The intuition for both results is straightforward. An increase in N_H/N_L increases the productivity of skilled workers relative to the productivity of the unskilled. Therefore, everything else equal, a society with more skilled workers benefits more from this type of technical change. However, this change also affects the relative scarcity, and therefore the relative price, of the two goods. In particular, the skill-intensive good becomes more abundant and its relative price falls. When $\sigma > 1$, this effect is weak, and the North still gains more in proportional terms, and the ratio of LDC income to income in the North falls. However, when $\sigma < 1$, the price effect is strong, and as a result, in proportional terms, the skill-abundant North benefits less than the LDCs.

Next, recall that when $\sigma > 1$, an increase in N_H/N_L corresponds to skill-biased technical change, while with $\sigma < 1$, it is a decrease in N_H/N_L that is skill biased. Therefore, irrespective of the value of σ, skill-biased technical change increases the income gap between the LDCs and the North. This extends the results in Acemoglu and Zilibotti (2001) to a slightly more general model, and also more importantly, to the case where the two factors are gross complements.

Now to highlight the role of directed technical change, contrast two polar opposite situations. First, suppose that there are no intellectual property rights enforced in the LDCs, so LDC firms pay no royalties to R&D firms in the North. This implies that new technologies are developed for the Northern market, equation (21) applies and $N_H/N_L \propto (H/L)^{\frac{\sigma-1}{\sigma}}$. Contrast this situation with an alternative where new technologies can also be sold in the LDC markets on exactly the same terms as in the North. In this case the equilibrium technologies would satisfy $N_H/N_L \propto (H^W/L^W)^{\frac{\sigma-1}{\sigma}}$ where $H^W/L^W = (H+\kappa H')/(L+\kappa L')$ is the world (effective) ratio of skilled to unskilled workers. The parameter κ features in this equation because it parameterizes the relative productivity/demand for machines of LDC workers.

By the assumption that $H'/L' < H/L$, we have $H^W/L^W < H/L$. This implies that, as a direct implication of the weak induced-bias hypothesis, equilibrium technologies will be too skill biased for the LDCs. In particular, when $\sigma > 1$, technologies developed in the North will feature too high a level of N_H/N_L for the LDCs' needs, and when $\sigma < 1$, technologies developed in the North will feature too low a level of N_H/N_L for the LDCs (which again corresponds to more skill-biased technology). Hence, irrespective of the value of the elasticity of substitution, directed technical change, by making technologies in the North too skill biased for the LDCs, creates a force towards a larger income gap between the rich and the poor. This result is intuitive: there are more skilled workers in the North, and directed technical change induces technology monopolists in the North to develop technologies appropriate for skilled workers (*i.e.* when $\sigma > 1$, higher N_H/N_L, and when $\sigma < 1$, lower N_H/N_L). These skill-biased technologies are less useful for LDCs, so LDCs benefit less than the North, and the income gap is larger than it would have been in the absence of directed technical change.[21]

5.3. *The effect of international trade on technical change*

The literature on trade and growth shows how patterns of international trade affect the rate of technical change (*e.g.* Grossman and Helpman (1991)). Similarly, when the direction of technical change is endogenous, changes in the amount of international trade may affect the type of technologies that are developed. This may be relevant for current economic concerns, for example because, as claimed by Wood (1994), trade opening may have affected wage inequality through its effect on "defensive" innovations.[22]

To address these issues, consider the same set-up as in the previous subsection with a set of LDCs that can copy innovations from the North and let us again focus on the case of no state dependence (*i.e.* $\delta = 0$ or the lab equipment specification). Assume first that there is no international trade, so the equilibrium characterized in the previous subsection applies. In particular, the relative price of skill-intensive goods in the North is given by equation (16), and the skill premium is given by an equation similar to (18). When skill bias of technology is endogenized, N_H/N_L is given by (21).

Next, suppose that there is opening to international trade, with both goods traded costlessly. Assume that the structure of intellectual property rights is unchanged as a result of this trade opening. International trade will generate a single world relative price of skill-intensive goods, p^W. To determine this price, note that the total supply of skill-intensive goods will be $(p_H^W)^{(1-\beta)/\beta} N_H^W (H + \kappa H')/(1-\beta)$, and the total supply of labour-intensive goods will be $(p_L^W)^{(1-\beta)/\beta} N_L^W (L + \kappa L')/(1-\beta)$. Using these expressions and equation (7), the world relative price is obtained as

$$p^W = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\beta\varepsilon}{\sigma}} \left(\frac{N_H^W (H+\kappa H')}{N_L^W (L+\kappa L')}\right)^{-\frac{\beta}{\sigma}} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\beta\varepsilon}{\sigma}} \left(\lambda^{-1}\frac{N_H^W H}{N_L^W L}\right)^{-\frac{\beta}{\sigma}}, \tag{32}$$

where the last equality defines $\lambda \equiv (H/L)/((H+\kappa H')/(L+\kappa L')) > 1$. The fact that $\lambda > 1$ follows from $H'/L' < H/L$. I also use the notation N_H^W and N_L^W to emphasize that world technologies may change from their pre-trade levels in the North as a result of international trade.

Since skills are scarcer in the world economy than in the North alone, trade opening will increase the relative price of skill-intensive goods in the North, *i.e.* $p^W > p$. This is a straightforward application of standard trade theory. What is different here is that this change in product prices will also affect the direction of technical change. Recall that the two forces shaping the direction of technical change are the market size effect and the price effect. Because trade does not affect the structure of intellectual property rights, the market sizes for different types of technologies do not change. But product prices are affected by trade, so the price effect will be operational. Since the price effect encourages innovations for the scarce factor, international

21. Reality presumably lies somewhere in between, with imperfect intellectual property rights enforcement in the LDCs. It is straightforward to generalize the results to that case, and show that technologies in that case will also be too skill biased for the LDCs compared to the case of full property rights enforcement (see Acemoglu (1999*a*)).

22. This point is also suggested in Acemoglu (1998) and analysed in detail in Acemoglu (1999*a*). Xu (2001) extends the set-up of Acemoglu (1999*a*), which I use here, to have both goods employ both factors.

trade, by making skills more scarce in the North, will induce more innovations directed at skilled workers.

To see this more formally, once again consider the case of no state dependence in R&D (*i.e.* the lab equipment specification, or (24) with $\delta = 0$). For the technology market to clear, we need condition (20) to be satisfied. Combining the relative price of skill-intensive goods, now given by (32), with technology market clearing condition (20), we obtain

$$\frac{N_H^W}{N_L^W} = \eta^\sigma \lambda \left(\frac{1-\gamma}{\gamma} \right)^\varepsilon \left(\frac{H}{L} \right)^{\sigma-1}. \tag{33}$$

Comparing this equation to (21), we see immediately that, because $\lambda > 1$, trade increases the physical productivity of skilled workers more than that of unskilled workers. As usual, this increase in N_H/N_L may not correspond to skill-biased technical change if $\sigma < 1$.

To study the effect of this induced change in technology on factor prices, first note that (32) implies that without a change in technology trade opening would increase the skill premium in the North to

$$\frac{w_H}{w_L} = \left(p^W\right)^{1/\beta} \frac{N_H}{N_L} = \left(\frac{1-\gamma}{\gamma} \right)^{\frac{\varepsilon}{\sigma}} \lambda^{\frac{1}{\sigma}} \left(\frac{N_H}{N_L} \right)^{\frac{\sigma-1}{\sigma}} \left(\frac{H}{L} \right)^{-\frac{1}{\sigma}}. \tag{34}$$

Comparing this equation to (18) from before, we see that trade opening necessarily increases the skill premium, which is a simple application of standard trade theory.

With directed technical change, the skill premium in the North, instead, changes to

$$\frac{w_H}{w_L} = \left(p^W\right)^{1/\beta} \frac{N_H^W}{N_L^W} = \eta^{\sigma-1} \lambda \left(\frac{1-\gamma}{\gamma} \right)^\varepsilon \left(\frac{H}{L} \right)^{\sigma-2}. \tag{35}$$

This equation differs from (34) because the skill bias of technology is now given by (33) rather than being held constant at the pre-trade level.

Comparing the post-trade skill premium, (35), to the skill premium before trade opening (*e.g.* equation (22)), we see that irrespective of whether $\sigma > 1$ or not, trade opening increases the skill premium. However, comparing equations (34) and (35) shows that the trade-induced technical change will increase the skill premium by more than predicted by the standard trade theory only when $\sigma > 1$. The intuition for this result is also simple: expression (33) shows that trade will induce technical change to increase the relative physical productivity of skilled workers. An increase in relative physical productivity translates into skill-biased technical change only when $\sigma > 1$. Therefore, we obtain the result that, as conjectured by Adrian Wood, trade opening could induce skill-biased technical change and increase wage inequality more than predicted by standard trade theory. Yet, this conclusion is obtained only when skilled and unskilled workers are gross substitutes (which appears to be the empirically relevant case).

5.4. *The Habakkuk hypothesis*

According to the Habakkuk (1962) hypothesis, the rapid technical change or technology adoption in the U.S. economy during the ninteenth century (especially relative to the British economy) resulted from the relative labour scarcity in the U.S., because labour scarcity increased wages, and encouraged firms to develop and adopt labour-saving technologies. Despite the prominence of this hypothesis in the economic history literature, there has been no widely accepted formalization of the argument, and we are consequently unaware of under what circumstances Habakkuk's conclusion would apply (see, *e.g.* David (1975)). In addition, standard economic theory suggests that higher wages may reduce investment, and via this channel, discourage innovation. Moreover, in contrast to the basic premise of the Habakkuk hypothesis, the endogenous growth literature emphasizes the presence of a scale effect which suggests that a larger workforce should encourage more innovation. Could the Habakkuk hypothesis be valid despite these countervailing effects?

To discuss this case, suppose that Z now stands for land. Moreover, assume that $\eta_Z = 0$, which implies that there are no land-complementary innovations. The only source of technical change is labour-complementary innovations. The rest of the set-up is the same as before, and to economize on space, I will only study the lab equipment specification. The question is: under what circumstances will greater labour scarcity (smaller L/Z) lead to a higher level of N_L—*i.e.* to more innovations?

From (14), the free-entry condition for technology monopolists, $\eta_L V_L = 1$, implies $\eta_L \beta p_L^{1/\beta} L/r = 1$. Now using (8) and (16), this condition gives

$$\eta_L \beta \left[\Lambda \left(\frac{N_L}{Z} \right)^{\frac{1-\sigma}{\sigma}} L^{-\frac{(1-\sigma)^2}{\sigma}} + \gamma^{\varepsilon} L^{\sigma-1} \right]^{\frac{1}{\sigma-1}} = r, \tag{36}$$

where Λ is a suitably defined constant. Inspection of equation (36) immediately shows that $\partial N_L/\partial Z > 0$. Therefore, a greater abundance of land (for a given level of employment) always encourages the creation of more labour-complementary technologies. So comparing the U.S. to Britain, as Habakkuk did, leads to the conclusion that there should be faster technical change in the more land-abundant U.S.

On the other hand, the sign of $\partial N_L/\partial L$—*i.e.* the effect of labour scarcity for a given supply of land—is ambiguous. If $\sigma > 1$, so that labour and land are gross substitutes, it can be verified that $\partial N_L/\partial L > 0$, hence in this case, greater scarcity of labour (for a given supply of land) discourages the development of new technologies. In contrast, if σ is sufficiently smaller than 1, *i.e.* if labour and land are sufficiently complementary, we can have $\partial N_L/\partial L < 0$. This is intuitive. A greater scarcity of labour creates two forces: the price of the labour-intensive good is higher, but also the market size for labour-complementary technology is smaller. Which force dominates depends on the strength of the price effect, which is again a function of the degree of substitutability. If $\sigma > 1$, the price effect is less powerful and we obtain the opposite of the result conjectured by Habakkuk—it is not labour scarcity, but labour abundance that spurs innovation (for a given quantity of land). However, with sufficient complementarity between labour and land, the model gives the result conjectured by Habakkuk—greater labour scarcity leads to faster innovation. Therefore, in this framework the Habakkuk hypothesis requires labour and land to be highly complementary. Since the focus of Habakkuk's analysis was technical change in agriculture and low-tech manufacturing, the assumption of a high degree of complementarity between labour and other inputs may be realistic.

6. TECHNICAL CHANGE WITH EXTREME STATE DEPENDENCE

I have so far discussed a number of applications of directed technical change with limited state dependence. In this section, I discuss two applications where extreme state dependence, $\delta = 1$, may be more appropriate.

6.1. *Why is long-run technical change labour augmenting?*

Consider another regularity discussed in the introduction: the share of labour and capital have been approximately constant in U.S. GDP, while the capital–labour ratio has been increasing steadily. This suggests that technical change has been mostly labour augmenting (unless the elasticity of substitution between capital and labour happens to be exactly equal to 1). Can a model of directed technical change be useful in thinking about why aggregate technical change appears to be labour augmenting?

To provide an answer to this question, suppose that Z corresponds to physical capital.[23] Assume also that $Z = K$ is growing over time due to capital accumulation. Let us start with the flexible innovation possibilities frontier given by (24), where δ parametrizes the degree of state dependence. Then equation (28), with K replacing Z, implies that unless $\sigma = 1$ (the elasticity of substitution exactly equal to 1), or $\delta = 1$ (extreme state dependence), factor shares will not be constant: with the capital–labour ratio growing, the share of capital will be contracting or expanding, and there exists no BGP. This implies that neither the lab equipment specification nor the knowledge-based R&D specification with limited state dependence are consistent with a constant capital share or with purely labour-augmenting technical change.

For all technical change to be (endogenously) labour augmenting, we therefore need $\delta = 1$. In Acemoglu (1999*b*), I show that when the innovation possibilities frontier takes the form in (24) with $\delta = 1$ and the elasticity of substitution is less than 1, there exists a unique equilibrium path tending to a BGP with only labour-augmenting technical change. This result can be interpreted as either a positive or negative one: on the positive side, it shows that it is possible to construct a model where equilibrium long-run technical change is labour augmenting, even though capital-augmenting technical change is also allowed. Moreover, the economy converges to this equilibrium, and on the transition path, there will typically be capital-augmenting technical change. On the negative side, it shows that this result obtains only when there is an extreme amount of state dependence in R&D, *i.e.* $\delta = 1$,[24] and when the elasticity of substitution between labour and capital is less than 1.[25]

6.2. *Wage push and technical change*

Finally, I use the above framework to investigate the technological implications of a "wage-push shock", which Blanchard (1997) and Caballero and Hammour (1998) argue took place in continental Europe starting in the late 1960's. Consider the above framework with Z interpreted as capital, K, and as in the previous subsection, focus on the case with the elasticity of substitution less than 1, *i.e.* $\sigma < 1$. In addition, in order to study the implications of wage push, let me introduce a quasi-labour supply function: $L = m(s_L)$, where s_L is the relative labour share, and I assume $m' > 0$, so that a higher labour share leads to greater labour supply.[26] To simplify the algebra, I further specialize this supply function, and assume that

$$L = \left(\mu \frac{s_L}{1 - s_L}\right)^{1/\tau} = \left(\mu \frac{s_L}{s_K}\right)^{1/\tau}. \tag{37}$$

Here $\tau > 0$, and μ is a shift parameter. A decrease in μ corresponds to a "wage-push" shock, since it increases the labour share at a given level of employment. A high level of τ corresponds

23. One has to be careful with this interpretation since the model already features "machines" which can also be thought as capital.

24. This leads to the question of how this result, which requires extreme state dependence, $\delta = 1$, can be reconciled with the findings in the previous subsection, which relied on limited state dependence, *i.e.* $\delta < 1$. One possibility is to develop a hybrid framework where the degree of state dependence varies across innovations directed at different factors (see Acemoglu (2001)).

25. An elasticity of substitution less than 1 between capital and labour appears reasonable. Using time-series data, Coen (1969), Eisner and Nadiri (1968) and Lucas (1969), for example, all estimate elasticities significantly less than 1. Although Berndt (1976) claims that the use of higher quality data leads to higher estimates of the elasticity of substitution, he does not control for a time trend in his estimation, biasing his results towards 1.

26. Note that this formulation captures both the labour supply decisions of workers and bargaining between firms and workers. For example, with a bargaining set-up, when unemployment is high (employment is low), workers will have a weaker bargaining position, and only receive a smaller share of output, leading to lower s_L.

to a more inelastic quasi-supply curve. This formulation is convenient since it yields a simple expression for s_L/s_K, for given bias of technology, from (18) to be combined with (37):

$$\frac{s_L}{s_K} = \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{\varepsilon}{\sigma}} \left(\frac{N_K}{N_L}\frac{K}{L}\right)^{\frac{1-\sigma}{\sigma}}. \tag{38}$$

In essence (38) shows that, when labour and K are gross complements, *i.e.* $\sigma < 1$, as we have assumed to be the case, an increase in the relative productivity, N_K/N_L, or relative abundance, K/L, of the other factor will raise the labour share. So with $\sigma < 1$, and wage-push shock will increase the labour share and reduce K/L. With $\sigma > 1$, we would obtain the opposite result.

What will happen in the long run? There are two margins of adjustment: the capital–labour ratio and technology. Even without any adjustment in technology, the labour share may return back to its initial level if the capital–labour ratio returns to its initial position. This will be the case, for example, when the long-run return to capital is given by a perfectly elastic supply of capital (*e.g.* CRRA preferences for the representative consumer). However, Blanchard (1997) shows that changes in capital–labour ratio do not account for the behaviour of the labour share. This raises the possibility that the major adjustment was due to changes in technology. To focus on this, I ignore capital accumulation and normalize $K = 1$ (see Acemoglu (1999*b*), for the analysis of the case where both technology and capital adjust).

Now combining (37) and (38), we obtain

$$L^{SR} = \mu^{\frac{\sigma}{\tau\sigma+1-\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{-\frac{\varepsilon}{\tau\sigma+1-\sigma}} \left(\frac{N_K}{N_L}\right)^{\frac{1-\sigma}{\tau\sigma+1-\sigma}}. \tag{39}$$

Since $\sigma < 1$, (39) defines a positive relationship between μ and L. I use the superscript "SR" to emphasize that this is the short-run response of employment for given bias of technology. From equation (39), a wage-push shock, *i.e.* a decrease in μ, will reduce employment. Equation (38), in turn, implies that as long as $\sigma < 1$, this shock will also increase the labour share in GDP. Therefore, for a given bias in technology, as argued by Blanchard and Caballero and Hammour, a wage push will reduce employment and increase the labour share.

The fundamental implication of the framework here is that the wage push will also affect the equilibrium bias of technical change. As in the previous subsection, consider the case with extreme state dependence, $\delta = 1$, and combine (37) with equation (28) from Section 4.2 with $Z = K = 1$ to obtain the long-run relative labour share as $(s_L/s_K)^{LR} = \eta^{-1}$. This equation states that the technology-market clearing condition can only be satisfied if the labour share in GDP returns back to its initial level. Using this, we also calculate the long-run employment level as

$$L^{LR} = \eta^{-\frac{1}{\tau}} \mu^{\frac{1}{\tau}}, \tag{40}$$

where the superscript "LR" shows that these expressions refer to the long-run equilibrium. Next, using equation (26), we obtain the relative technology level consistent with long-run equilibrium as

$$\frac{N_K}{N_L} = \eta^{\frac{\sigma\tau-(1-\sigma)}{\tau(1-\sigma)}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\sigma}} \mu^{\frac{1}{\tau}}. \tag{41}$$

Comparison of (39) and (40) immediately implies that the elasticity of employment with respect to μ is greater in the long run than in the short run ($\sigma/(\tau\sigma + 1 - \sigma) < 1/\tau$).[27] This again can be thought of as an application of the LeChatelier principle. Intuitively, with $\sigma < 1$, despite the

27. When $\sigma > 1$, equation (40) no longer gives the long-run labour demand, since, in this case, the economy is unstable and will tend to an equilibrium with only one type of innovation.

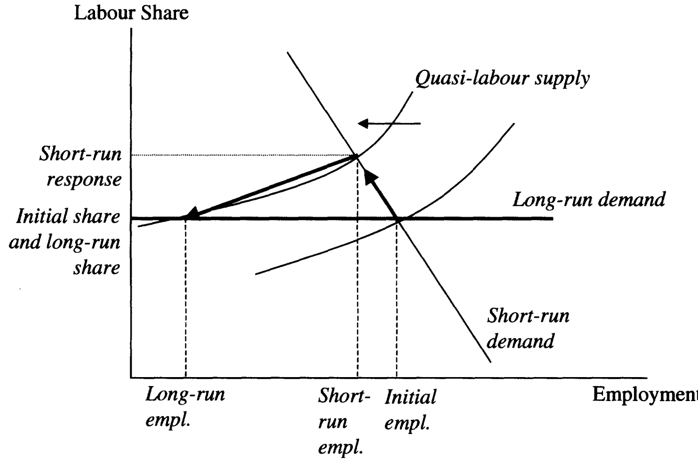


FIGURE 4

Short-run and long-run responses of the labour share to a shift in the quasi-labour supply when capital and labour are gross complements

increase in the cost of labour resulting from the wage push, firms cannot substantially reduce their labour demand in the short run because of the low elasticity of substitution. However, the change in factor prices induces technical change that would allow firms to be less dependent on labour. Once this technical change takes place, firms gradually reduce their labour demand.[28] Inspection of equation (41) also implies that a decline in μ—*i.e.* an adverse labour supply shock—will reduce N_K/N_L. This corresponds to capital-biased technical change since capital and labour are gross complements.[29] Figure 4 draws this case diagrammatically, assuming that technology is given in the short run when the wage push first occurs, and then traces the adjustment of the economy to the shock. It shows diagrammatically how, as in the case of European economies, the wage push will first increase the labour share, and then gradually reduce it by creating capital-biased technical change. Throughout the process, employment falls.

7. CONCLUDING REMARKS

For many problems in macroeconomics, development economics, labour economics, and international trade, whether technical change is biased towards particular factors is of central importance. This paper synthesized some recent research on the determinants of biased technical change. The presumption is that the same economic forces—profit incentives—that affect the amount of technical change will also shape the direction of technical change, and therefore determine the equilibrium bias of technology. I argued that this perspective helps us understand a number of otherwise puzzling patterns.

I also highlighted the various modelling choices involved in thinking about the direction of technical change. I demonstrated that there are two forces affecting equilibrium bias: the price

28. This idea is related to that proposed by Caballero and Hammour (1998). They suggest that during the 1980's, in order to avoid expropriation by labour, European firms chose more capital intensive production techniques, reducing their labour demand.

29. However, notice that in this model, the labour share simply returns back to its original level. So to explain why the labour share may have fallen below its original level in many European countries, we may also need some of the adverse labour supply shocks to have reversed themselves (*i.e.* μ to have fallen back towards its initial level).

effect and the market size effect. The elasticity of substitution between different factors regulates how powerful these effects are, and this has implications for how technical change and factor prices respond to changes in relative supplies.

Another important determinant of equilibrium bias of technology is the form of the innovation possibilities frontier—how relative costs of different types of innovation change with the current state of technology. The choice here is between a specification that emphasizes state dependence, whereby past innovations complementing a factor make current innovations directed at that factor cheaper, and a specification without state dependence. It appears that in thinking about skill-biased technical change a specification with only limited state dependence leads to more plausible results, while in the case of capital- and labour-augmenting technical change, a specification with extreme state dependence appears more appropriate. Although it is possible to combine these features to have a unified framework that can be applied both to analysing the degree of skill bias and capital bias of technology (see Acemoglu (2001)), this framework is currently highly *ad hoc* and without micro-foundations. Future work towards such a unified model would be very useful.

Whether technical change exhibits this type of state dependence, and whether this state dependence affects different factors differentially, is ultimately an empirical question. Existing evidence does not enable us to reach firm conclusions. Nevertheless, data on patent citations seem to provide a useful starting place, and empirical work that can inform modelling choices in this field is another area for fruitful future research.

Acknowledgements. This paper is a revised version of the Review of Economic Studies lecture delivered at the 2001 Royal Economic Society meeting in Durham. The author thanks Rubén Segura-Cayuela for excellent research assistance and Orazio Attanasio, David Autor, Olivier Blanchard, Nobu Kiyotaki, Mark Sanders, Rubén Segura-Cayuela, Jaume Ventura, three anonymous referees, and seminar participants at MIT, Institute for International Economics at Stockholm and the Royal Economics Society for comments.

REFERENCES

ACEMOGLU, D. (1998), "Why do New Technologies Complement Skills? Directed Technical Change and Wage Inequality", *Quarterly Journal of Economics*, **CXIII**, 1055–1090.

ACEMOGLU, D. (1999*a*), "Patterns of Skill Premia" (NBER Working Paper No. 7018).

ACEMOGLU, D. (1999*b*), "Labor- and Capital-augmenting Technical Change" (NBER Working Paper No. 7544).

ACEMOGLU, D. (2001), "Directed Technical Change" (NBER Working Paper No. 8287).

ACEMOGLU, D. (2002), "Technical Change, Inequality and the Labor Market", *Journal of Economic Literature*, **XL**, 7–72.

ACEMOGLU, D. and ZILIBOTTI, F. (2001), "Productivity Differences", *Quarterly Journal of Economics*, **CXVI**, 563–606.

AGHION, P. and HOWITT, P. (1992), "A Model of Growth Through Creative Destruction", *Econometrica*, **LX**, 323–351.

ANGRIST, J. (1995), "The Economic Returns to Schooling in the West Bank and Gaza Strip", *American Economic Review*, **85**, 1065–1087.

ATKINSON, A. B. and STIGLITZ, J. E. (1969), "A New View of Technological Change", *Economic Journal*, **79**, 573–578.

AUTOR, D., KRUEGER, A. and KATZ, L. (1998), "Computing Inequality: Have Computers Changed the Labor Market?", *Quarterly Journal of Economics*, **CXIII**, 1169–1214.

BAIROCH, P. (1988) *Cities and Economic Development: From the Dawn of History to the Present* (Chicago: University of Chicago Press) translated by Christopher Braider.

BASU, S. and WEIL, D. N. (1998), "Appropriate Technology and Growth", *Quarterly Journal of Economics*, **CXIII**, 1025–1054.

BERNDT, E. (1976), "Reconciling Alternative Estimates of the Elasticity of Substitution", *Review of Economics and Statistics*, **58** (1), 59–68.

BLANCHARD, O. (1997), "The Medium Run", *Brookings Papers on Economic Activity*, 89–158.

BRESNAHAN, T. F. and TRAJTENBERG, M. (1995), "General Purpose Technologies: 'Engines of Growth'?", *Journal of Econometrics*, **65** (1), 83–108.

CABALLERO, R. J. and HAMMOUR, M. (1998), "Jobless Growth: Appropriability, Factor Substitution and Unemployment", *Carnegie-Rochester Conference Proceedings,* Vol. 48, 51–94.

CABALLERO, R. J. and JAFFE, A. B. (1993), "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth", in *NBER Macroeconomics Annual 1993* (Cambridge, MA: MIT Press) 15–74.

COEN, R. M. (1969), "Tax Policy and Investment Behavior: A Comment", *American Economic Review*, 370–379.

DAVID, P. (1975) *Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century* (London: Cambridge University Press).

DRANDAKIS, E. and PHELPS, E. (1965), "A Model of Induced Invention, Growth and Distribution", *Economic Journal*, **76**, 823–840.

EISNER, R. and NADIRI, M. I. (1968), "Investment Behavior and the Neoclassical Theory", *Review of Economics and Statistics*, 369–382.

FELLNER, W. (1961), "Two Propositions in the Theory of Induced Innovations", *The Economic Journal*, **71** (282), 305–308.

FREEMAN, R. (1986), "Demand for Education", in A. Orley and L. Richard (eds.) *Handbook of Labor Economics*, Chapter 6, Vol. I (North-Holland) 357–386.

GALOR, O. and MAOV, O. (2000), "Ability Biased Technological Transition, Wage Inequality and Economic Growth", *Quarterly Journal of Economics*, **115**, 469–497.

GOLDIN, C. and KATZ, L. (1995), "The Decline of Non-competing Groups: Changes in the Premium to Education, 1890 to 1940"(NBER Working Paper No. 5202).

GOLDIN, C. and KATZ, L. F. (1998), "The Origins of Technology-skill Complementarity", *Quarterly Journal of Economics*, **113**, 693–732.

GROSSMAN, G. and HELPMAN, E. (1991) *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press).

HABAKKUK, H. J. (1962) *American and British Technology in the Nineteenth Century: Search for Labor Saving Inventions* (Cambridge University Press).

HECKMAN, J., LOCHNER, L. and TABER, C. (1998), "Tax Policy and Human Capital Formation", *American Economic Review Papers and Proceedings*, **88**, 293–297.

HICKS, J. (1932) *The Theory of Wages* (London: Macmillan).

JAFFE, A. B., TRAJTENBERG, M. and HENDERSON, R. (1993), "Geographic Localizaion of Knowledge Spillovers as Evidenced by Patent Citations", *Quarterly Journal of Economics*, **CVIII**, 577–598.

JAMES, J. A. and SKINNER, J. S. (1985), "The Resolution of the Labor-Scarcity Paradox", *Journal of Economic History*, **45**, 513–540.

JONES, C. I. (1995), "R&D-based Models of Economic Growth", *Journal of Political Economy*, **103**, 759–784.

KATZ, L. and AUTOR, D. (2000), "Changes in the Wage Structure and Earnings Inequality", in O. Ashenfelter and D. Card (eds.) *The Handbook of Labor Economics*, Vol. III (Amsterdam: Elsevier).

KENNEDY, C. (1964), "Induced Bias in Innovation and the Theory of Distribution", *Economic Journal*, **LXXIV**, 541–547.

KILEY, M. (1999), "The Supply of Skilled Labor and Skill-Biased Technological Progress", *Economc Journal*.

KRUSELL, P., OHANIAN, L., RIOS-RULL, V. and VIOLANTE, G. (2000), "Capital Skill Complementary and Inequality", *Econometrica*, **68** (5), 1029–1053.

LLOYD-ELLIS, H. (1999), "Endogenous Technological Change and Wage Inequality", *American Economic Review*, **89**, 47–77.

LUCAS, R. E. (1969), "Labor–Capital Substitution in U.S. Manufacturing", in A. Harberger and M. J. Bailey (eds.) *The Taxation of Income from Capital* (Washington, DC: The Brookings Institution) 223–274.

MOKYR, J. (1990) *The Levers of Riches: Technological Creativity and Economic Progress* (New York: Oxford University Press).

NELSON, R. and PHELPS, E. (1966), "Investment in Humans, Technological Diffusion and Economic Growth", *American Economic Association Papers and Proceedings,* Vol. 56, 69–75.

NORDHAUS, W. (1973), "Some Skeptical Thoughts on the Theory of Induced Innovation", *Quarterly Journal of Economics*, **LXXXVII**, 208–219.

REBELO, S. (1991), "Long-run Policy Analysis and Long-run Growth", *Journal of Political Economy*, **99**, 500–521.

RIVERA-BATIZ, L. A. and ROMER, P. M. (1991), "Economic Integration and Endogenous Growth", *Quarterly Journal of Economics*, **106**, 531–555.

ROMER, P. M. (1990), "Endogenous Technological Change", *Journal of Political Economy*, **IIC**, S71–S102.

ROTHBARTH, E. (1946), "Causes of the Superior Efficiency of the USA Industry as Compared with British Industry", *Economics Journal*.

SALTER, W. E. G. (1966) *Productivity and Technical Change* 2nd edition (Cambridge, U.K.: Cambridge University Press).

SAMUELSON, P. (1965), "A Theory of Induced Innovations along Kennedy–Weisacker Lines", *Review of Economics and Statistics*, **XLVII**, 444–464.

SCHMOOKLER, J. (1966) *Invention and Economic Growth* (Cambridge: Harvard University Press).

SCHULTZ, T. (1975), "The Value of the Ability to Deal with Disequalibria", *Jouranl of Economic Literature*, **13**, 827–846.

SEGERSTROM, P. S., ANANT, T. and DINOPOULOS, E. (1990), "A Schumpeterian Model of the Product Life Cycle", *American Economic Review*, **80**, 1077–1092.

STEWART, F. (1977) *Technology and Underdevelopment* (London, U.K.: The Macmillan Press Ltd.).

TRAJTENBERG, M., HENDERSON, R. and JAFFE, A. B. (1992), "Ivory Tower Vs. Corporate Lab: An Empirical Study of Basic Research and Appropriatability" (NBER Working Paper 4146).

WILLIAMSON, J. (1990) *Coping with City Growth during the British Industrial Revolution* (Cambridge: Cambridge University Press).

WOOD, A. (1994) *North–South Trade, Employment and Inequality: Changing Fortunes in a Skill Driven World* (Oxford: Clarendon Press).

XU, B. (2001), "Endogenous Technology Bias, International Trade and Relative Wages" (Mimeo, University of Florida).

YOUNG, A. (1993), "Invention and Bounded Learning by Doing", *Journal of Political Economy*, **101**, 443–472.