Factor Prices and Technical Change: From Induced Innovations to Recent Debates*

Daron Acemoglu
Massachusetts Institute of Technology

October 22, 2001

Abstract

This paper revisits the induced innovation literature of the 1960s to which Phelps was a major contributor (Drandakis and Phelps, 1965). This literature was the first systematic study of the determinants of technical change and also the first investigation of the relationship between factor prices and technical change. I present a modern reformulation of this literature based on the tools developed by the endogenous growth literature. This reformulation confirms many of the insights of the induced innovations literature, but reveals a new force, which I refer to as the market size effect: there will be more technical change directed at more abundant factors.

I use this modern reformulation to shed light on two recent debates: (1) why is technical change often skill biased, and why has it become more skill biased during recent decades? (2) what is the role of human capital differences in accounting for income differences across countries? Interestingly, an application of this modern reformulation to these debates also reiterates some of the insights of another important paper by Phelps, Nelson and Phelps (1966).

JEL Classification: E25, J31, O30, O31, O33.

Keywords: Biased technical change, endogenous technical change, factor distribution of income, growth, innovation, skill-biased technical change, technology.

*This paper is prepared in honor of Edmund Phelps. I thank Nancy Stokey, Jaume Ventura and participants at the Festschrift conference for comments and Rubén Segura-Cayuela for excellent research assistance.

1 Introduction

In many ways, one can see the "induced innovations" literature of the 1960s as the harbinger of the endogenous growth literature of the 1980s and 1990s. While the endogenous growth literature studies the process of growth at the aggregate, the induced innovations literature attempted to understand what type of innovations the economy would generate, and the relationship between factor prices and technical change. Although it was Hicks in *The Theory of Wages* (1932) who first discussed the issue of induced innovation,[1] the important advances were made during the 1960s by Kennedy (1964), Samuelson (1965), and Drandakis and Phelps (1965), who studied the link between factor prices and technical change.

However, during the 1960s the economics profession did not possess all the tools necessary for a systematic study of these issues. In particular, when firms choose technology in addition to capital and labor, the notion of competitive equilibrium needs to be modified or refined either by introducing technological externalities (Romer, 1986, and Lucas, 1988) or by introducing monopolistic competition (Romer, 1990, Grossman and Helpman, 1991, and Aghion and Howitt, 1992).[2] The absence of the appropriate tools forced this literature to take a number of shortcuts, and ultimately rely on heuristic arguments rather than fully micro-founded models.

In this paper, I briefly survey the approach taken and the problems faced by the induced innovations literature, and then recast the results of this literature in terms of models of endogenous technology (or directed technical change), based on my own recent work (e.g., Acemoglu, 1998, 1999, 2001, and Acemoglu and Zilibotti, 2001, as well as Kiley, 1999). This modeling exercise not only formalizes the contribution of the induced innovations literature, but also highlights a new economic force that did not feature in this literature. In particular, like Hicks, the induced innovations literature emphasized

[1]In particular, Hicks argued that technical change would attempt to replace the more expensive factor. 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..."(p. 124).

[2]In addition, see the work by Segerstrom, Anant and Dinopoulos (1990), Jones (1995), Stokey (1995) and Young (1993), and the survey in Aghion and Howitt (1998).

relative prices as the key determinant of which factors new technologies would save on. In models of endogenous technology, as emphasized by Romer (1990), there is a market size effect because new technologies, once developed, can be used by many firms and workers. This market size effect also features in the analysis of the direction and bias of technical change: there will be greater incentives to develop technologies for more abundant factors. Th2 market size effect is not only of theoretical interest, but turns out to play an important role in a number of recent debates.

In the second part of the paper, I discuss how the directed technical change model, the modern reformulation of its induced innovations literature, sheds light on two recent debates. The first relates to the question of why the demand for skills has been increasing throughout the 20th century, and even accelerated over the past 30 years. The second concerns the role of human capital in economic development. I will argue that a model of directed technical change provides an explanation both for why we should expect technical change to be skill biased in general, and for why this skill biased may have accelerated over recent decades. I will also show that this model points out an interesting interaction between human capital and technology, and via this channel, it suggests why human capital differences may be more important in explaining differences in income per capita across countries than standard models imply.

These two debates are interesting not only because they have been active areas of recent research, but also because they relate to another important contribution of Phelps, Nelson and Phelps (1966). Nelson and Phelps (1966) postulated that human capital is essential for the adoption of new technologies. This view has at least two important implications: first, the demand for skills will increase as new technologies are introduced, and second, economies with high human capital will effectively possess better technologies. These insights are relevant for the two debates mentioned above, since they provide a theory for why technical change may be skill biased, and why human capital differences can be essential in accounting for differences in income per capita across countries. Nelson and Phelps developed this insight in a reduced form model. Interestingly, the directed technical model change provides a framework to derive related results from a more micro-founded model. But these results differ in interesting and empirically testable ways from the predictions of the Nelson and Phelps approach.

The rest of the paper is organized as follows. In the next section, I briefly survey the induced innovations approach of the 1960s. In Section 3, I outline a simple model of directed technical change based on my own research, especially, Acemoglu (2001). I show how this framework captures many of the insights of the induced innovations literature in a micro-founded model, and without encountering some of the problems that this earlier literature faced. In Section 4, I use the directed technical change model to investigate when and why the demand for skills will increase. In Section 5, I discuss the role of human capital in accounting for differences in income and output per worker across countries, and then I use the directed technical change model to highlight how the interaction between technology and skills can lead to large differences in income per capita across countries. Section 6 concludes.

2 A Simple Induced Innovation Model

In this section, I outline a version of the model considered by Drandakis and Phelps (1965), which in turn builds on Kennedy (1964). The economy is populated by a mass of consumers with constant savings rate θ.

All firms have access to a constant-elasticity of substitution (CES) constant returns to scale production function,

$$Y = \left[\gamma \left(N_L L\right)^{\frac{\sigma-1}{\sigma}} + (1-\gamma)\left(N_K K\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where L is labor, which is assumed constant throughout the paper , K is capital, and N_L and N_K are labor- and capital-augmenting technology terms, which are controlled by each individual firm. σ is the elasticity of substitution between labor and capital.[3] For simplicity, there is no depreciation of capital. Although firms hire the profit-maximizing amount of labor and capital, they choose their technologies to maximize "*the current rate of cost reduction*" for given factor proportions (see, Kennedy, 1964, p. 543, Drandakis and Phelps, 1965, p. 824). This is equivalent to maximizing the instantaneous rate of

[3] I choose the CES form to simplify the discussion.

output growth, R, taking K and L as given. To calculate this rate of output growth, simply take logs and differentiate (1) with respect to time, holding K and L constant. This gives:

$$R = s\frac{\dot{N}_L}{N_L} + (1-s)\frac{\dot{N}_K}{N_K}, \tag{2}$$

where s is the share of labor in GDP.

In maximizing R, firms face a constraint first introduced by Kennedy (1964), which I will refer to as the "innovation possibilities frontier".[4] This frontier is of central importance for the arguments of the induced innovations literature, as it specifies the technologically-determined constraints on how labor-augmenting and capital-augmenting technical change can be traded off. Let me for now take a general form

$$\frac{\dot{N}_L}{N_L} = \Gamma\left(\frac{\dot{N}_K}{N_K}\right), \tag{3}$$

where Γ is a strictly decreasing a differentiable and concave function. This frontier captures the intuitive notion that firms (or the economy) have to trade-off a higher rate of labor-augmenting technical change with a lower rate of capital-augmenting technical change. As a result, the innovation possibilities frontier traces a downward sloping locus as shown in Figure 1.

The solution to maximizing (2) subject to (3) takes a simple form satisfying the first-order condition

$$(1-s) + s\Gamma'\left(\frac{\dot{N}_K}{N_K}\right) = 0. \tag{4}$$

Diagrammatically, this solution is drawn in Figure 1 as the tangency of the counters for the instantaneous rate of output growth, (2), and of the innovation facilities frontier, (3), as shown by point A. Comparative statics are straightforward. A greater share of labor in GDP, s, makes labor-augmenting technology more valuable, and increases $\frac{\dot{N}_L}{N_L}$ and reduces $\frac{\dot{N}_K}{N_K}$. Put differently, technical change, very much like Hicks conjectured, tries to replace the expensive factor. A greater s corresponds to labor being more expensive, and as a result, technical change becomes more labor-augmenting. This is the first important insight of the induced innovations literature, which I highlight for future reference:

[4] Kennedy (1964) called this "innovation possibility function", while Drandakis and Phelps (1965) called it "invention possibility frontier".

Result 1: There will be greater technical change augmenting the factor that is more "expensive".

Next note that the growth rate of output is given by

$$\frac{\dot{Y}}{Y}=\frac{\gamma\frac{\dot{N}_L}{N_L}\left(N_L L\right)^{\frac{\sigma-1}{\sigma}}+(1-\gamma)\left(\theta\frac{Y}{K}+\frac{\dot{N}_K}{N_K}\right)\left(N_K K\right)^{\frac{\sigma-1}{\sigma}}}{\gamma\left(N_L L\right)^{\frac{\sigma-1}{\sigma}}+(1-\gamma)\left(N_K K\right)^{\frac{\sigma-1}{\sigma}}}$$

where I have replaced $\frac{\dot{K}}{K}$ by $\theta\frac{Y}{K}$ using the constant saving rule. Let us now look for an equilibrium satisfying the Kaldor facts that the output to capital ratio, $\frac{Y}{K}$, is constant, and output grows at a constant rate. This necessarily implies that $\frac{\dot{N}_K}{N_K}=0$ and $\frac{\dot{N}_L}{N_L}=\theta\frac{Y}{K}=\frac{\dot{Y}}{Y}$. Therefore, technical change has to be purely labor-augmenting (or the tangency point has to be at B as drawn in Figure 1). Moreover, the fact that $\frac{\dot{N}_K}{N_K}=0$ immediately pins down the equilibrium labor share from (4) as:[5]

$$(1-s)+s\Gamma'(0)=0.$$

Intuitively, there are two ways of performing capital-like tasks in this economy: accumulate more capital, or increase the productivity of capital. In contrast, there is only one way of increasing the productivity of labor, via technical change.[6] In equilibrium, all technical change is directed to labor, while capital accumulation increases the supply of capital-like tasks. Therefore, this model not only gives us a theory of the type of technical change, but also a theory of factor shares. This gives a second major result:

Result 2: Induced innovations and capital accumulation determine equilibrium factor shares, ensure that all technical change will be labor augmenting and imply a constant share of labor in GDP in the long run.

Finally, Drandakis and Phelps (1965) also addressed the question of whether the economy would tend to this long-run equilibrium. They showed that as long as the

[5] Also this equation indirectly determines the capital-output ratio to satisfy the requirement that $\frac{\dot{N}_L}{N_L}=\theta\frac{Y}{K}$.

[6] Interestingly, Hicks also anticipated this result. He wrote "The general tendency to a more rapid increase in capital than labor which has marked European history during the last few centuries has naturally provided a stimulus to labor-saving invention" (pp. 125).

elasticity of substitution (between labor and capital) is less than 1, the economy is stable. Intuitively, a greater share of labor in GDP encourages more labor-augmenting technical change as shown by the first-order condition (4). When the elasticity of substitution is greater than 1, this labor-augmenting technical change increases the share of labor even further, destabilizing the system. In contrast, when the elasticity of substitution is less than 1, the share of labor falls and the economy converges to the steady state. This leads to the third major result of the induced innovations literature:

Result 3: As long as the elasticity of substitution between capital and labor is less than 1, the economy converges to the steady state with constant factor shares.

Overall, the induced innovations literature provided the first systematic study of the the determinants of technical change, and the relationship between factor prices and technical change. Moreover, this literature obtained a number of important results that I summarized in Results 1-3.

But there are also some problems with the approach of this literature. The most important lies in the assumption that firms simply maximize (2). Why can't we have profit-maximizing firms instead? The answer relates to the problems that Romer (1986) had to confront in order to construct a model of long-run growth. If aggregate technology has increasing-returns-type characteristics, there does not exist a competitive equilibrium with complete markets. In this context, profit-maximizing firms would solve

$$\max_{K,L,N_K,N_L} \int \exp\left(-rt\right) \left(\left[\gamma \left(N_L L\right)^{\frac{\sigma-1}{\sigma}} + \left(1-\gamma\right) \left(N_K K\right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} - wL - rK \right) dt,$$

subject to (3), and taking the factor prices, w and r, as given. But this problem does not have an interior solution, since the production function exhibits increasing returns to scale. Therefore, to go beyond the heuristics of maximizing the instantaneous rate of cost reduction, we need a micro-founded model of innovation.[7]

[7] See Salter (1960) and Nordhaus (1973) for some of the early criticisms.

3 A Model of Directed Technical Change[8]

How do we incorporate profit-maximizing firms in an economy with increasing returns, and maintain the flavor of a competitive/decentralized equilibrium? Romer (1986) and Lucas (1988) achieved this by introducing technological externalities. In these models, investments (in either physical or human capital) increase the productivity of other firms in the economy, and because individual firms don't internalize this effect, they are subject to constant "private returns" (in the sense that when they double all factors, their profits double, while total output may increase by more). As a result, despite increasing returns at the aggregate level, a competitive equilibrium continues to exist. Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992) developed a different formulation by introducing monopolistic competition, and an explicit discussion of endogenous technical change. In these models final good producers (users of technology) are competitive, but suppliers of technology command market power. As a result, when these monopolistic producers double their inputs, total output increases by more than two folds, but their profits only double because of the decline in the prices they face. Here I will build on this class of models, and extend them to discuss the central issues of the induced innovations literature, the possibility of innovations benefiting factors differentially.

3.1 Basics

Consider an economy that admits a representative consumer with logarithmic preferences

$$\int_0^\infty \ln C e^{-\rho t} dt, \tag{5}$$

where C is constant, ρ is the rate of time preference and the budget constraint is

$$C + I \leq Y \equiv \left[\gamma Y_L^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma) Y_Z^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{6}$$

where I denotes investment. I also impose the usual no-Ponzi game condition, requiring the lifetime budget constraint of the representative consumer to be satisfied. The pro-

[8] The material here liberally borrows from Acemoglu (2001). I omit many of the details and opt for a heuristic presentation to save space and minimize repetition.

duction function in (6) implies that consumption and investment goods are "produced" from an output of two other (intermediate) goods, a labor-intensive good, Y_L, and a good that uses another factor, Z, Y_Z (or alternatively, utility is defined over to a CES aggregate of Y_L, and Y_Z). I am being deliberately vague about the other factor of production, Z, as I want to think of it as skilled workers or capital in different applications.

The two intermediate goods have the following production functions

$$Y_L = \frac{1}{1-\beta}\left(\int_0^{N_L} x_L(j)^{1-\beta} dj\right) L^\beta \text{ and } Y_Z = \frac{1}{1-\beta}\left(\int_0^{N_Z} x_Z(j)^{1-\beta} dj\right) Z^\beta \quad (7)$$

where $\beta \in (0,1)$. The labor-intensive good is produced from labor and a range of labor-complementary machines. $x_L(j)$ denotes the amount of the j-th labor-complementary (labor-augmenting) machine used in production. The range of machines that can be used with labor is denoted by N_L. The production function for the other intermediate uses Z-complementary machines and is explained similarly. It is important that these two sets of machines are different, enabling me to model the fact that some technologies will be augmenting labor, while others increase the productivity of Z.

Although, for given N_L and N_Z, the production functions in (7) exhibit constant returns to scale, when N_L and N_Z are chosen by the firms in the economy, there will be increasing returns in the aggregate.

I assume that machines in both sectors are supplied by profit-maximizing "technology monopolists". Each monopolist will set a rental price $\chi_L(j)$ or $\chi_Z(j)$ for the machine it supplies to the market in order to maximize its profits. For simplicity, I assume 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.

First suppose that N_L and N_Z are given. Then an equilibrium consisis of machine prices $\chi_L(j)$ or $\chi_Z(j)$ that maximize the profits of technology monopolists, machine demands from the two intermediate goods sectors $x_L(j)$ or $x_Z(j)$ that maximize intermediate good producers' profits, and factor and product prices w_L, w_Z, p_L, and p_Z that clear markets.

This equilibrium is straightforward to characterize. Since the product markets for the two intermediates are competitive, product prices satisfy

$$p \equiv \frac{p_Z}{p_L} = \frac{1-\gamma}{\gamma}\left(\frac{Y_Z}{Y_L}\right)^{-\frac{1}{\varepsilon}}. \tag{8}$$

The greater the supply of Y_Z relative to Y_L, the lower is its relative price, p.

Firms in the labor-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, $\chi_L(j)$, as well as the range of machines, N_L, as given. The maximization problem facing firms in the Z sector is similar. The first-order conditions for these firms give machine demands as:

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

The important point that comes out from these expressions is that a greater level of employment of a factor raises the demand for machines complementing that factor, because it creates a greater market size for the machines. This observation underlies the market size effect that was emphasized in the introduction.

The profits of a monopolist supplying labor-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 markup over marginal cost: $\chi_L(j) = \frac{\psi}{1-\beta}$. To simplify the algebra, I normalize the marginal cost to $\psi \equiv 1-\beta$. This implies that in equilibrium all machine prices will be given by $\chi_L(j) = \chi_Z(j) = 1$.

Using this price, machine demands given by (10) and the assumption of competitive factor markets, we can also calculate relative factor rewards as:

$$\frac{w_Z}{w_L} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\sigma}} \left(\frac{N_Z}{N_L}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{Z}{L}\right)^{-\frac{1}{\sigma}}, \tag{11}$$

where

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

Inspection of (11) immediately shows that σ is the elasticity of substitution between factors, which in turn is derived from the elasticity of substitution between the goods in consumers' utility function, ε.

For a given state of technology as captured by N_Z/N_L, the relative factor reward, w_Z/w_L, is decreasing in the relative factor supply, Z/L. This is the usual substitution effect: the more abundant factor is substituted for the less abundant one, and its relative marginal product falls.

To determine the direction of technical change, we need to calculate the profitability of different types of innovations. Using the profit-maximizing machine prices, $\chi_L(j) = \chi_Z(j) = 1$, and machine demands given by (10), we find the monopoly profits as:

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

Then, the discounted net present value of technology monopolists are given by standard Bellman equations:

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

where r is the interest rate. These value functions have a familiar intuitive explanation, equating the discounted value, rV_L or rV_Z, to the flow returns, which consist of profits, π_L or π_Z, and the appreciation of the asset at hand, $\dot{V}_L$ or $\dot{V}_Z$.

In steady state, the prices, p_L or p_Z, and the interest rate will be constant, r, so $\dot{V}_L = \dot{V}_Z = 0$ and

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

The greater V_Z is relative to V_L, the greater are the incentives to develop Z-complementary machines, N_Z, rather than labor-complementary machines, N_L. Equation (14) highlights two effects on the direction of technical change:

1. The price effect: there will be greater incentives to invent technologies producing more expensive goods (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. The market size effect encourages innovation for the more abundant factor (V_L and V_Z are increasing in Z and L, the total supplies of the factors combined with these technologies).

The price effect is the analogue of Result 1 of the induced innovations literature, showing that there will be more technical change directed towards more "expensive" factors (note that $w_L = \beta N_L p_L^{1/\beta} / (1-\beta)$ and $w_Z = \beta N_Z p_Z^{1/\beta} / (1-\beta)$). In addition, this model introduces a new force, the market size effect, which will play an important role in the applications below, and is essential for a new result discussed below, that an increase in the supply of a factor induces technical change biased toward that factor.

3.2 The Innovation Possibilities Frontier

Instead of (3), which specified an innovation possibilities frontier trading off the two types of innovations, we now need a frontier that transforms actual resources into either of the two types of innovations. This frontier will embed a specific form of (3). In addition, we have to be careful in the choice of the form of the innovation possibilities frontier, since, as is well known in the endogenous growth literature, many specifications will not enable long-run growth. In particular, when there are scarce factors used for R&D, sustained growth requires these factors to become more and more productive over time, for example due to spillovers from past research.[9]

Here I assume that R&D is carried out by scientists, and there is a constant supply of scientists equal to S. With only one sector, sustained growth requires $\dot{N}/N$ to be proportional to S: that is, current research benefits from the stock of past innovations, N. With two sectors, instead, there is a variety of specifications, with possible interactions between the two sectors. In particular, each sector can benefit either from its own stock of past innovations, or from a combination of its own stock and the stock of the other sector.

To clarify the issues, it is useful to first introduce the concept of "state dependence". The degree of state dependence relates to how future relative costs of innovation are affected by the current composition of R&D. When current R&D in a sector benefits more from its own stock of past innovations than past innovations complementing the other factor, I refer to the innovation possibilities frontier as "state dependent". A

[9] See Acemoglu (1998, 2001) for models of directed technical change where R&D uses final output, and there are no spillovers.

flexible formulation is:

$$\dot{N}_L = \eta_L N_L^{(1+\delta)/2} N_Z^{(1-\delta)/2} S_L \text{ and } \dot{N}_Z = \eta_Z N_L^{(1-\delta)/2} N_Z^{(1+\delta)/2} S_Z, \tag{15}$$

where $\delta \in (0,1)$ measures the degree of state dependence: when $\delta = 0$, there is no state dependence since both N_L and N_Z create symmetric spillovers for current research in the two sectors. In contrast, when $\delta = 1$, there is an extreme amount of state dependence because an increase in the stock of labor-complementary machines today makes future labor-complementary innovations cheaper, but has no effect on the cost of Z-complementary innovations.[10]

To find the steady state equilibrium of this economy, we need to equate the relative profitability of Z-augmenting technical change to its relative cost. The relative profitability is V_Z/V_L given by (14), while the relative cost is

$$\left(\partial \dot{N}_Z / \partial S_Z\right) / \left(\partial \dot{N}_L / \partial S_L\right) = \eta_Z N_Z^{\delta} / \eta_L N_L^{\delta}.$$

The steady-state equilibrium condition is then:

$$\eta_L N_L^{\delta} \pi_L = \eta_Z N_Z^{\delta} \pi_Z. \tag{16}$$

Now solving condition (16) together with (12), we obtain the equilibrium relative technology as

$$\frac{N_Z}{N_L} = \left(\frac{\eta_Z}{\eta_L}\right)^{\frac{\sigma}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{1-\delta\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-1}{1-\delta\sigma}}. \tag{17}$$

In Acemoglu (2001), I show that the equilibrium will be stable as long as $\sigma < \delta^{-1}$, and here I simply assume that this condition is satisfied, so $1 - \delta\sigma > 0$.

There are a number of important results that follow from equation (17). First, the relative degree of Z-augmenting technology simply depends on the relative supply of Z. This captures both the price and the market size effects emphasized above, since in equilibrium prices are also determined by relative supplies (from equation (8)). Second, in the case when the elasticity of substitution between the two factors is greater than

[10] With existing data it is not possible to determine the extent of state dependence. In Acemoglu (2001), I argued that results from the patent citations literature, e.g., Trajtenberg, Henderson, and Jaffe (1992), suggest that there is at least some degree of state dependence in the innovation process.

1, i.e., $\sigma > 1$, an increase in the relative supply of Z increases N_Z/N_L. That is, greater Z/L leads to Z-augmenting technical change. This is a consequence of the market size effect. Greater relative supply of Z creates more demand for machines complementing this factor. This in turn increases the productivity of Z further.

One way to illustrate the implications of the market size effect is to compare the constant-technology relative demand curve for Z, which keeps N_Z/N_L constant (equation (11)), to the endogenous-technology relative demand where N_Z/N_L is given by (17). With this purpose, substitute (17) into (11) to obtain the endogenous-technology relative demand:

$$\frac{w_Z}{w_L} = \left(\frac{\eta_Z}{\eta_L}\right)^{\frac{\sigma-1}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{(1-\delta)\varepsilon}{1-\delta\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-2+\delta}{1-\delta\sigma}}. \tag{18}$$

Figure 2 draws the endogenous-technology relative demand curve given by (18), ET_1, together with the constant-technology demand curve, CT. It is straightforward to verify that with $\sigma > 1$, ET_1 is flatter than CT—the increase in Z/L raises N_Z/N_L and the relative demand for Z. This is because, as pointed out above, changes in technology increase the demand for the factor that has become more abundant.

Now contrast the previous case to the one where the elasticity of substitution, σ, is less than 1. Because with $\sigma < 1$, an increase in the relative supply of Z now reduces N_Z/N_L, making technology more labor-augmenting. This is the price effect dominates the market size effect, and new technologies are directed at the factor that has become more scarce. Interestingly, however, the endogenous-technology relative demand curve is still flatter, as drawn in Figure 2 (again simply compare (18) to (11)). Why is this? Because when $\sigma < 1$, a decline in N_Z/N_L is actually *biased* towards Z (see equation (11), and the discussion in Acemoglu, 2001). So irrespective of the elasticity of substitution (as long as it is not equal to 1, i.e., as long as we are not in the Cobb-Douglas case), the long-run relative demand curve is flatter than the short-run relative demand curve. At some level, this is an application of the LeChatelier principle, which states that factor demands become flatter when all other factors adjust (here these other factors correspond to “technology”).

But there is more to this framework than the LeChatelier principle. Somewhat surprisingly, the relative demand curve can be upward sloping as shown by ET_2 in

Figure 2. This happens when

$$\sigma > 2 - \delta. \tag{19}$$

Intuitively, the increase in the relative supply of Z raises the demand for Z-complementary innovations, causing Z-biased technical change. This biased technical change increases the marginal product of Z more than that of labor, and despite the substitution effect, the relative reward for the factor that has become more abundant increases. Inspection of (19) shows that this can only happen when $\sigma > 1$. That is, for an upward-sloping relative demand curve, we need the market size effect to be strong enough. The possibility of an upward-sloping relative demand curve for skills is of interest not only to show the strength of directed technical change, but also because it will play an important role in the first application of this framework below.

So we now have a micro-founded framework that can be used to study the direction of technical change and to determine towards which factors new technologies will be biased. This framework also leads to an analogue of Result 1 of the induced innovations literature. Does it also capture the insights of the induced innovations literature related to the behavior of the shares of capital and labor in GDP? To answer this question, let us look at the implications of directed technical change for factor shares. Multiplying both sides of equation (18) by Z/L, we obtain relative factor shares as:

$$\frac{s_Z}{s_L} \equiv \frac{w_Z Z}{w_L L} = \left(\frac{\eta_Z}{\eta_L}\right)^{\frac{\sigma-1}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{(1-\delta)\varepsilon}{1-\delta\sigma}} \left(\frac{Z}{L}\right)^{\frac{\sigma-1+\delta-\delta\sigma}{1-\delta\sigma}}. \tag{20}$$

This equation shows that there is no reason to expect endogenous technical change to keep factor shares constant. In fact, generally as Z/L changes (e.g., due to capital accumulation as in the simple induced innovation model of Section 2), s_Z/s_L will change.

However, factor shares *will be* constant in an interesting, and perhaps empirically relevant, special case where there is full state dependence in the innovation possibilities frontier, that is, when $\delta = 1$ in terms of equation (15). In this special case, the accumulation equations take the familiar-looking simple form

$$\frac{\dot{N}_L}{N_L} = \eta_L S_L \text{ and } \frac{\dot{N}_Z}{N_Z} = \eta_Z S_Z. \tag{21}$$

That is, research effort devoted to one sector leads to proportional improvements in that sector, and has no effect on the other sector. Using this formulation of the innovation possibilities frontier and multiplying both sides of (11) by Z/L, we now have:

$$\frac{s_Z}{s_L} = \frac{\eta_L}{\eta_Z}. \tag{22}$$

Hence, in this case, directed technical change works to stablize factor shares. This formulation of the innovation possibilities frontier therefore leads to the second major result, Result 2, of the induced innovations literature, but with a micro-founded model where firms maximize profits and with the caveat that this result only obtains for a specific formulation of the innovation possibilities frontier.

Now imagine that Z stands for capital as in the canonical model of the induced innovations literature. Capital accumulation is given by the optimal consumption decision of the representative consumer. Hence we have the Euler equation

$$\frac{\dot{C}}{C} = r - \rho, \tag{23}$$

where r is the interest rate, and capital accumulation is given by

$$\dot{Z} = Y - C. \tag{24}$$

In steady state, equation (22) ensures that factor shares are constant. In addition, capital accumulation, which follows from (24), implies that there will be labor-augmenting technical change. This can be seen by first noting that the relative share of capital is given by

$$\frac{s_Z}{s_L} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\sigma}} \left(\frac{N_Z}{N_L}\frac{Z}{L}\right)^{\frac{\sigma-1}{\sigma}}.$$

Inspection of this equation shows that this relative factor share can only remain constant if $\frac{N_L}{N_Z}$ is growing at the same rate as capital for worker, $\frac{Z}{L}$. So technical change has to be labor-augmenting. In fact, the result here is stronger than this: in steady state the interest rate has to remain constant in the long run, and so N_Z has to remain constant (see Acemoglu, 1999).[11] This model then predicts that the share of capital should

[11] Briefly, the interest rate is $r = \beta N_Z p_Z^{1/\beta}/(1-\beta)$ and to ensure balanced growth, the interest rate has to be constant. Along the balance growth path, p_Z is constant, so N_Z has to remain constant.

stay constant along the steady-state equilibrium, and technical change should be purely labor-augmenting.[12]

Is this steady-state equilibrium stable? Recall that in this framework stability requires $\sigma < \delta^{-1}$. In addition, to ensure the constancy of factor shares in the long run, we now have $\delta = 1$. Therefore, as in Drandakis and Phelps (1965), stability requires the elasticity of substitution to be less than 1. The intuition is also similar to that in the induced innovations literature: when $\frac{s_Z}{s_L} > \frac{\eta_L}{\eta_Z}$, the share of the Z factor is sufficiently large that all firms want to undertake Z-augmenting technical change. If the introduction of new technologies increases $\frac{s_Z}{s_L}$ further, the economy will diverge away from steady state. Therefore, for the steady state to be stable, we need Z-augmenting technical change to reduce the relative factor share of Z. This requires the two factors to be gross complements, i.e., $\sigma < 1$ (see Acemoglu, 1999, for details). Therefore, Result 3 of the induced innovations literature also follows from this more micro-founded framework.

I next discuss how the directed technical change model, the modern reformulation of the induced innovation literature, sheds new light on two recent debates.

4 Debate 1: Why Is Technical Change Skilled Biased?

4.1 Skill-Biased Technical Change

The general consensus among labor and macro economists is that technical change has been skill-biased over the past 60 years, and most probably throughout the 20th century.

Figure 3 summarizes the most powerful argument for why we should think that technical change has been favoring skilled workers more than unskilled workers. It plots a measure of the relative supply of skills (the number of college equivalent workers

[12] The reader may notice one difference between the reasoning for constant factor shares in the induced innovations model and in my model. In the induced innovations model, the constancy of the factor shares follows from the requirement that there has to be steady capital accumulation. Otherwise, equation (4) is consistent with different factor shares. In contrast, in my model equation (22) implies that "any technology equilibrium" has to give constant factor shares. Capital accumulation can then be introduced into this framework, as discussed briefly above and much more in detail in Acemoglu (1999), and leads to the conclusion that all technical change has to be labor-augmenting.

divided by noncollege equivalents) and a measure of the return to skills (the college premium).[13] 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. If technical change had not been biased toward skilled workers (and presuming that skilled and unskilled workers are imperfect substitutes), we would expect a large decline in the returns to skills. On the contrary, over this time period, returns to skills, as proxied by the college premium, appear to have increased. The figure also shows that despite the rapid increase in the supply of skills, the college premium has been increasing very rapidly over the recent decades. Most economists attribute this pattern to an acceleration in the skill bias of new technologies.

Why has technical change been skill biased throughout the 20th century? And why has it become more skill biased over the recent decades? There are two popular explanations. In the first explanation, technical change is sometimes skill-biased, but there is no theory for when we should expect more skill bias. This approach can obviously account for the patterns we observe, since it has enough degrees of freedom. And for this reason, it is not a particularly attractive approach. Moreover, according to this explanation the pattern whereby technical change appears to have become more skill-biased during the past 25 years, precisely when the supply of skills has increased very rapidly, has to be viewed as a coincidence.

The second explanation, interestingly, builds on another seminal paper by Phelps, Nelson and Phelps (1966). Nelson and Phelps suggested that human capital and skills are important for the absorption and use of new technologies. (I outline a simple version of the Nelson and Phelps model in the Appendix). According to this view, the demand for skills has been increasing because there have been more (or more high-tech) new technologies during the 20th century, and perhaps because the rate of technical change

[13] The samples are constructed as in Katz and Autor (2000). I thank David Autor for providing me with data from this study. Data from 1939, 1949 and 1959 come from 1940, 1950 and 1960 censuses. The rest of the data come from 1964-1997 March CPSs. The college premium is the coefficient on workers with a college degree or more relative to high school graduates in a log weekly wage regression. The relative supply of skills is calculated from a sample that includes all workers between the ages of 18 and 65. It is defined as the ratio of college equivalents to non-college equivalents, calculated as an Autor, Katz and Krueger (1998) using weeks worked as weights.

has increased during the past 30 years. This explanation also has some drawbacks. First, the original Nelson-Phelps model and its modern reformulations (e.g., Greenwood and Yorukoglu, 1997, Galor and Maov, 2000, Aghion, Howitt and Violante, 2000) are very "reduced form": the adoption and use of new technologies is simply assumed to be skill-intensive. And yet, one can imagine new technologies simplifying previously complex tasks, such as scanners. Second, there are many historical examples of skill-replacing technologies, such as weaving machines, the factory system, the interchangeable parts technology, and the assembly line. So the presumption that new technologies always increase the demand for skills is not entirely compelling.

I will next show that an approach based on directed technical change provides an explanation for why technical change has been skill biased over the past century and became more skilled biased over the past 30 years, without assuming technical change to be always and everywhere skill biased.

4.2 Directed Technical Change and Skill Bias[14]

Consider the model of Section 3, with Z interpreted as the number of skilled workers, and L as the number of unskilled workers. Then, equation (17) gives the degree of skill bias of technology, and leads to a number of interesting implications.

First, an increase in the relative supply of skills will encourage the development of skill-biased technologies. Throughout the 20th century, the relative supply of skills has increased very rapidly, both in the U.S. and in most other OECD economies. Therefore, the framework here suggests that this increase in the supply of skills should have induced new technologies to become more and more skill biased.

Second, with the same reasoning, when the supply of skills accelerates, we expect the degree of skill bias of new technologies to accelerate. Moreover, recall that when condition (19) is satisfied, equation (18) traces an upward-sloping long-run relative demand curve for skills. Therefore, the induced skill bias of new technologies can be sufficiently pronounced that the skill premium may increase in response to a large increase in the supply of skills. In fact, this model, together with condition (19), provides an attrac-

[14] This material builds on Acemoglu (1998).

tive explanation for the behavior of the college premium over the past 30 years shown in Figure 3. Because technology is slow to change (i.e., because N_Z and N_L are state variables), it is reasonable to expect the first response of the college premium to a large increase in supplies to be along the constant-technology demand curve. That is, in response to the increase in the supply of skills, returns to skills will initially decline. Then once technology starts adjusting, the economy will move to the upward-sloping long-run relative demand curve, and returns to skills will increase sharply. Figure 4 draws this case diagrammatically.

Therefore, this theory can explain the secular skill bias of technical change, why technical change has become more skill biased during recent decades, and also why in response to the large increase in the supply of skills of the 1970s, the college premium first fell, and then started a sharp increase in the 1980s.[15]

5 Debate 2: The Role of Human Capital in Cross-Country Income Differences

5.1 Human Capital and Income Differences

There are large differences in human capital across countries. For example, while the average years of schooling of the population in 1985 was just under 12 in the United States and New Zealand, it was less than 2 years in much of sub-Saharan Africa. Can these differences in educational attainment (more generally, human capital) be the proximate or the ultimate cause of the large differences in income per capita across countries? A number of economists, including Lucas (1988), Azariadis and Drazen (1990), Becker, Murphy and Tamura (1990), Stokey (1991), and Mankiw, Romer and Weil (1992), have emphasized the role of human capital in cross-country differences in income levels and growth rates.

Can these effects be quantitatively large? The recent literature on decomposing

[15] This framework also suggests a reason why many technologies developed during the early 19th century may have been skill-replacing. This is because, during this time period, there was a large increase in the supply of unskilled labor in British cities during that time period. See Acemoglu (2001) for a more detailed discussion.

differences in income per capita or output per worker across countries into different components concludes that the answer is no. Both Klenow and Rodriguez (1997) and Hall and Jones (1999) find that differences in human capital, or even differences in physical capital, can account for only a fraction of the differences across countries, with the rest due to differences in efficiency of factor use (or due to differences in "technology").

Let me reiterate this point somewhat differently. Figure 5 plots the logarithm of output per worker relative to the U.S. for 103 countries against average years of schooling in 1985. The figure shows a strong correlation between output per worker and schooling. In fact, the bivariate regression line plotted in Figure 5 has a slope coefficient of 0.29 (standard error 0.02) and an R-square of 65 percent.[16] So in a regression sense, it appears that there could be a lot in the differences in human capital. However, a simple calculation suggests that it is difficult to rationalize educational attainment raising income as steeply as suggested by Figure 5.

To see this, note that the "private" return to schooling—the increase in individual earnings resulting from an additional year of schooling—is about 6-10 percent (e.g., Card, 1999). In the absence of human capital externalities, the contribution of a one-year increase in average schooling to total output would be of roughly the same magnitude.[17] But then differences in schooling can explain little of the cross-country variation in income. More specifically, the difference in average schooling between the top and bottom deciles of the world education distribution in 1985 is less than 8 years. With the returns to schooling around 10 percent, we would expect the top decile countries to produce about twice as much per worker as the bottom decile countries. In practice, the output per worker gap between these deciles is approximately 15.

So how could we justify human capital differences playing a more important role in the world distribution of income? There are a number of possible avenues. First, formal schooling is only one component of human capital, and differences in formal schooling may be understating the true differences in human capital. This could be because workers acquire much more on-the-job training in some countries than others, or because

[16] Data on output per worker are from Summers and Heston (1991), with the Hall and Jones (1999) correction. Education data are from Barro and Lee (1993).

[17] And if capital were in scarce supply, it would be lower.

the quality of schooling varies substantially across countries. There is undoubtedly much truth to both of these points, but it appears unlikely that they can be the whole story. For example, workers who migrate to the U.S. from other countries quickly converge to the earning levels similar to those of U.S. workers with similar schooling, suggesting that quality differences are not the major factor behind the differences in income across countries. Moreover, even without controlling for education, these workers earn not much less than U.S. workers (certainly nothing comparable to the output gaps we observe across countries), so differences in output across countries must have more to do with physical capital or technology differences (see Hendricks, 2001).[18]

Second, perhaps the above calculation understates the importance of human capital because it ignores human capital externalities. After all, it's quite plausible that the whole society benefits indirectly from the greater human capital investment of a worker. In fact, externalities were the centerpiece of Lucas' (1988) paper which argued for the importance of human capital differences, and have been a major building block for many of the papers in the endogenous growth literature.

How large do human capital externalities need to be to justify a strong "causal" effect of human capital differences on income differences consistent with the bivariate relationship shown in Figure 5? A back-of-the-envelope calculation suggests that in order to justify the magnitudes implied by Figure 5, human capital externalities need to be very large. The slope of the line in Figure 5, 0.29, is consistent with (total or social) returns to education of approximately 34 percent ($\exp(0.29) - 1 \simeq 0.34$). Or alternatively, the comparison of the top and bottom decile countries implies returns on the order of 40 percent. To rationalize Figure 5, we therefore need human capital externalities of 25-30 percent on top of the 6-10 percent private returns. In other words, external returns created by education needs to be of the order of three to four times the private returns—very large human capital externalities indeed!

What is the evidence? There have been a number of studies attempting to estimate human capital externalities by exploiting differences in average schooling across cities or

[18] These results have to be interpreted with some caution, since it may be workers with relatively high observed or unobserved skills who migrate to the U.S..

states in the U.S. (e.g., Rauch, 1993). However, these studies face serious identification problems. Cities with greater average schooling may also have higher wages for a variety of other reasons. To solve this problem, one needs to find "quasi-exogenous" differences in schooling across labor markets. In Acemoglu and Angrist (2000), Josh Angrist and I attempted to do this by looking at variations in compulsory attendance laws and child labor laws in U.S. states between 1920 and 1960. It turns out that these laws were quite important earlier in the century in determining educational attainment, especially high school graduation rates. For example, we found that a person growing up in a state with strict child labor laws was 5 percent more likely to graduate from high school than a person growing up in a state with the most permissive child labor laws. Moreover, these differences in laws do translate into substantial differences in average schooling across states. These laws therefore provide an attractive source of variation to identify human capital externalities.

So how high are human capital externalities? Contrary to my expectations, we found very small human capital externalities, often not significantly different from 0. Our baseline estimates are around 1 or 2 percent (in other words, they imply plausible externalities of the order of 20 percent of the private returns). These are magnitudes far short of what is required for human capital to be a major ultimate or proximate cause of the differences in income per capita.

Are we then to conclude that human capital differences across countries are more of a symptom than the cause of the differences in income? Perhaps this is the correct conclusion to draw. But it is too early to jump to this conclusion, since there is one more line of attack: human capital differences can affect the type of technologies that a county adopts, and how effectively these technologies are being utilized. At an intuitive level, human capital differences have a much larger effect on income when they interact with technology choices. This is the insight that comes out of Nelson and Phelps (1966), which is discussed in the Appendix, and also follows from the modern reformulation of the induced innovations literature that I discussed in Section 3 (see below).[19]

Can human capital play a much more important role in accounting for cross-country

[19] This point is developed in my joint work with Fabrizio Zilibotti (see Acemoglu and Zilibotti, 2001).

income differences when it interacts with technology? Although existing evidence does not enable us to answer this question, there are a few pieces of evidence that suggest that there might be something here. First, Benhabib and Spiegel (1994) present cross-country regression evidence consistent with this view. Second, there is micro evidence consistent with the notion that human capital facilitates the absorption and use of new and more productive technologies (e.g., Foster and Rosenzweig, 1996). Third, in Acemoglu and Zilibotti (2001), Fabrizio and I undertook a simple calibration of a model with skill-technology mismatch and found that it can account for a large fraction of the actual differences in income per capita with the differences in human capital across countries.

Finally, a quick look at the cross-country data shows that there is a strong correlation between "technology" and human capital, especially, a measure of the relative supply of high human capital workers. For example, Figure 6 plots the TFP measure (relative to the U.S.) calculated by Hall and Jones (1999) from a cross-country levels accounting exercise against the ratio of college to noncollege ratio of workers in the population in 1985 (from Barro and Lee, 1993). This correlation might of course reflect the effect of some other factors on both variables, for example, institutional differences across countries as emphasized by Hall and Jones (1999) and Acemoglu, Johnson and Robinson (2001), or the effect of higher (exogenous) productivity on human capital investments. Nevertheless, it suggests that a more careful look at the relationship between human capital and technology is required. This paper, naturally, is not the right forum for this, so here instead I will simply develop an alternative approach to analyze the interaction between human capital and technology based on the insights of the directed technical change/induced innovations literature.

5.2 Directed Technical Change, Appropriate Technology and Human Capital

The notion of directed technical change implies that new technologies in the U.S. or in the OECD countries, which are often used by LDCs, will be designed to work best with the conditions and factor supplies in these technologically more advanced nations. Because these nations are more abundant in human capital, frontier technologies they develop will typically require highly skilled workers. Lack of skilled personnel in LDCs will then create a technology-skill mismatch, making it difficult for these countries to benefit from frontier technologies. In other words, these technologies will be, at least to some degree, "inappropriate" to the LDCs' needs. This insight follows from an application of the directed technical change model, but is also quite related to the insights of Nelson and Phelps (1966).

To develop these ideas more formally, consider the model of Section 3 applied to a world economy. A technology leader that I refer to as the North, e.g., the U.S., produces the technologies N_Z and N_L, while the other countries, the South or the LDCs, simply copy these. For simplicity, I take all of the LDCs to be identical, with L' unskilled workers and Z' skilled workers. A key characteristic of the LDCs is that they are less abundant in skilled workers than the North, that is:

$$\frac{Z'}{L'} < \frac{Z}{L}.$$

Assume that LDCs can copy new machine varieties invented in the North, without paying royalties to Northern technology monopolists because of lack of intellectual property rights. This assumption implies that the relevant markets for the technology monopolists will be given by the factor supplies in the North. I also assume that the cost of producing machines in the LDCs may be higher, $\kappa^{-\beta/(1-\beta)}$ rather than $\psi \equiv (1-\beta)$ 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. I also assume that there is free entry to copying Northern machines. This implies that all machines will sell at marginal cost in the LDCs, that is, at the price $\kappa^{-\beta/(1-\beta)}$. It is natural to think of κ as less than 1, so that machines are more expensive in the South than in the North. Finally, there is no international trade.

Using the above expressions, we obtain the output levels of the two final goods in the North as:

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

while in the LDCs, we have:

$$Y_L' = \frac{1}{1-\beta}\,(p_L')^{(1-\beta)/\beta}\,\kappa N_L L' \text{ and } Y_Z' = \frac{1}{1-\beta}\,(p_Z')^{(1-\beta)/\beta}\,\kappa N_Z Z',$$

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 costs are different in the South. Notice also that the technology terms, N_L and N_Z, are the same as in the North, since these technologies are copied from the North.

The ratio of aggregate income in the South to that in the North can be written as:

$$\frac{Y'}{Y} = \kappa \left(\frac{\gamma \left((p'_L)^{(1-\beta)/\beta} N_L L' \right)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma) \left((p'_Z)^{(1-\beta)/\beta} N_Z Z' \right)^{\frac{\varepsilon-1}{\varepsilon}}}{\gamma \left(p_L^{(1-\beta)/\beta} N_L L \right)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma) \left(p_Z^{(1-\beta)/\beta} N_Z Z \right)^{\frac{\varepsilon-1}{\varepsilon}}} \right)^{\frac{\varepsilon}{\varepsilon-1}}. \qquad (25)$$

Simple differentiation and algebra show that (see Acemoglu, 2001):

$$\frac{\partial Y'/Y}{\partial N_Z/N_L} \propto \left(\frac{N_Z}{N_L} \right)^{-\frac{1}{\sigma}} \left[\left(\frac{Z'}{L'} \right)^{\frac{\sigma-1}{\sigma}} - \left(\frac{Z}{L} \right)^{\frac{\sigma-1}{\sigma}} \right]. \qquad (26)$$

Since $Z'/L' < Z/L$, this expression implies that when $\sigma > 1$, i.e., when the two factors are gross substitutes, an increase in N_Z/N_L raises the income gap between the LDCs and the North (i.e. reduces Y'/Y). This implies that as technologies produced by the North become more and more directed towards skilled workers, there will be a tendency for the income gap between the North and the South to increase. Intuitively, these new technologies are less and less appropriate to the needs of the LDCs who do not have a sufficient number of skilled workers to make best use of these technologies.[20]

What are the implications of directed technical change for income differences across countries? Equation (26) shows that a greater skill bias of technology will increase the income gap between rich and poor countries (between countries with different levels of skill abundance). Directed technical change implies that technologies developed in the

[20] In contrast, when $\sigma < 1$, an increase in N_Z/N_L narrows the income gap since the term in square brackets in (26) is now negative. However, when $\sigma < 1$, a lower N_Z/N_L increases the demand for Z relative to the demand for labor—that is, it corresponds to Z-biased technical change. Moreover, in this case, the North, which is more abundant in Z, will invest in technologies with lower N_Z/N_L than is appropriate for the South. In other words, exactly as in the case with $\sigma > 1$, the North will choose "too skill-biased" technologies, since with $\sigma < 1$, lower N_Z/N_L corresponds to greater skill bias. 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, i.e., $\sigma < 1$. See Acemoglu (2001) for a more detailed discussion.

U.S. or the OECD will be catered to their own needs, so typically more skill biased than would have been otherwise. As a result, directed technical change will increase the income gap across countries.

To gain further insight, consider a special case with $\sigma = 2$ (this was the case that was derived as an equilibrium of a more detailed model with many sectors in Acemoglu and Zilibotti, 2001), and suppose that $\delta = 0$ in terms of equation (15). In general, equation (25) is complicated because domestic prices in the North and the South differ depending on domestic factor supplies and world technology. However in the case with $\sigma = 2$, (25) simplifies to

$$\frac{Y'}{Y} = \kappa \left(\frac{\gamma^{\frac{\varepsilon}{2}} (L')^{1/2} + (1-\gamma)^{\frac{\varepsilon}{2}} \left(\frac{N_Z}{N_L} Z' \right)^{1/2}}{\gamma^{\frac{\varepsilon}{2}} (L)^{1/2} + (1-\gamma)^{\frac{\varepsilon}{2}} \left(\frac{N_Z}{N_L} Z \right)^{1/2}} \right)^2 . \tag{27}$$

This expression shows explicitly that the extent to which less developed country can take advantage of new technologies depends on the number of skilled workers. A country with few skilled workers will have difficulty adapting to the world technology, especially when this technology is highly biased towards skilled workers, i.e., when N_Z/N_L is high. To develop this point further, note that when $\sigma = 2$ and $\delta = 0$, the endogenous technology equation (17) implies

$$\frac{N_Z}{N_L} = \left(\frac{\eta_Z}{\eta_L} \right)^2 \left(\frac{1-\gamma}{\gamma} \right)^{\varepsilon} \left(\frac{Z}{L} \right)^2 . \tag{28}$$

Substituting this into (27) we obtain:

$$\frac{Y'}{Y} = \kappa \left(\frac{\gamma^{\frac{\varepsilon}{2}} (L')^{1/2} + \gamma_0 \frac{Z}{L} (Z')^{1/2}}{\gamma^{\frac{\varepsilon}{2}} (L)^{1/2} + \gamma_0 \frac{Z}{L} (Z)^{1/2}} \right)^2 ,$$

where γ_0 is a suitably defined constant. So the income gap between the North and the South will depend on the human capital gap between the North and the South, very much as in Nelson and Phelps (1966).

It is also instructive at this point to compare this approach based on endogenous technology choice to Nelson and Phelps' original contribution. Despite the similarities, there are also a number of important differences between this approach based on the idea of directed technical change and appropriate technology, and the Nelson and Phelps' approach. The first difference is more apparent than real. It may appear that in Nelson and Phelps (1966), human capital differences translate into technology differences, whereas in the simple version of the model of Acemoglu and Zilibotti (2001) which I outlined here, all countries have access to the same technology frontier, so there should not be differences in TFP/technology. This conclusion is not correct, however, because TFP differences are calculated as a residual from assuming a specific relationship between human capital, physical capital and output. In particular, it is typically assumed that $Y = K^{.33} (AH)^{0.67}$, and A, the TFP term, is calculated as the residual. In the presence of technology-skill mismatch as in my model with Fabrizio Zilibotti, the effect of human capital on output will be counted as part of TFP. Therefore, in both the model discussed here and in Nelson and Phelps (1966), human capital differences contribute to differences in output per worker or income per capita because of their interaction with technology adoption and the efficiency of technology use.

The second difference between the approaches is more interesting. Inspection of equation (30) from the Nelson-Phelps model given in the Appendix shows that human capital should matter more when the growth rate of the world technology, g, is greater. In contrast, (27) (or less transparently, (25)) shows that human capital should matter more when new technologies are more skill-biased. This is an interesting area to investigate. A recent paper by Easterly (2001) finds that low human capital countries lagged especially behind the richer countries over the past 25 years. Interestingly, this period has been one of slow world growth, so according to Nelson-Phelps' view, these low human capital countries should have benefited relative to the technology leaders. On the other hand, the past 25 years have also been characterized by very rapid skill-biased technical change, which suggests that, according to the model in Acemoglu and Zilibotti (2001), low human capital countries should have lagged further behind. But of course, many other factors could account for this pattern, and more work is required to get a better understanding of these issues.

6 Conclusion

This paper revisited the induced innovations literature of the 1960s to which Phelps was a major contributor (Drandakis and Phelps, 1965). This literature provided the first systematic study of the determinants of technical change, and also investigated the relationship between factor prices and technical change. I presented a modern reformulation of this literature based on the tools developed by the endogenous growth literature. This reformulation confirms many of the insights of the induced innovations literature, but reveals a new force, the market size effect: there will be more technical change directed at more abundant factors.

This modern reformulation sheds light on two recent debates: why is technical change often skill biased, and why has it become more skill biased during recent decades? And what is the role of human capital differences in accounting for income differences across countries? Interestingly, an application of this modern reformulation to these important debates also reiterates some of the insights of another important paper by Phelps, Nelson and Phelps (1966). Despite the similarities, there are also different implications of the Nelson-Phelps approach and those of an approach based on the direct technical change model. To investigate these differences empirically appears a fruitful area for future research.

7 Appendix: The Nelson-Phelps Model

I now briefly outline a version of the second model of Nelson and Phelps (1966).[21] Imagine that there is a world technology frontier, $T(t)$, advancing at an exogenous rate g, i.e.,

$$T(t) = T(0)\exp(gt).$$

Countries can benefit from this world technology by incorporating it into their production processes. But this is a human capital-intensive task. For example, a country needs highly skilled engineers to adapt world technologies to their conditions, to fill key positions in the implementation of these technologies and to train workers in the use of these new techniques. So Nelson and Phelps (1966) postulated that the technology of country j, $A_j(t)$, would evolve according to the differential equation

$$\frac{\dot{A}_j(t)}{A_j(t)} = \frac{\phi(h_j)(T(t) - A_j(t))}{A_j(t)}, \tag{29}$$

where h_j is the human capital in country j, which is assumed to be timing variant (see their equation (8')). This equation states that the farther a country is from the world technology frontier, the faster is its rate of progress. Most plausibly, this would be because there is more technology out there to be absorbed. But also $\phi'(h_j) > 0$ so that, the greater the human capital of a country is, the faster will this convergence be. Here h_j can be years of schooling, or the fraction of high skilled individuals, such as university graduates, or engineers, or some other feature of the human capital distribution.

The first implication of (29) is that

$$\frac{\partial^2 \dot{A}_j(t)/A_j(t)}{\partial T(t)\,\partial\phi(h_j)} > 0,$$

so that human capital becomes more valuable when technology is more advanced. This is the reason why the Nelson-Phelps approach has provided the foundation for a number of recent papers linking the demand for skills to the speed and extent of technical change (e.g., Greenwood and Yorukoglu, 1997, Galor and Maov, 2000, Aghion, Howitt and Violante, 2000).

[21] See also Shultz (1975) and Welch (1970).

Second, note that although equation (29) is in terms of technological progress, it does have a unique stable stationary distribution as long as $\phi(h_j) > 0$ for all countries. In the stationary state, all $A_j(t)$'s will grow at the same rate g, and this stationary cross-country distribution is given by

$$A_j(t) = \frac{\phi(h_j)}{g + \phi(h_j)} T(t). \tag{30}$$

Suppose now that output in each country is proportional to $A_j(t)$. Equation (30) then implies that countries with low human capital will be poor, because they will absorb less of the frontier technology. This effect is in addition to the direct productive contribution of human capital to output, and suggests that human capital differences across countries can be more important in causing income differences than calculations based on private returns to schooling might suggest.

8 References

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality" *Quarterly Journal of Economics*, CXIII (1998), 1055-1090.

Acemoglu, Daron, "Labor- And Capital-Augmenting Technical Change" NBER Working Paper no. 7544 (1999).

Acemoglu, Daron, "Directed Technical Change" NBER Working Paper 8287 (2001).

Acemoglu, Daron and Joshua D. Angrist, "How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws" *NBER Macro Annual 2000*, pages 9-71.

Acemoglu, Daron, Simon Johnson and James A. Robinson "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, forthcoming, (2001).

Acemoglu, Daron and Fabrizio Zilibotti, "Productivity Differences" (2001), *Quarterly Journal of Economics*, 116, pp.563-606.

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction" *Econometrica*, LX (1992), 323-351.

Aghion, Philippe and Peter Howitt, *Endogenous Growth Theory*, Cambridge, MA, MIT Press, 1998.

Aghion, Philippe, Peter Howitt and Gianluca Violante "Technology, Knowledge, and Inequality" Harvard mimeo, (2000).

Ahmad, Syed, "On The Theory of Induced Invention," *Economic Journal* LXXVI, (1966), 344-357.

Autor, David, Alan Krueger and Lawrence Katz, "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics*, CXIII (1998), 1169-1214.

Azariadis, Costas and Alan Drazen, "Threshold Externalities in Economic Development" *Quarterly Journal of Economics*, vol. 105, 501-26.

Barro, Robert J. and Jong-Wha Lee, "International Comparisons of Educational Attainment," *Journal of Monetary Economics* 32 (1993), 363-94.

Becker, Gary, Kevin M. Murphy and Robert Tamura, "Human Capital, Fertility and Economic Growth" *Journal of Political Economy*, vol. 98, (1990), S12-37.

Benhabib, Jess and Mark M. Spiegel, "The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data," *Journal of Monetary Economics* 34, (1994), 143-73.

Card, David E., "The Causal Effect of Education on Earnings," in O. Ashenfelter and D. Card, eds., *The Handbook of Labor Economics* Volume III, Amsterdam: Elsevier (1999).

Drandakis, E. and Edmund Phelps, "A Model of Induced Invention, Growth and Distribution" *Economic Journal*, 76 (1965), 823-840.

Easterly, William, "The Lost Decades" *Journal of Economic Growth* (2001).

Foster, Andrew and Mark Rosenzweig, "Technical Change in Human Capital Return and Investments:Evidence from the Green Revolution" *American Economic Review*, 86, (1996) 931-953.

Freeman, Richard "Demand For Education" Chapter 6 in Orley Ashenfelter and Richard Layard (editors) *Handbook of Labor Economics*, North Holland, Vol I (1986), 357-386.

Galor, Oded and Omer Maov "Ability Biased Technological Transition,Wage Inequality and Economic Growth" *Quarterly Journal of Economics*, CXV (2000), pp. 469-499.

Greenwood, Jeremy and Mehmet Yorukoglu "1974" *Carnegie-Rochester Conference Series on Public Policy*, XLVI (1997), pp. 49-95.

Grossman, Gene and Elhanan Helpman, *Innovation and Growth in the Global Economy,* Cambridge, MA, MIT Press, 1991.

Hall, Robert and Charles I. Jones, "Why Do Some Countries Produce So Much More Output per Worker Than Others?," *Quarterly Journal of Economics*, (1999).

Hendricks, Lutz "Cross-Country Income Differences: Technology Gaps of Human Capital Gaps? Evidence from Immigrants Earnings" forthcoming *American Economic Review.*

Hicks, John *The Theory of Wages*, Macmillan, London, 1932.

Jones, Charles I., "R & D-Based Models of Economic Growth," *Journal of Political*

Economy 103, (1995), 759-784.

Kennedy, Charles, "Induced Bias in Innovation and the Theory of Distribution" *Economic Journal,* LXXIV (1964), 541-547.

Kiley, Michael, "The Supply of Skilled Labor and Skill-Biased Technological Progress" 1999, *Economc Journal.*

Klenow, Peter J. and Andres Rodriquez-Clare, "The Neoclassical revival in Growth Economics: Has It Gone Too Far?," *NBER Macroeconomics Annual*, (1997), pages 73-103.

Lucas, Robert, "On the mechanics of economic development." *Journal of Monetary Economics*, 22 (1988), 3-42.

Mankiw, N. Gregory, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107, (1992), 407-37.

Nelson, Richard and Edmund Phelps. 1966. "Investment in Humans, Technological Diffusion and Economic Growth." *American Economic Association Papers and Proceedings.* 56, pp. 69-75.

Nordhaus, William; "Some Skeptical Thoughts on the Theory of Induced Innovation" *Quarterly Journal of Economics*, LXXXVII, (1973), 208-219.

Rauch, James E., "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities" *Journal of Urban Economics* 34, (1993), 380-400.

Romer, Paul M., "Increasing Returns and Long-Run Growth" *Journal of Political Economy* 94 (1986), 1002-1037.

Romer, Paul M., "Endogenous Technological Change" *Journal of Political Economy*, IIC (1990), S71-S102.

Salter W. E.G., *Productivity and Technical Change*, 2nd edition, Cambridge University Press, (1966) Cambridge, United Kingdom.

Samuelson, Paul, "A Theory of Induced Innovations Along Kennedy-Weisacker Lines" *Review of Economics and Statistics*, XLVII (1965), 444-464.

Schultz, Theodore. 1975. "The Value of the Ability to Deal with Disequalibria." *Jouranl of Economic Literature.* 13, pp. 827-846.

Segerstrom, P. S., T. Anant, and E. Dinopoulos, "A Schumpeterian Model of the

Product Life Cycle" *American Economic Review* 80 (1990), 1077-1092.

Stokey, Nancy, "Human Capital, Product Quality and Growth" *Quarterly Journal of Economics*, vol. 106, (1991), 587-616.

Stokey, Nancy, "R&D and Economic Growth" *Review of Economic Studies*, vol. 62, (1995) 469-490.

Summers, Lawrence and Alan Heston, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics*, 106, (1991), 327-368.

Trajtenberg, Manuel Rebecca Henderson, and Jaffe, Adam B. "Ivory Tower Vs. Corporate Lab: an Empirical Study of Basic Research and Appropriatability" NBER working paper 4146,1992.

Welch, Finis, "Education in Production," *Journal of Political Economy* 78 (1970), 312-327.

Young, Alwyn "Growth without Scale Effect" *Journal of Political Economy*, 106 (1998), 41-63.

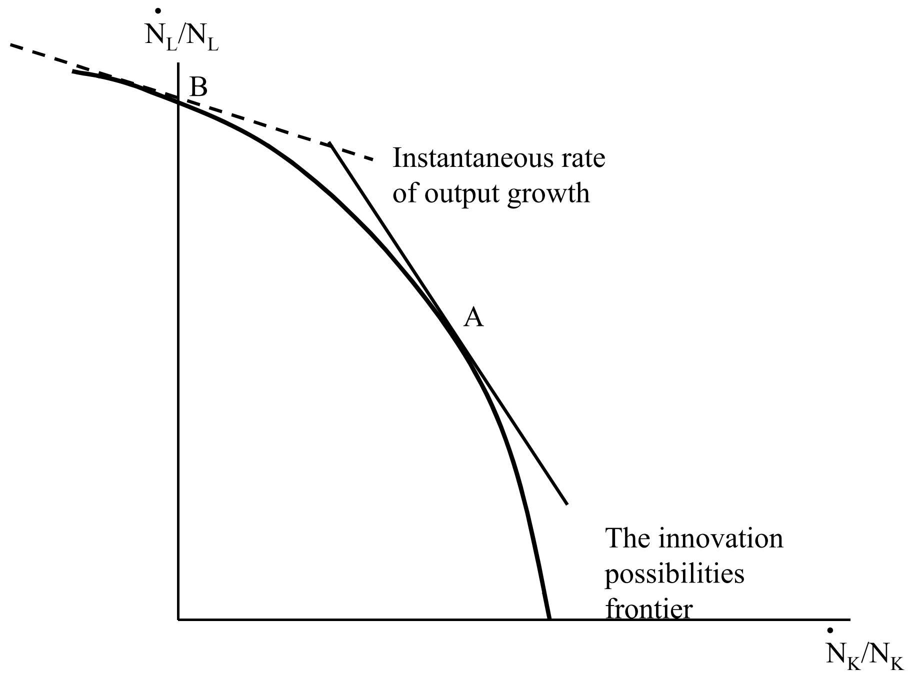


Figure 1: Equilibrium in the induced innovations model. Point B corresponds to the equilibrium with capital accumulation.

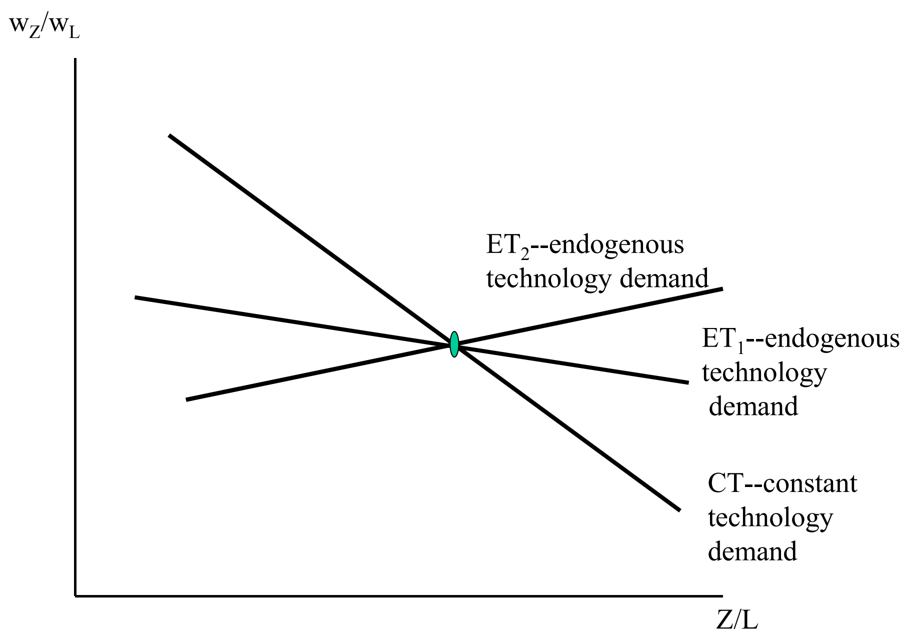
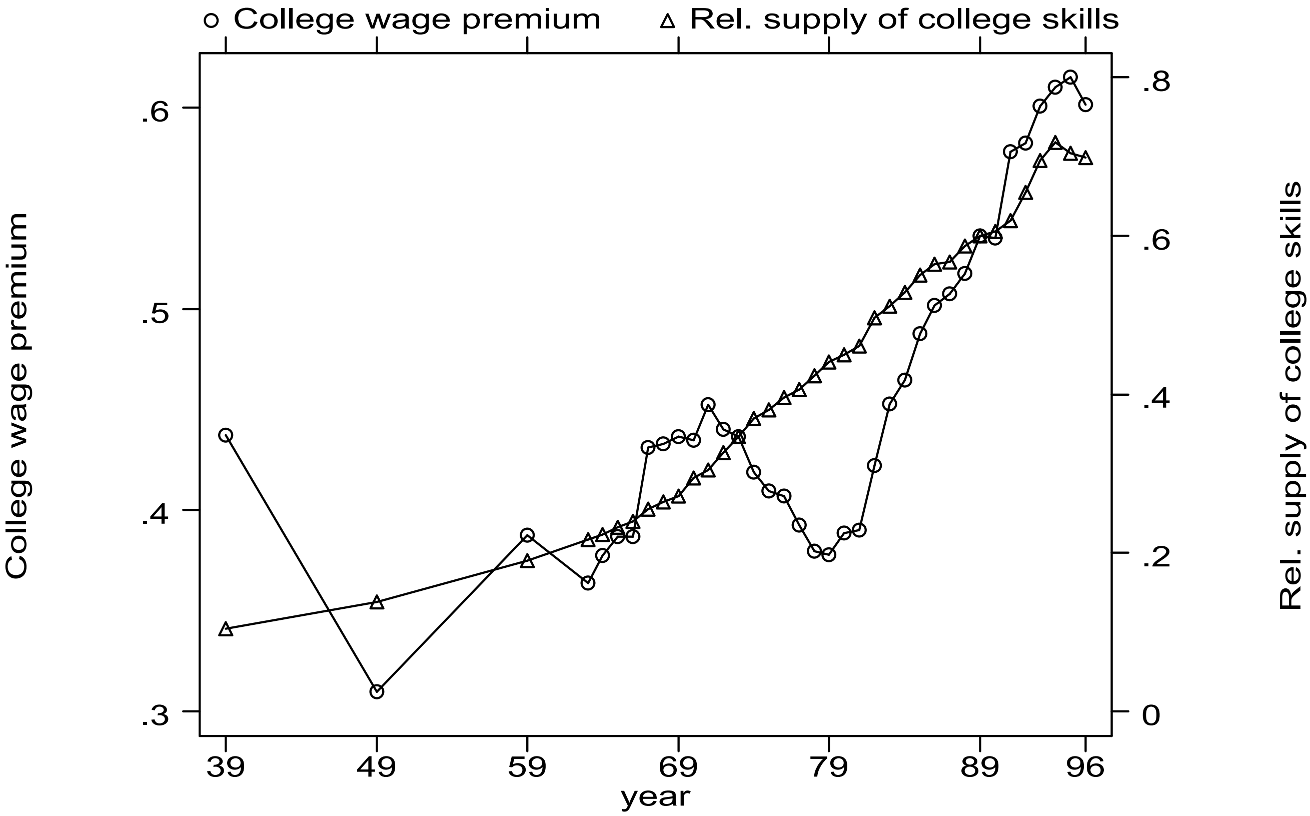


Figure 2: Constant-technology and endogenous-technology relative demand curves.



Relative Supply of College Skills and College Premium

Figure 3: Relative supply and relative demand for college skills in the U.S. labor market 1939-1996, author's calculations from Census and Current Population Surveys data.

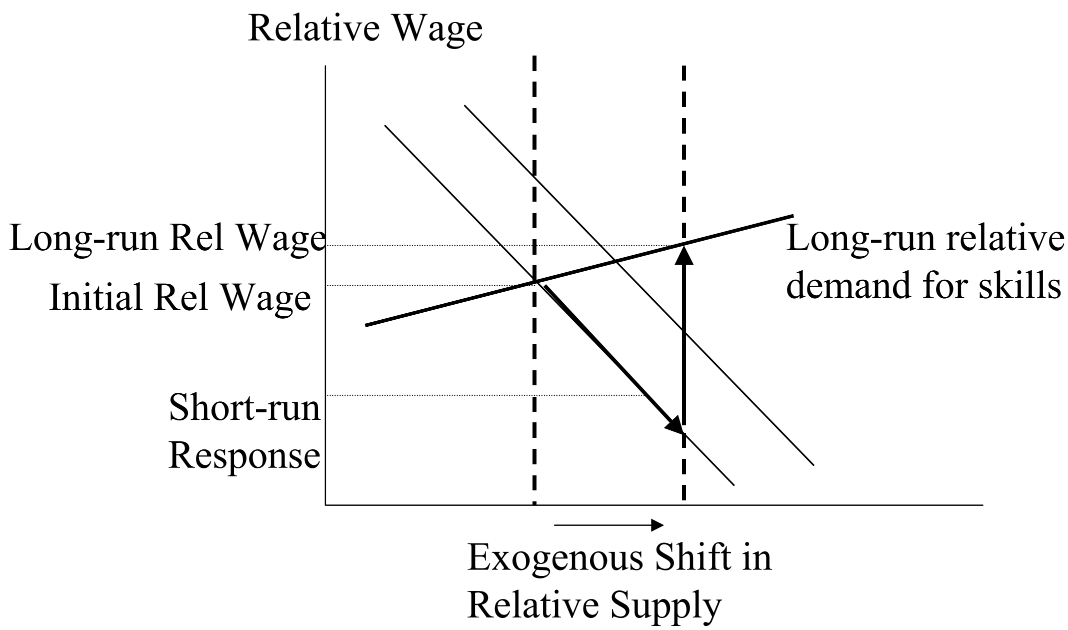


Figure 4: Short-run and long-run response of the skill premium to an increase in the supply of skills.

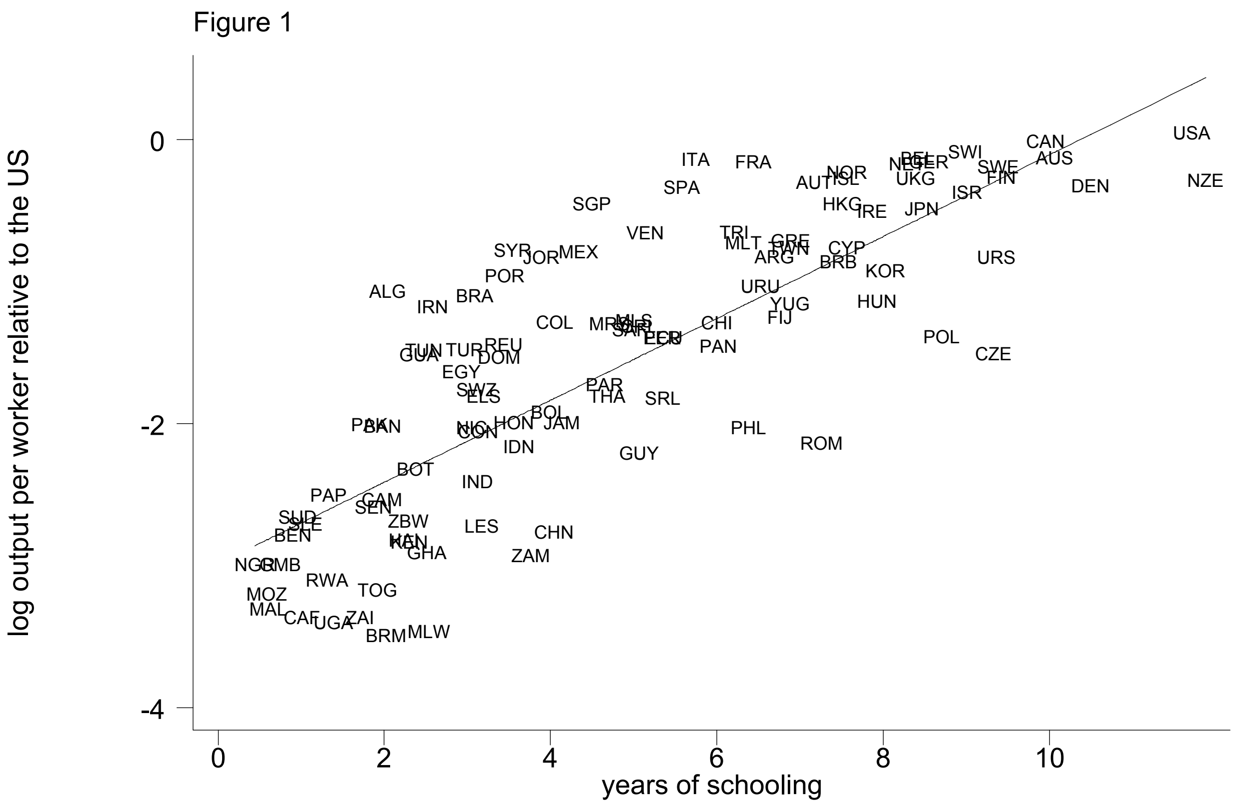


Figure 5: Cross-country relationship between years of schooling and logarithm of output per worker. Data from Barro and Lee (1993) and Hall and Jones (1999).

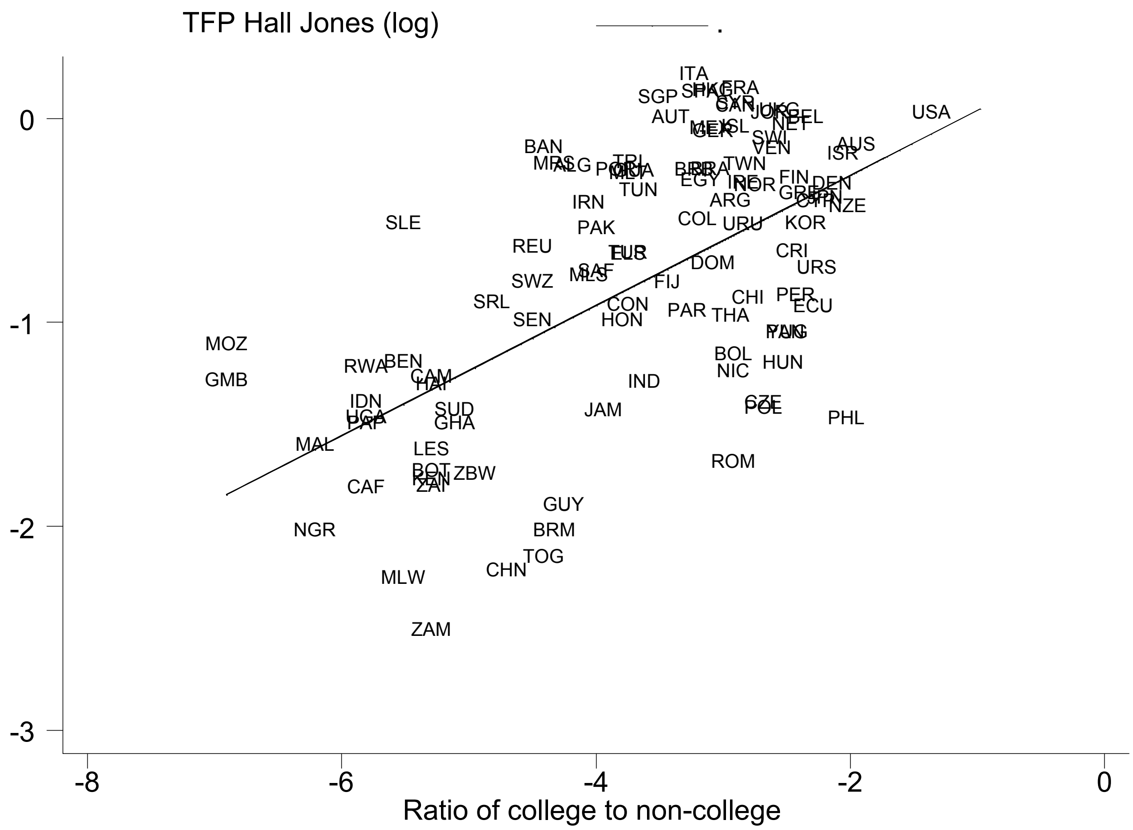


Figure 6: Cross-country relationship between the ratio of college graduates to non-college graduates and total factor productivity. Data from Barro and Lee (1993) and Hall and Jones (1999).