Review of Economic Dynamics **4,** 127–158 (2001)
doi:10.1006/redy.2000.0104, available online at http://www.idealibrary.com on IDEAL®

Schooling and Distortions in a Vintage Capital Model[1]

Xavier Mateos-Planas

Department of Economics, University of Southampton, Highfield SO17 1BJ, Southampton, United Kingdom
E-mail: fxmp@soton.ac.uk

Received January 21, 1999

This paper integrates the analysis of choices on education and on technology adoption to study international economic disparities. Two candidate explanations are considered: differences in distortions that affect the cost of technology adoption and differences in the effectiveness of schools. The implications of these two factors for differences in output per capita, educational attainment, and the age of technologies across countries are assessed in a vintage capital model with technology-specific learning-by-doing. Predictions are obtained for a parameterized economy that matches U.S. aggregate observations and evidence on learning. Differences in investment distortions produce plausible correlations only if the major role of education is to improve the ability to learn technologies. On the other hand, differences in school effectiveness produce plausible results only if the role of education is to provide a productive ability that is independent of learning. *Journal of Economic Literature* Classification Numbers: I2, O1, O3, O4.

1. INTRODUCTION

There are persistent differences in income per capita across countries. A major task of economic analysis is to identify the factor or set of factors that constitute ultimate causes of these differences. Some recent studies—which include Parente and Prescott (1994) and Jovanovic and

[1]This paper is based on Chapter 3 of my Ph.D. dissertation at Universitat Pompeu Fabra under the supervision of Ramon Marimon. My thanks go to him for useful advice as well as to Jordi Caballé, Antonio Ciccone, Javier Díaz-Jiménez, Stephen L. Parente, and Gilles Saint-Paul for useful discussions and criticism. In particular, S. Parente's comments have led to substantial improvements in various parts of the paper. The paper has also benefited from constructive comments by two referees and the editor Boyan Jovanovic. All remaining errors are my sole responsibility. This paper was presented at the SED 1999 Meeting in Sardinia. A previous version had been circulated under the title "Education, Technology Adoption and Productivity."

1094-2025/01 $35.00

Rob (1998)—focus on the role of factors that affect the adoption of technologies. In the growth literature, other papers—including Lucas (1988), Mankiw *et al.* (1992), and Grossman and Helpman (1991, Chapter 5)—have emphasized the importance of the incentives for education. This paper integrates the analysis of both education and technology adoption in the context of a vintage capital model with technology-specific learning-by-doing. The objective is to assess different factors as explanations of observed international disparities.

The paper focuses on the role of two such factors motivated by some recent empirical studies. The first factor is differences in policies that affect the resource cost of capital equipment that embodies new technology. Several papers such as Jones (1994), Chari *et al.* (1997), and Jovanovic and Rob (1998) find the price of capital to be negatively correlated with international income levels. The second factor is differences in the effectiveness of schools. Measures of school inputs—e.g., teacher/pupil ratio, government expenditures, number of school hours—in Barro and Lee (1997) are positively related with the level of development. The idea of school effectiveness here may also include the presence of policies and institutional distortions that affect educational outcomes for given measured inputs.

The importance of these two factors as explanations of international income differences is evaluated by studying the implied signs of the relationships between income per capita, educational attainment, and the age distribution of machines. More precisely, the criterion adopted in this paper is that, for a factor to qualify as a sensible explanation of disparities, the implied differences in output per capita must be positively related to differences in measures of educational attainment in population[2] and negatively related to differences in the average age of machines.[3] One objective of this paper is to explore whether and when the effectiveness of education and/or distortions that affect the cost of technology adoption pass this qualitative test.

[2] In 1960, years of schooling averaged 1.39 in the poorest third of countries, 2.91 in the middle third of countries, and 5.6 in the top third group of countries. The figures for 1985 are 2.18, 4.67, and 7.94 years of schooling, respectively. The figures for years of schooling are calculated from Barro and Lee (1993). Countries have been ranked according to figures on GDP per capita in Summers and Heston (1991, Mark 5.0).

[3] There are not aggregate measures of the age of capital available for a large set of countries. For the more developed economies, the figures reported in Wolff (1991) show that changes in economic leadership over the last century are related to changes in dominance in new capital vintages. On the other hand, there is microeconomic evidence—for example, Welch (1970), Bartel and Litchenberg (1987), and Foster and Rosenzweig (1996)—that education affects the choice of technology.

The key feature of the model is the choice by agents on the length of schooling and the sequence of subsequent technology adoptions in the presence of exogenous embodied technological change and technology-specific learning-by-doing. Two different roles for education are considered. One assumption is the *learning hypothesis*, according to which education improves the ability of economic agents to learn new technologies. This assumption is motivated by early works on human capital by Nelson and Phelps (1966), Welch (1970), and Schultz (1975) which emphasize that education has a value in a changing environment that requires the ability to adapt to (or to learn) new technologies. The other assumption postulates that education provides a productivity advantage which is independent of experience. It will be referred to as the *neutrality hypothesis*. This assumption is found in much of the growth literature, for example, Lucas (1988), Grosman and Helpman (1991, Chapter 5), and Bils and Klenow (1998).

Balanced-growth output comparisons are conducted for economies that differ in school effectiveness and investment distortions. The benchmark economy roughly matches U.S. aggregate observations and evidence on learning processes. The findings from the numerical exercise are as follows. Under the *learning hypothesis*, the model predicts positive correlations between per capita output, education, and a lower machine age only when economies differ in adoption costs. In this case, the response of the pattern of technology adoption is key to the predicted effects. Under the *neutrality* view, only variation in school effectiveness predicts the right signs for these correlations. In this case, the decisions on technology adoption play a mere supportive role. The results are informative as to mechanisms that may be important ingredients in a theory of international disparities.

In that the interaction between education and technology is analyzed to explain economic differences, this paper relates to a body of literature that includes Stokey (1991), Ciccone (1996), Keller (1996), Restuccia (1997), and Jovanovic (1998). In the present paper, the nature of this interaction is as follows. How much education individuals acquire depends on the expected pace of adoption of new technologies they will have to learn and operate after their education; on its part, the pattern of adoption of new technologies is influenced by the level of educational skills of the workers and managers that have to learn and operate those technologies. Here this interaction can be analyzed because the problem of adoption/replacement of technologies over time with technology-specific learning has been explicitly worked out. This differs from the other papers cited that do not analyze this type of problem and cannot, therefore, analyze this connection. They assume, rather than derive, some form of technology-skill complementarity that brings automatically together the skilled agents and the advanced technologies. In this sense, one contribution of the present analysis is to provide a model with stronger micro-foundations that allows us to examine how the relation between education and technology may be altered by policies and the role of education.

The model also shares features with recent analyses of the choice of replacement/adoption of technologies in dynamic settings in Parente (1994, 2000), Cooley *et al.* (1997), Greenwood and Yorukoglu (1996), Klenow (1998), and Jovanovic and Rob (1998). In Parente (1994, 2000), learning is not completely technology-specific and education is treated parametrically. Jovanovic and Rob (1998) do not consider learning-by-doing or education but endogenize the creation of new technologies. In Klenow (1998) there is no education either but there is variable labor intensity in production which affects learning-by-doing; here the technology is of fixed coefficients and learning depends on the passage of time. Cooley *et al.* (1997) introduce education, but it affects the creation of technologies rather than their adoption. In Greenwood and Yorukoglu (1996) firms hire skilled workers to accelerate learning, which resembles the *learning* assumption made here.

Of all these papers, only Jovanovic and Rob (1998) and Parente (2000) analyze sources of long-run disparities in international income levels. Both papers quantify the level effects of investment distortions which are found to be small, although somewhat larger in Parente (2000) where adoption of below-frontier technologies is allowed. While introducing the endogenous response of education, the present paper does not improve the quantitative implications found in these other papers. Parente (2000) also explores the level effects of parameters that are associated with the assumptions on education made in the present paper.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 analyzes the optimal choices on technology adoption and education. Section 4 defines and characterizes balanced-growth equilibria. Section 5 calibrates the model and reports steady-state comparisons associated with differences in adoption costs and school effectiveness under both assumptions on the role of education. Conclusions and remarks about directions for future research in Section 6 conclude the paper.

2. THE MODEL

Demographics. The economy is populated by a continuum of overlapping generations of agents. Each agent faces a constant probability of death per unit of time, p, which is independent of age. The size of population is normalized to one by letting the size of each new cohort be equal to p. Then $pe^{-p(t-\tau)}$ is the measure of agents born at τ that are alive at time t. In the first part of his life, an agent goes to school for a length of time s. In the remaining part of his life, he produces output by operating a firm.

Output Technology. There is a single homogeneous good that can be used for consumption or investment. Each firm produces output using one machine. The flow of output of a firm depends on the quality of the machine in use, on the agent's technology-specific expertise, and on the agent's general skill. The quality of the machine is given by the technology embodied in it and I index technologies over the positive real line by a. Expertise in a technology, q, can take on only two values, 1 and δ, with $\delta > 1$. General skill is represented by a positive real number h. Output of a firm that operates a machine of quality a with technology-specific expertise q and general skill h is

$$h \cdot q \cdot a, \tag{1}$$

with $q \in \{1, \delta\}$ and $a, h \in R_+$.

At any instant of time, a firm may either switch to a more advanced technology or continue to use the present one. I call *technology adoption* the decision to operate a new technology by replacing the current machine with another of different quality.

If the firm's experience in the use of its current technology is shorter than a period of length μ, its level of expertise in this technology is 1. Thereafter, its level of expertise in this technology increases to δ. Figure 1 depicts the typical path of technology-specific expertise.

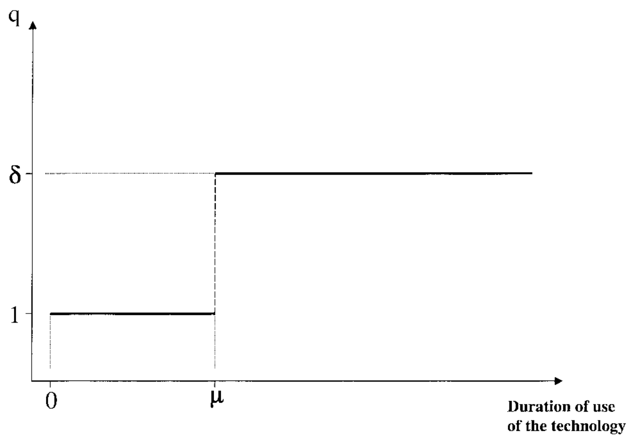

FIG. 1. Evolution of technology-specific expertise.

This learning-by-doing is technology-specific. Thus if the firm decides to switch technologies, no part of the expertise in the previous technology can be carried over to the new one.[4]

The upper bound on the technologies that can be used by any firm at time t is denoted by $A(t)$. This frontier technology grows at a constant and exogenous rate γ over time.

Switching to a technology a requires an expenditure of $\pi \cdot a$ units of output in the machine embodying that technology. The parameter π is a policy parameter that affects the resource cost of a machine.

Education. The length of time, μ, that it takes for a firm to accumulate expertise in a technology is determined by the schooling of the agent that runs it. The years of schooling of an agent is denoted by s. The value of μ depends negatively on the time an agent spends in school. In particular, μ is assumed to be a nonincreasing function of s, $\mu(s)$. More specifically,

$$\mu = \mu(s) \equiv \mu_0 e^{-\lambda s}, \tag{2}$$

with μ_0 and λ being nonnegative constants. This view of schooling is referred to as the *learning hypothesis*.

Schooling can also affect an agent's general level of skill, h. I assume h to be a nondecreasing function, $h(s)$, of s. Specifically,

$$h = h(s) \equiv h_0 s^{\beta}, \tag{3}$$

with h_0 and β being nonnegative constants. This view of schooling is referred to as the *neutrality hypothesis*.[5]

[4] This is a simplification with respect to Parente (1994) and Jovanovic and Nyarko (1996) who allow for different degrees of transferability of knowledge across vintages. In terms of Jovanovic and Nyarko, I am in case $\alpha = 0$ (in their Eq. (3)).

[5] Rosenzweig (1995) identifies two roles of education in Jovanovic and Nyarko's (1996) Bayesian learning target-input model. First, the more educated may obtain more information from each use of the technology as reflected in a lower variance of the optimal target which speeds up the gain in the precision of optimal decisions. This is similar to my *learning hypothesis* that education lowers μ. Second, education may also improve access to information sources as reflected in a lower variance of the prior about the input target which improves the initial precision of decisions in a new technology. This *information prior* assumption does not correspond to my *neutrality hypothesis*. It would be picked up in my model by letting education raise h—just like under the *neutrality* assumption—and, in addition, replacing δ by δ/h. Intuitively, if the initial productivity in a new technology is higher, then the gain from experience is smaller for given full-potential productivity $a\delta$.

I think that my *neutrality* assumption could be accommodated in a version of Jovanovic and Nyarko's information setup where the posterior variance is bounded above zero (possibly introducing an additional source of noise) and education reduces this lower bound as well as the prior variance. To my knowledge, this sort of analysis has not yet been carried out and I have to accept that this hypothesis is less well founded on an information-theoretic model but seems to capture the practice in part of the literature on human capital.

I will interpret λ and β as measuring the effectiveness of individual time spent in the classroom. This school effectiveness may certainly depend on the quality and quantity of educational inputs other than individual time. But it may also include policy distortions and institutional factors that spur or hamper the acquisition of the relevant skills. For instance, national educational programs that emphasize contents not directly usable in production will have a smaller β; schemes that emphasize information acquisition rather than analytical skills and critical awareness will be associated with smaller λ.

Preferences, Markets, and the Distribution of Wealth. An agent belonging to cohort τ has preferences defined over lifetime streams of consumption. Because of lifetime uncertainty, utility is evaluated in expected terms and is represented by

$$E\left[\frac{1}{1-\sigma}\int_{\tau}^{\tau+T} e^{-\rho(t-\tau)} c(t,\tau)^{1-\sigma}\,dt\right],$$

where T is time-until-death, $c(t,\tau)$ is the flow of consumption to an agent of cohort τ at time t, ρ is the subjective time-discount rate, and σ is the relative rate of risk aversion. Since the probability of death, p, is constant, T is an exponential random variable with density pe^{-pT} and utility can be rewritten as

$$\frac{1}{1-\sigma}\int_{\tau}^{\infty} e^{-(\rho+p)(t-\tau)} c(t,\tau)^{1-\sigma}\,dt. \qquad (4)$$

There is a perfect capital market where agents can borrow and lend. There is a market for insurance where insurance companies make premium payments to the living in exchange for receipt of their assets in the event they die. Free-entry in this market implies that the insurance premium equals p per unit of time. When an agent dies, the firm she is operating is dissolved and there is no market for discontinued firms.

The agent is assumed to have a wealth given by the lifetime present value of per capita output minus adoption costs. The correct interpretation is to think of the agent as belonging to an extended family where all members of all ages share overall family output at every point in time.[6]

[6] This is a simplifying assumption that facilitates exposition. Another possible assumption is that an agent's wealth is given by the present value of the firm he or she operates. Section 5 shows that this more natural assumption has no substantive effect.

3. OPTIMAL INDIVIDUAL CHOICES

The objective of an agent is to maximize the utility function in Eq. (4) taking prices as given. Since I focus on balanced-growth equilibria, I consider a constant interest rate r. This problem has two parts. The first part is that the agent maximizes the present lifetime value of the firm he operates. This choice maximizes the agent's wealth since it maximizes its contribution to the present value of net output per capita. Toward this end, an agent of cohort τ has two choices to make. First, he chooses how long to attend school, s. Second, after school has been completed, the agent decides which technologies to use at every instant over his productive life. The choices that maximize a firm's present value can be found in two steps. First, given a choice s, and hence μ and h, it is possible to solve for the optimal pattern of technology adoption. Having obtained the optimal technology adoption decisions as a function of s, it is possible to find the optimal schooling level that maximizes the present value of the firm. The second part of the agent's problem is that the agent, as a consumer, chooses a path for consumption that maximizes utility subject to the constraint that the discounted present value of consumption does not exceed his wealth.

In this section, I characterize the optimal technology adoption and schooling decisions of an agent born at time τ. Section 3.1 shows that optimal technology adoption results in a sequence of evenly spaced dates at which the firm switches to the frontier technology and stays there until the next upgrade. The effect of parameters on the frequency of technology upgrades is analyzed. The optimal schooling choice can then be determined. Section 3.2 analyzes this decision and highlights the mechanisms underlying the response of schooling and technology adoption to changes in investment distortions and school effectiveness under two views about the role of education.

3.1. Technology Adoption

I begin by assuming that the agent has already chosen s and so the values of μ and h have already been determined. Before describing this problem formally I define a feasible adoption plan. Exposition is simplified by noting that the assumption $\pi > 0$ rules out the optimality of plans characterized by continuous switching of technologies. Thus, without any loss, I ignore this possibility in the following definition.

DEFINITION 1. Given the path for the frontier technology $A(t)$, a *feasible adoption plan* for an agent born at time τ with schooling s is defined by:

(i) A sequence $\{x_j\}$ representing the dates at which each jth adoption occurs for $j = 1, 2, \ldots, J$, such that $x_j \geq \tau + s$ and $x_{j+1} > x_j$.

The number of adoptions, J, may be either a positive integer, in which case I adopt the convention that $x_{J+1} = +\infty$, or infinity.

(ii) A path for technologies represented by a function $a(t, \tau)$ defined for $t \geq \tau + s$ such that $a(t, \tau) \leq A(t)$, and $a(t, \tau)$ constant for $t \in (x_j, x_{j+1})$ all $j = 1, 2, \ldots, J$.

The technology in (1) and the structure of learning imply that a feasible adoption plan generates the following path for the firm's output, $y(t, \tau)$,

$$y(t,\tau) = \begin{cases} h \cdot a(x_j, \tau) & \text{if } t \in \left[x_j, \min\{x_{j+1}, x_j + \mu\}\right) \\ h \cdot \delta \cdot a(x_j, \tau) & \text{if } t \in \left[\min\{x_{j+1}, x_j + \mu\}, x_{j+1}\right) \end{cases} \tag{5}$$

Now the problem of a firm consists of maximizing the present value of output net of adoption costs by choosing a feasible adoption plan or making no adoption at all. To characterize the solution it is convenient to set up the objective more formally.

The existence of an insurance premium implies that the agent discounts values at the rate $r + p$. Define by m_j the length of the period over which the jth technology is operated in a feasible adoption plan, i.e., $m_j \equiv x_{j+1} - x_j$. The present discounted value as of date x_j of net output accruing between x_j and x_{j+1} can then be expressed as $h \cdot a(x_j, \tau) \cdot W(m_j)$ where

$$\begin{aligned} W(m_j) &= \frac{1}{a(x_j,\tau)\cdot h} \cdot \left[\int_{x_j}^{x_j+m_j} e^{-(r+p)(t-x_j)} y(t,\tau)\, dt - \pi \cdot a(x_j,\tau)\right] \\ &= \begin{cases} \dfrac{1}{r+p}\left[1 - e^{-(r+p)m_j}\right] - \dfrac{\pi}{h} & \text{if } m_j < \mu \\ \dfrac{1}{r+p}\left[1 + (\delta - 1)e^{-(r+p)\mu} - \delta e^{-(r+p)m_j}\right] - \dfrac{\pi}{h} & \text{if } m_j \geq \mu \end{cases} \end{aligned} \tag{6}$$

is an increasing, bounded, continuous and piece-wise differentiable function of m_j. Thus the present value as of date $\tau + s$ associated with a feasible adoption plan can now be expressed as the following discounted sum of net values accruing from the adoption and use of different technologies,

$$h \cdot \sum_{j=1}^{J} W(x_{j+1} - x_j) \cdot a(x_j, \tau) \cdot e^{-(r+p)(x_j - (\tau + s))} \tag{7}$$

The problem of the firm is to choose the adoption plan that maximizes the value of expression (7) if it is positive or else remain inactive and produce zero. I will characterize the solution when the size of investment distortions is not too large and the frontier technology does not advance too fast. This is the case when the two following conditions hold.

C1:

$$\lim_{m \to +\infty} W(m) = (1 + (\delta - 1)\exp(-(r+p)\mu))/(r+p) - \pi/h > 0.$$

C2: $r + p - \gamma > 0.$

Assuming C1 rules out inactivity as an optimal choice, C2 then restricts the class of feasible adoption plans that are optimal. Proposition 1 states this result more precisely.[7]

PROPOSITION 1. *If conditions* $C1$ *and* $C2$ *hold then*:

a. *An optimal choice involves a feasible adoption plan with adoption of the frontier technology, i.e.,* $a(x_j, \tau) = A(x_j)$ *all* $j = 1, 2, \ldots, J$.

b. *In an optimal feasible adoption plan the first adoption occurs at* $\tau + s$, *i.e.,* $x_1 = \tau + s$.

Since the technology to be adopted when upgrading is always the frontier, solving for the optimal timing of adoptions will complete the characterization of the adoption choice. By Proposition 1 the objective in (7) can be cast in recursive form. This leads to the following result.

PROPOSITION 2. *Assume* $C1$ *and* $C2$ *hold. In an optimal adoption plan* $x_{j+1} - x_j = m$ *all* $j = 1, 2, \ldots, J$ *where* $m \in R_+$ *solves,*

$$V = \max_m \{W(m) + e^{-(r+p-\gamma)m} V\}, \tag{8}$$

and the resulting V is such that

$$A(\tau + s) \cdot h \cdot V \tag{9}$$

is the value of the expression in (7) *in an optimal plan.*

This proposition says that optimal upgrades occur at a constant frequency because the firm faces the very same problem (8) any time it has to decide how long to hold out for a new machine. Still, the number of adoptions, J, can be either 1 [if $m = +\infty$] or infinity [if $m < +\infty$]. The solution also delivers the value V which is $W(m)/(1 - \exp(-(r + p -$

[7] With C1 but not C2, it is optimal to wait infinity to start producing. Without C1, it is optimal to produce zero. Propositions 1, 2, and 3 are proved in Appendix A.

$\gamma)m))$ when evaluated at the optimal m. Then, the maximized value of the objective (7) can be calculated as (9).

I now proceed to characterize the solution to problem (8). This problem is a one-stage optimization program. The only difficulty arises because the assumption of learning as a step function creates a discontinuity in the derivative of the objective at $m = \mu$. Nonetheless, the optimal m can still be characterized as a solution to a first-order condition. Proposition 3 establishes this result as well as the procedure to find the solution.

PROPOSITION 3. *Assume $C1$ and $C2$ hold. Then:*

a. *A solution to* (8) *implies $m < +\infty$ and must satisfy*

$$\Delta(m) \equiv W'(m) - (r + p - \gamma)\frac{e^{-(r+p-\gamma)m}}{1 - e^{-(r+p-\gamma)m}}W(m) = 0. \quad (10)$$

b. *If* $\lim_{\mu^-} \Delta(m) > 0$ *then the solution to* (8) *is the unique root of* (10) *with $m > \mu$. If* $\lim_{\mu^-} \Delta(m) < 0$ *and* $\lim_{\mu^+} \Delta(m) < 0$ *then the solution to* (8) *is the unique root of* (10) *with $m < \mu$. If* $\lim_{\mu^-} \Delta(m) < 0$ *and* $\lim_{\mu^+} \Delta(m) > 0$ *then there are two roots of* (10), *one on each side of μ. The solution to problem* (8) *is the one root that yields the highest value of* $W(m)/(1 - \exp(-(r + p - \gamma)m))$.

According to this proposition, solving (8) only requires finding the values of m that solve the first-order condition (10). The procedure in part b indicates that there are at most two such values and exactly determines when the solution is to be found for values that are smaller than μ, larger than μ, or both. To understand the statements, it helps to depict $\Delta(m)$ as a piece-wise concave function that starts positive, eventually becomes monotonically decreasing, and exhibits a positive right jump at μ where, nevertheless, the slope remains continuous.

I will now use the first-order condition (10) to shed some light on the determinants of the time between adoptions m. Proposition 3b recognizes that a change in parameters may induce a discontinuous response of m if it implies a move from a plan where learning occurs [i.e., $m > \mu$] to a plan where more frequent upgrades imply that learning does not occur [i.e., $m < \mu$], or the other way around. This notwithstanding, here I will focus on changes that produce a local response when $m > \mu$. It helps intuition to use (6) to rearrange (10) as

$$\delta = (r + p - \gamma)\frac{e^{\gamma m}}{1 - e^{-(r+p-\gamma)m}}W(m) \quad (11)$$

The two sides of this equation contain, respectively, the normalized values for the marginal benefit and marginal cost of m. As I am looking at a situation where $m > \mu$, the marginal benefit of m consists of the current flow of normalized output on the current technology, δ. The marginal cost consists of the discounted foregone present value that would arise on switching to a new machine whose quality exceeds the current one's by a factor $\exp(\gamma m)$. It is optimal to pick m so that these two values are equalized.

For the purposes of this paper, it is important to understand how the pattern of technology adoption is affected by investment distortions, π, and education, h, and μ. These parameters show up within $W(m)$ in condition (11). Furthermore, by (6), the role of h and π for technology adoption is entirely captured through the ratio π/h. In other words, general skill dampens the presence of investment distortions. One can figure out the effect of slower learning-by-doing, lower general skill, or higher investment distortions on optimal m. By reducing $W(m)$, both larger μ and larger π/h produce a downward shift of the marginal cost leading to larger m, which implies less frequent technology adoptions.

3.2. Schooling Choice

The length of schooling attendance, s, is decided optimally taking into account its effects on the value of the firm which, in turn, depends on the response of technology adoption choices. To analyze this problem, it is useful to represent explicitly the dependence of the value of the firm on μ and π/h. Let $V(\mu, \pi/h)$ denote the value that solves (8) as a function of these two variables. Then, by (9) in Proposition 2, the value of the firm can be written as $A(\tau + s) \cdot h \cdot V(\mu, \pi/h)$. The agent's goal is to maximize the present discounted value of this expression as of date τ. The technology of education described in (2) and (3) and the assumption that $A(t)$ grows at rate γ therefore imply that the agent solves

$$\max_{s \geq 0} e^{-(r+p-\gamma)s} h(s) V(\mu(s), \pi/h(s)) \tag{12}$$

$$\text{s.t.} \quad \mu(s) = \mu_0 e^{-\lambda s}$$

$$h(s) = h_0 s^{\beta}$$

Since Proposition 3 implies that m in program (8) can be bounded, application of the maximum principle implies that the objective of this program is continuous in s. Without an upper bound on s, existence of a solution cannot be generally established though.

Next, the specialized problems that arise under two specific assumptions on education are considered. The analysis will focus on each case separately to draw the implications of changes in investment distortions, π, and measures of the effectiveness of schooling, λ and β, for the optimal level of schooling s that solves (12).

3.2.1. *The Learning Hypothesis*

In this case, the parameter β is assumed to be zero so that h is not affected by s. Without loss of generality, I normalize $h = 1$ by setting $h_0 = 1$. Assume that a solution to (12) is interior, implies $m > \mu$, and is characterized by a first-order condition which must read

$$\frac{dV(\mu(s), \pi/h)}{d\mu}\mu'(s) = (r + p - \gamma)V(\mu(s), \pi/h) \tag{13}$$

which equates marginal benefit (left) and marginal cost (right) of s.[8]

Now, consider the effect of an increase in π on this equation. The marginal cost on the right, $V(\mu, \pi/h)$, is reduced, which tends to dictate higher schooling s. This is an opportunity-cost effect. On the other hand, m goes up so adoptions of technology become less frequent. This reduces the marginal benefit of education on the left-hand side of (13): less frequent adoptions make the ability to learn new technologies provided by education less valuable. This tends to reduce the optimal time spent at school. Thus investment distortions may have a negative impact on educational attainment only because of this latter effect.

On its part, a lower quality of education, λ, also produces effects on both sides of (13). Again, the opportunity cost on the right-hand side is reduced, which works in favor of longer schooling. On the left-hand side of (13), the marginal benefit falls as larger $\mu(s)$ as well as the consequent rise of m reduce the marginal gain of speeding up learning through schooling (the $-dV/d\mu$ term). This tends to reduce schooling. Finally, there is the effect from the direct impact on μ' that may have either sign on the marginal benefit.[9]

3.2.2. *The Neutrality Hypothesis*

In this case, the parameter λ is assumed to be zero so that μ is not affected by s. As before, assume that a solution to (12) is interior, implies

[8] The discussion that follows relies on the following derivation. Use the envelope theorem [i.e., Eq. (11)] to show that the first left-hand side term in (13) can be written as

$$\frac{dV(\mu(s), \pi/h)}{d\mu} = -(\delta - 1)\frac{e^{-(r+p)\mu(s)}}{1 - e^{-(r+p-\gamma)m}}.$$

[9] If lower school effectiveness was represented by larger μ_0 rather than smaller λ, then $-\mu'$ would go up, which works for larger s.

$m > \mu$, and is characterized by a first-order condition which must read

$$-\frac{dV(\mu, \pi/h(s))}{d(\pi/h(s))}\pi\frac{h'(s)}{h(s)} + h'(s)V(\mu, \pi/h(s))$$
$$= (r + p - \gamma)h(s)V(\mu, \pi/h(s)) \qquad (14)$$

which equates marginal benefit (left) and marginal cost (right) of s. The marginal benefit includes the increase in disembodied productivity h as well as the effect on the value of the firm through the term π/h.[10]

Consider a rise in π. Ignoring for the moment the response of the first term on the left-hand side, the effect on $V(\mu, \pi/h(s))$ shifts both marginal cost and marginal benefit downward, but the absolute shift of marginal cost is bigger, which tends to increase s.[11] But the first component of the marginal benefit will also change. In particular, the fact that m increases, and hence adoptions become less frequent, brings about a reduction of the marginal benefit of reducing the adoption cost term, π/h, through education. This tends to dictate lower education.

A reduction in the effectiveness of schools also reduces the second term on the marginal benefit side as well as the marginal cost. The first term of the marginal benefit will also decrease except for the possibility that the rise in the marginal return of education due to larger π/h is quantitatively substantial.

4. EQUILIBRIUM

The exercises conducted in the previous section suggest that when the choice of technology is endogenized, the response of schooling to changes in various parameters cannot generally be determined. I now proceed with the characterization of a balanced-growth path equilibrium. I will then be able to pin down reasonable parameter values and therefore obtain

[10] For the discussion that follows it is useful to calculate

$$-\frac{dV(\mu, \pi/h(s))}{d(\pi/h(s))}\pi\frac{h'(s)}{h(s)} = h'(s)\left[\frac{1}{1 - e^{-(r+p-\gamma)m}}\frac{\pi}{h(s)} + V(\mu, \pi/h(s))\right].$$

Then the derivative of this expression with respect to π is

$$-\frac{r + p + \gamma}{1 - e^{-(r+p-\gamma)m}}h'(s)\frac{\pi}{h(s)}\frac{dm}{d\pi} < 0.$$

[11] If V did not depend on h, then the two effects would exactly cancel out.

predictions for differences in output per capita, educational attainment, and the age of capital produced by differences in the level of investment distortions and the effectiveness of schools.

I study balanced-growth equilibrium paths along which all aggregate quantities grow at a constant rate and the interest rate is constant. I define such an equilibrium as follows.

DEFINITION 2. Given $\sigma, \rho, p, \delta, \gamma, \pi, \mu(\cdot), h(\cdot)$, and a distribution of wealth $Z(\tau)$, a *competitive balanced growth equilibrium* is a constant m denoting the spacing between adoptions, a constant level of education s, a distribution of output per firm $y(t,\tau)$, a path for per capita output $y(t)$, a distribution of individual consumption $c(t,\tau)$, a path for per capita consumption $c(t)$, and a constant interest rate r, such that:

1. Taking r, s, $\mu(s)$, and $h(s)$ as given, $y(t,\tau)$ is determined by (5) for an optimal technology adoption plan by a firm of vintage $\tau + s$ characterized in Propositions 1 through 3, with m being the optimal length of time between adoptions.

2. Taking r as given, s is optimal in that it solves (12).

3. Given r, $c(t,\tau)$ maximizes utility of the agent of cohort τ in (4) subject to the constraint that the present value of consumption cannot exceed his wealth $Z(\tau)$,

$$\int_{\tau}^{\infty} c(t,\tau) e^{-(r+p)(t-\tau)}\, dt = Z(\tau).$$

4. Market clearing and aggregate consistency. Aggregate consumption equals aggregate output minus investment,[12]

$$c(t) = y(t) - \frac{pe^{-ps}}{1 - e^{-pm}} \pi A(t), \tag{15}$$

with

$$c(t) = \int_{-\infty}^{t} pe^{-p(t-\tau)} c(t,\tau)\, d\tau,$$

and

$$y(t) = \int_{-\infty}^{t-s} pe^{-p(t-\tau)} y(t,\tau)\, d\tau.$$

[12] The demographic assumptions made imply that $p\exp(-pu)$ is the size of population aged u. Therefore the measure of individuals adopting the frontier technology at any instant along a balanced-growth path is $p\exp(-ps) + p\exp(-p(s+m)) + p\exp(-p(s+2m)) + p\exp(-p(s+3m)) + \cdots = [p\exp(-ps)]/[1-\exp(-pm)]$. Investment is this number times the cost of a new machine $\pi A(t)$.

5. $c(t)$ and $y(t)$ grow at a constant rate.
6. Conditions C1 and C2 hold.

Points 1 and 2 in the definition require that agents behave optimally. By point 6 and the results in Section 3, the individual path of output then results from writing (5) for a feasible plan with $x_{j+1} - x_j = m$ all j and $a(x_j, \tau) = A(x_j)$. This yields,

$$y(t,\tau) = \begin{cases} h(s)A(\tau + s + (j-1)m) \\ \quad \text{if } \tau + s + (j-1)m \le t < \tau + s + (j-1)m + \mu \\ h(s)\delta A(\tau + s + (j-1)m) \\ \quad \text{if } \tau + s + (j-1)m + \mu \le t < \tau + s + jm \end{cases}$$

where m and s solve (8) and (12), respectively. The formula for aggregate output in point 4, which takes into account the age structure of population, can then be developed to obtain

$$\begin{aligned} y(t) &= \sum_{j=1}^{\infty} \int_{t-s-jm}^{t-s-(j-1)m} pe^{-p(t-\tau)} y(t,\tau)\, d\tau \\ &= A(t) \frac{p}{\gamma + p} \frac{e^{-ps}}{1 - e^{-pm}} R(m,s) \end{aligned} \tag{16}$$

with

$$R(m,s) = \begin{cases} 1 - e^{-(\gamma+p)m} & \text{if } m < \mu(s) \\ (1 - e^{-(\gamma+p)\mu(s)}) + \delta(e^{-(\gamma+p)\mu(s)} - e^{-(\gamma+p)m}) & \text{if } m > \mu(s) \end{cases}$$

Hence, aggregate output grows at the same rate as the frontier technology γ and, by the market-clearing condition (15) in point 4, so does aggregate consumption and, hence, point 5 in the definition holds. It remains to find for which interest rate r, if any, per capita consumption, $c(t)$, calculated by aggregation of individual consumption, $c(t, \tau)$, is consistent with the market clearing condition in (15). With the preferences described by (4) and the assumption of free borrowing and lending, point 3 implies that individual consumption, $c(t, \tau)$, grows at a constant rate γ_c with

$$\gamma_c = \frac{r - \rho}{\sigma}. \tag{17}$$

To add up consumption of the different agents, it is necessary to be specific about the distribution of wealth across cohorts $Z(\tau)$. The assumption of the model is that $Z(\tau)$ is the present value of net output per capita over the agent's lifetime,

$$Z(\tau) = \int_{\tau}^{\infty}\left[y(t) - \frac{pe^{-ps}}{1-e^{-pm}}\pi A(t)\right]e^{-(r+p)(t-\tau)}\,dt. \tag{18}$$

Using the intertemporal budget constraint in point 3 and constant growth rates γ_c and γ for individual consumption and aggregate net output, respectively, the condition for aggregate consumption in point 4 can be rewritten as

$$c(t) = p\frac{r+p-\gamma_c}{r+p-\gamma}\,\frac{1}{p-\gamma_c+\gamma}\left[y(t) - \frac{pe^{-ps}}{1-e^{-pm}}\pi A(t)\right]. \tag{19}$$

In a balanced-growth equilibrium, the market clearing condition (15) must be satisfied for consumption as given in (19) and growth of individual consumption as in (17). Simple inspection indicates this holds for $r = \gamma$ and for $\gamma_c = \gamma$. This leads to the following proposition.

PROPOSITION 4. *Assume $Z(\tau)$ is given by* (18) *and conditions* $C1$ *and* $C2$ *hold for $r = \gamma$ and $r = \sigma\gamma + \rho$. Then the economy has two balanced-growth equilibria. The type*-1 *equilibrium has $r = \sigma\gamma + \rho$ and $\gamma_c = \gamma$. The type*-2 *equilibrium has $r = \gamma$ and $\gamma_c = (r-\rho)/\sigma$.*

To develop the intuition, notice that on a balanced-growth equilibrium individual consumption must grow at a constant rate. There are different such paths that satisfy the intertemporal budget constraint and market clearing. In the type-1 equilibrium, individual consumption is growing at the same rate as per capita net output at every time. In this case, all agents of all ages have the same consumption level equal to per capita net output. Another path consistent with market clearing is associated with the type-2 equilibrium. If, as I will assume later, $\sigma \geq 1$, then consumption declines relative to net individual income over the life-cycle so that young agents borrow and old agents lend. A complete analysis of the stability properties of the two equilibria is beyond the scope of this paper.

5. NUMERICAL EXPERIMENTS

I will analyze numerically the effect of differences in investment distortions and the effectiveness of schools on the two types of balanced-growth equilibria. To this purpose, I need to assign values to the parameters of the model. To impose some discipline, in Section 5.1 parameter values will be selected so that the equilibrium of the benchmark economy mimics observed features of the U.S. economy. Then, in Section 5.2, I will be able to compare balanced-growth path equilibria of different economies. Solving for an equilibrium involves, first, calculating the interest rate as in Proposition 4, second, computing optimal adoptions and schooling by solving (8) and (12), and, finally, computing output in (16) and a measure of the average age of machines in operation.[13] Section 5.3 considers modifications to the basic model. The findings will be discussed in Section 5.4.

5.1. Calibration

In order to be able to match observations on the interest rate and output growth, the type-1 equilibrium is chosen to select the benchmark parameters. The parameters of the model are: γ, p, ρ, σ, δ, π, h_0, β, μ_0, and λ. I will calibrate these parameters under two different assumptions about the role of education. These two assumptions are the *learning* and the *neutrality* hypotheses discussed in Section 3. For each of the two calibrations, I use the following procedure.

1. Set calibration targets for long-run growth rate of output per capita, risk-free interest rate, life-expectancy, speed of learning-by-doing, progress ratio, investment/output ratio, years of schooling, and general skill.
2. Choose γ, σ, and ρ so that, using Proposition 4, the targets for the growth rate of output per capita and the risk-free interest rate in step 1 are matched. Choose p so that $1/p$ matches the target for life-expectancy in 1. Choose δ consistent with the progress ratio in 1.
3. Fix values of s and h consistent with the targets in step 1.
4. Pick a value for π.
5. Pick a value for μ.
6. Compute the optimal adoption length m.
7. Check that outcome is consistent with target for speed of learning in 1. If not, update μ and go back to 6.
8. Check that outcome is consistent with target for investment/output ratio in 1. If not, update π and go back to 5.
9. Choose parameters h_0, β, μ_0, and λ so that the optimal choice of schooling s and $h(s)$ matches the values s and h fixed in step 3, and $\mu(s)$ matches the μ found as output of steps 4–8.

I will now describe the choice of values for the targets in step 1. I use an average growth rate of output per capita of 2% and a risk-free interest rate

[13] The average age of machines is computed as $(1/\gamma)\log(A(t)/\overline{A}(t))$ with $\overline{A}(t) = A(t)(p/(\gamma+p))(1-\exp(\gamma+p)m)/(1-\exp(-pm))$.

of 4.5%. Life in the model starts after early childhood so life-expectancy is set at 66.66 years, which is in line with figures in Barro and Lee (1994). Following the choice in Klenow (1998), the progress-ratio is fixed a priori to 2. Studies on learning-by-doing indicate that progress ratios vary widely across plants, sectors, and industries. The choice made here can be justified as a rough average of various pieces of evidence.[14] To measure the speed of learning I draw on evidence in Bahk and Gort (1993) and Bessen (1997) that learning is exhausted after approximately 5 years. For the investment/output ratio I use 7.3% share of equipment investment in GNP from NIPA. Given that the model excludes investment on the intensive margin, I take as satisfactory an investment/output ratio around 6.4%. The target figure for educational attainment is 10 years. Average education in the U.S. population over age 25 was 9.36 years in 1965 and 11.78 in 1985. Finally, the value of general skill is normalized to 1.

In step 7, I want to verify if the implied choices are consistent with the target set for the speed of learning in step 1. Since in the model learning is a step function, to use that evidence an approximation must be done. Consider a parameterized version of the learning curve used in Parente (1994) which expresses output as a function of time, $2-(2-1)\exp(-1.4565t)$. This continuous-learning curve involves the target progress ratio of 2 and is roughly consistent with the evidence in that, after 5 years, learning involves growth less than 0.05%.[15] I will move on to step 8 when μ and π are such that the model's learning process is equivalent to

[14] Rapping (1965) found a 40% yearly productivity growth in U.S. ship-building yards over 1941–1944. These figures suggest a lower bound for the progress ratio around 3.3. Jovanovic and Nyarko (1995) fit learning curves to productivity data on a number of new activities. The data they use include progress ratios ranging from 111 to 3, but for many activities more modest learning rates are found. Other studies seem to suggest that progress ratios are generally of the order or several percentage points only, much smaller than the several-fold gains found in case studies. In Bahk and Gort (1993) the total productivity gain is around 15% for recently born plants (average 1% per year over 14 years). On a sample of new plant start-ups in U.S. manufacturing industries, Bessen (1997) estimates an average progress ratio after learning is exhausted (5 years) of 5.7%.

As Bessen (1997) points out, the case study literature shows that most productivity gains from learning occur initially over a short span of time. Thus low frequency data and the loss of information on initial months of activity may explain why some studies—such as Bahk and Gort (1993)—tend to underestimate the measures of learning or, in other words, to overestimate initial productivity.

[15] As mentioned in footnote 13, it seems that most learning accrues during the first few months of an activity. Bessen (1997) argues that assuming that, on average, reported annual data drops the first 6 months of activity helps to reconcile the differing findings across empirical studies on learning. The continuous-learning approximation to the current calibration implies a half-life of 0.47, which means that after the first 6 months of activity about a 30% increase in productivity is left to be learned. This is more in line with the industry figures found in Bahk and Gort (1993) and Bessen (1997). Parente (2000) assumes a larger progress ratio and a somehow higher speed of learning.

TABLE I
Benchmark Economy

		Learning hypothesis	Neutrality hypothesis
$\sigma = 1.25$	$\rho = 0.02$	$\mu_0 = 5000$	$\mu_0 = 0.6728$
$\gamma = 0.02$	$p = 0.015$	$\lambda = 0.8914$	$\lambda = 0.0$
$\delta = 2.0$	$\pi = 1.12$	$h_0 = 1.0$	$h_0 = 0.43$
		$\beta = 0.0$	$\beta = 0.37$

this continuous learning process in the sense that, for the model's optimal m, the two processes yield the same value for the firm.[16]

Finally, in step 9 the computation of the optimal choice of s is done on a discrete grid. In all the calculations, the objective in problem (12) is well defined with the possibility (under the learning hypothesis) of a local maximum at $s = 0$. Table I summarizes the choices of parameters.

5.2. Balanced-Growth Comparisons

This section considers the balanced-growth effect of increases in the size of investment distortions, π, and decreases in school effectiveness, λ or β, relative to the benchmark economy. Figures on educational attainment s, relative income per capita y/y_{US}, and the average age of machines are reported. The length of technology adoptions, m, is also reported. The fact that the interest rate remains unaffected in either type of equilibrium means that the analysis in Section 3 carries over to interpret the results.

5.2.1. *Type*-1 *Equilibrium*

Table II reports figures associated with various values of π under the two assumptions about the role of education for type-1 equilibria. The first row corresponds to the benchmark economies.

Under the *learning hypothesis*, the return to education is closely linked to the pattern of technology adoptions. As discussed in Section 3.2.1, higher size of distortions that affect the cost of equipment tends to reduce the frequency of technology updates and the opportunities for learning. Thus education that improves learning ability is less valuable and agents reduce the period of time spent in school. In the parameterized economy analyzed

[16] In other words,

$$W(m) = \int_0^m e^{-(r+p)t}\left[2 - (2-1)e^{-1.4565t}\right] dt - \frac{\pi}{h},$$

where m is the model's optimal time between adoptions and $W(m)$ is as in Eq. (6) with $\delta = 2$.

TABLE II
Type-1 Equilibrium. Changes in Investment Distortions

Investment distortions	Learning hypothesis				Neutrality hypothesis			
π	s	m	y/y_{US}	age	s	m	y/y_{US}	age
1.12	10.0	10.85	1.0	5.18	10.0	10.85	1.0	5.18
1.72	9.9	13.06	0.987	6.17	10.3	12.81	1.0	6.06
2.32	9.8	15.11	0.973	7.08	10.5	14.61	0.987	6.86
2.92	9.8	16.86	0.960	7.84	10.8	16.26	0.979	7.58
3.52	9.7	18.74	0.946	8.64	11.0	17.85	0.973	8.27

here, this mechanism dominates the effect of a lower opportunity costs of time. A lower frequency of technology adoptions increases the average age of technology and reduces aggregate productivity. Hence, in the calibrated model, cross-country variation in investment distortions creates a positive correlation between output per capita and educational attainment, and a negative correlation with the age of the machines in operation.

Under the *neutrality hypothesis*, the prediction for the correlation output-years of schooling is the opposite. Section 3.2.2 showed that, much like under the learning assumption, the higher cost of equipment reduces the frequency of technology upgrades. Thus education, to the extent that it contributes to reduce the perceived cost of adoptions (through the term π/h), becomes less valuable which tends to shorten the optimal schooling period. But in this economy, this effect is too weak as compared with the response to the reduction in the opportunity cost of schooling time.

The effect of changes in measures of the effectiveness of schools is illustrated in Table III.

Under the *learning hypothesis* for education, differences in λ predict a negative association between educational attainment and output. In this economy, a lower λ leads to a reduction in the opportunity cost of schooling time that overwhelms the other effects discussed in Section 3.2.1.

TABLE III
Type-1 Equilibrium. Changes in School Effectiveness

Effectiveness	Learning hypothesis				Effectiveness	Neutrality hypothesis			
λ	s	m	y/y_{US}	age	β	s	m	y/y_{US}	age
0.89	10.0	10.85	1.0	5.18	0.37	10.0	10.85	1.0	5.18
0.78	11.3	11.16	0.979	5.32	0.27	7.4	12.47	0.758	5.81
0.70	12.3	11.58	0.953	5.51	0.17	4.8	13.79	0.591	6.50
0.59	14.3	12.28	0.913	5.83	0.07	2.0	15.27	0.490	7.15

In particular, it is true that the induced rise in m tends to reduce the value of education, but in this case this effect is too weak.

Under the *neutrality* view of education, as discussed in Section 3.2.2, a reduction in β reduces the direct marginal return of education as well as the indirect return from reducing the effective cost of technology adoption π/h. Therefore, it creates incentives for shorter schooling that dominate the response to a lower opportunity costs of schooling time. Thus lower effectiveness of schools reduces schooling and output per capita. In addition, it increases the effective adoption cost, $\pi/h(s)$, thereby raising the average age of capital and reducing output per capita further.

5.2.2. *Type-2 Equilibrium*

With these parameters, type-2 equilibria are associated with a lower interest rate than type-1 equilibria. Hence the model produces higher levels of investment in education and output. Comparisons of type-2 balanced growth equilibria arising from differences in π, λ and β under the two assumptions on the role of education lead to exactly the same type of conclusions as for type-1 equilibria. Tables IV and V report the results of the computations for type-2 equilibria under the parameters in Table I. The two scenarios that deliver plausible correlations between output,

TABLE IV
Type-2 Equilibrium. Changes in Investment Distortions

Investment distortions	Learning hypothesis				Neutrality hypothesis			
π	s	m	y/y_{US}	age	s	m	y/y_{US}	age
1.12	10.9	9.15	1.00	4.40	26.0	9.26	1.00	4.45
1.72	10.9	11.11	0.980	5.30	26.5	10.63	0.994	5.08
2.32	10.9	12.85	0.967	6.08	27.0	11.86	0.982	5.64
2.92	10.8	14.52	0.954	6.82	27.4	13.00	0.976	6.15
3.52	10.7	16.08	0.941	7.50	27.8	14.07	0.964	6.63

TABLE V
Type-2 Equilibrium. Changes in School Effectiveness

Effectiveness	Learning hypothesis				Effectiveness	Neutrality hypothesis			
λ	s	m	y/y_{US}	age	β	s	m	y/y_{US}	age
0.89	11.20	8.90	1.00	4.29	0.37	24.90	9.31	1.00	4.47
0.78	12.60	9.07	0.974	4.36	0.27	19.30	10.57	0.734	5.05
0.70	13.80	9.19	0.954	4.42	0.17	12.40	12.16	0.556	5.77
0.59	14.90	10.72	0.902	5.12	0.07	5.20	13.88	0.444	6.54

educational attainment, and the age of capital are the same as when comparing type-1 steady states in Section 5.2.1.

5.3. Robustness and Extensions

With parameters calibrated to the type-2 steady state, the qualitative findings remain the same. Also, the nature of the results seems robust to alternative parameterizations and interpretations of parameters within the model. Concerning the latter, the interpretation of μ_0 and h_0 as measures of the effectiveness of schools leads to the same results. However, the model itself contains simplifying assumptions that at this stage deserve to be examined more carefully. The first is the assumption on the distribution of wealth as expressed in Eq. (18). The second is the assumption of learning as a discrete process. In the remaining part of this section, I will examine alternative assumptions and demonstrate that the results derived so far remain largely intact.

5.3.1. *The Distribution of Wealth*

An alternative to (18) is the more natural assumption that an agent's lifetime wealth, $Z(\tau)$, consists of the value of the firm he or she operates. This amounts to replacing (18) by the following,

$$Z(\tau) = A(\tau)e^{-(r+p-\gamma)s}h(s)V(\mu(s), \pi/h(s)), \tag{20}$$

where the notation is the same as in problem (12). The equilibrium conditions apply to this case. Furthermore, a modest amount of work shows that, under the assumption on the distribution of wealth contained in (20), the characterization of balanced-growth equilibria is similar to that provided in Proposition 4. As in that proposition, $r = \gamma$ is an equilibrium. Thus the type-2 equilibrium characterizes a balanced-growth path irrespective of the wealth distribution. Moreover, for all the economies analyzed under the assumption in (20), I have found that there is another equilibrium that, in terms of the value of r relative to the one in the type-2 equilibrium, resembles the type-1 equilibrium of Proposition 4. However, differently from the type-1 equilibrium under assumption (18), when (20) holds the interest rate cannot be characterized analytically and, in general, will change when parameters of the model change. Hence in this economy there will be richer general equilibrium effects from changes in π, λ, or β. I study the scope of these new effects for economies that have the benchmark parameters in Table I under the learning hypothesis. For the sake of comparability, for the *neutrality hypothesis* I have adjusted $\beta = 0.425$ and $h_0 = 0.38$. Now computing equilibria requires an additional round of iterations on the interest rate.

TABLE VI
Type-1 Equilibrium. Changes in Investment Distortions

Alternative assumption on wealth distribution										
Investment distortions	Learning hypothesis					Neutrality hypothesis				
π	s	m	y/y_{US}	age	$r\%$	s	m	y/y_{US}	age	$r\%$
1.12	9.8	11.44	1.0	5.45	5.18	9.8	11.42	1.0	5.44	5.17
1.72	9.7	13.71	0.986	6.46	5.18	10.1	13.40	1.0	6.33	5.20
2.32	9.7	15.61	0.973	7.30	5.19	10.3	15.24	0.993	7.14	5.22
2.92	9.6	17.66	0.960	8.18	5.20	10.5	16.98	0.986	7.89	5.24
3.52	9.6	19.42	0.946	8.93	5.21	10.7	18.64	0.980	8.60	5.26

TABLE VII
Type-1 Equilibrium. Changes in School Effectiveness

Alternative assumption on wealth distribution											
Effec-tiveness	Learning hypothesis					Effec-tiveness	Neutrality hypothesis				
λ	s	m	y/y_{US}	age	$r\%$	β	s	m	y/y_{US}	age	$r\%$
0.89	9.8	11.44	1.0	5.45	5.18	0.42	9.8	11.42	1.0	5.44	5.17
0.78	11.0	11.98	0.973	5.69	5.24	0.33	7.8	12.76	0.756	6.04	5.06
0.70	12.0	12.44	0.953	5.90	5.30	0.23	5.7	14.28	0.588	6.71	4.94
0.63	13.2	13.05	0.919	6.17	5.36	0.13	3.4	15.86	0.475	7.41	4.80

The results of the experiments for type-1 equilibria under (20) are illustrated in Tables VI and VII. The signs of the correlations do not differ from the ones found with a constant interest rate. The model also delivers predictions for the correlation between output and the interest rate. This correlation is negative except for the changes in the school parameter under the neutrality assumption.

5.3.2. *Continuous Learning*

Assume learning-by-doing is a smooth function of time. In particular, let technology-specific skill, $q(m)$, be governed by the following version of the learning curve in Parente (1994),

$$q(m) = \delta - (\delta - 1)e^{-\mu m},$$

where δ still represents the progress ratio and μ measures the speed of learning. I replace Eq. (2) with the specification $\mu(s) = \mu_0 s^{\lambda}$. I maintain the other assumptions of the model. In this case, Propositions 1 and 2 still hold and the solution to the adoption problem is the unique solution to a smooth first-order condition. Different from the case with discreet learning, now the qualitative effect of μ on the optimal frequency of adoptions is ambiguous. The same arguments as before lead to the two balanced-growth path equilibria in Proposition 4. I calibrate this model to the same targets as the discreet-learning model for the type-1 equilibrium. This results in exactly the same values for parameters other than those of the schooling technology in Table I. The parameters of the schooling technology that differ from those in Table I are as follows. Under the *learning hypothesis*, $\mu_0 = 1.4566 \times 10^{-9}$ and $\lambda = 9.0$. Under the *neutrality hypothesis* $\mu_0 = 1.4566$.

Figures for comparisons of balanced-growth paths for different π, λ and β are displayed in Tables VIII and IX. They are virtually identical to those for the model with learning as a step function in Tables II and III.

TABLE VIII
Type-1 Equilibrium. Changes in Investment Distortions

Continuous learning								
Investment distortions	Learning hypothesis				Neutrality hypothesis			
π	s	m	y/y_{US}	Age	s	m	y/y_{US}	Age
1.12	10.0	10.85	1.0	5.18	10.0	10.85	1.0	5.18
1.72	9.9	13.06	0.987	6.17	10.3	12.81	0.993	6.06
2.32	9.8	15.12	0.973	7.08	10.5	14.61	0.987	6.86
2.92	9.8	16.86	0.956	7.84	10.8	16.26	0.970	7.58
3.52	9.7	18.74	0.946	8.64	10.9	17.34	0.973	8.05

TABLE IX
Type-1 Equilibrium. Changes in School Effectiveness

Continuous learning									
Effectiveness	Learning hypothesis				Effectiveness	Neutrality hypothesis			
λ	s	m	y/y_{US}	Age	β	s	m	y/y_{US}	Age
9.0	10.0	10.85	1.0	5.18	0.37	10.0	10.85	1.0	5.18
8.56	11.0	11.36	0.980	5.41	0.27	7.4	12.25	0.756	5.81
8.11	12.3	11.80	0.946	5.61	0.17	4.8	13.79	0.591	6.50
7.44	14.6	13.17	0.886	6.22	0.07	2.0	15.27	0.490	7.15

5.4. Discussion

A theory of cross-country disparities must be consistent with the fact that richer countries tend to be more educated and use more advanced technologies. Under the *learning* assumption of education, it is found that cross-country disparity caused by variation in the size of investment distortions (but not in school effectiveness) is consistent with a positive association between output per capita and the level of educational attainment. Under the *neutrality hypothesis*, variation in school effectiveness (but not in investment distortions) has implications consistent with these relationships. Furthermore, under these two scenarios, the average age of technologies in operation declines with the level of income per capita.

These results therefore illustrate the possible conflicting implications of different assumptions on the role of education. A relevant question at this stage is to decide which of the two assumptions considered here is more reasonable. The *learning hypothesis* is explicit about the sorts of skills provided by education and has been invoked in a number of empirical studies such as Welch (1970), Bartel and Litchenberg (1987), Foster and Rosenzweig (1996), and Rosenzweig (1995). In addition, this view can be rationalized, as Rosenzweig (1995) does, in the context of the information-theoretic model of learning in Jovanovic and Nyarko (1996). This approach creates a link between technology and the returns to schooling, and this paper has formalized this link and the analysis of its implications. The analysis of technology adoption proves to be essential to the argument.

On its part, what I have called the *neutrality hypothesis*, however, appears to be less explicit about what the skills involved are and where they come from. In this view, the connection between technology and the returns to schooling is weaker and plays just a supportive role. In other words, the predictions for schooling and output in a model that ignores the technology adoption problem—formally, with $V(\mu, \pi/h)$ replaced by a constant in problem (12)—would be essentially the same.

All this notwithstanding, this paper cannot discriminate the two explanations of disparities across countries as the level effects associated with differences in schooling effectiveness and investment distortions are small under the two assumptions. This is not surprising for the case of changes in investment distortions under the *learning hypothesis*. In this model, as in Jovanovic and Rob (1998) and Parente (2000), differences in the age distribution of technologies alone cannot produce wide differences in income per capita. Only large differences in speeds of learning-by-doing coupled with a large learning progress ratio can amplify the effects. However, the response of education in the model seems unable to bring about sizeable differences in the speed of learning and thereby in output even when large progress ratios are considered.

The quantitative properties of the model are surely related to its implications for the returns to schooling. For the benchmark economy, the return to one more year of education under the *neutrality* assumption is around 4% for the type-1 equilibrium.[17] Under the *learning* assumption, the figure lies between 3.4 and 6.7%. Bils and Klenow (1998) document an average Mincerian return to schooling of 9.9% on a sample of 52 countries. Hence the calibration of the present model seems to grossly underestimate the return to education. This is more so for the type-2 equilibrium where the interest rate is unrealistically low.

6. CONCLUSION AND FINAL REMARKS

This paper integrates the analysis of choices on education and on technology adoption to study international disparities. The analysis is conducted in a vintage capital model with technology specific learning-by-doing. Unlike previous studies, this allows us to evaluate the importance of distortions that affect the cost of technology adoption and the effectiveness of education by testing their implications for output per capita, educational attainment, and the age of technologies in use. Remarkably, the analysis of technology adoption allows consideration of alternative assumptions on the role of education.

The findings for a reasonably parameterized version of the model are as follows. If the role of education is to improve the ability to learn technologies, only differences in the cost of technology adoption produce consistent correlations. On the other hand, if the role of education is to provide an absolute productive advantage that is independent of learning, then only differences in school effectiveness produce consistent correlations.

Two main conclusions emerge from the analysis of this paper. First, predictions that can be obtained under supposedly simplifying assumptions on the role of education—such as the *neutrality hypothesis*—may differ dramatically from the predictions derived under other less-simplifying assumptions, such as the *learning hypothesis*. Second, in the model analyzed changes in different factors yield different qualitative predictions. These differing implications can potentially be used to assess the importance of different factors in development. However, the present paper stands far on the way toward drawing stronger conclusions in this sense; not the least of

[17] This number has been calculated numerically from the impact on aggregate output per firm of changing s exogenously from 9 to 10 and from 10 to 11. It is close to the percentage change in h, β/s, thus suggesting that the effect of education on output through π/h is small in the margin.

the reasons is that the model's quantitative predictions are not in accord with the data.

Directions for further research involve natural extensions of the present framework. The model contains assumptions that impose tight bounds on the size of the differences in income that can be produced. Notably, all agents adopt the frontier technology and learning is completely technology specific. I think that a model analyzing the transmission of knowledge across technologies, rather than uniquely its accumulation within each single technology, offers reasonable expectations of better matching the data. In such a framework, the mechanisms analyzed here would still play an important role.

APPENDIX: PROOF OF PROPOSITIONS 1 THROUGH 3

Proof of Proposition 1. C1 implies that the solution is a feasible adoption plan $(J, \{x_j\}, a(t, \tau))$ since a feasible adoption plan with $J = 1$ exists where the expression in (7) has a positive value.

I will now prove that in an optimal plan $W(m_j) > 0$ for all $j = 1, 2, \ldots, J$. In an optimal adoption plan $W(m_{j'}) > 0$ for some j'. Then $W(m_1) > 0$ since, otherwise, the value in (7) would increase by holding on until $x_1 = x_{j'}$. Finally, assume that, for some j in an optimal plan, $W(m_j) \leq 0$. Since $W(\cdot)$ is monotonically increasing, the value in (7) would increase by holding out the $j - 1$th adoption until x_{j+1}. But then x_j cannot be the jth adoption which contradicts the optimality of the plan containing x_j.

Now part (a) follows from inspection of (7). To prove part (b), use that $A(x_j) = A(\tau + s)\exp(\gamma(x_j - (\tau + s)))$ to rewrite (7) as

$$h \cdot A(\tau + s) \cdot \sum_{j=1}^{J} W(x_{j+1} - x_j) e^{-(r+p-\gamma)(x_j - (\tau+s))}$$

$$= h \cdot A(\tau + s) \cdot \sum_{j=1}^{J} W(m_j) e^{-(r+p-\gamma)(x_1 + \sum_{i=1}^{j-1} m_i - (\tau+s))}$$

where the equality follows from the notation $m_j \equiv x_{j+1} - x_j$. Therefore, for any given sequence of tenures $\{m_j\}$, by anticipating the start of production [i.e., reducing x_1] the value of (7) increases unambiguously by assumption C2. That $\tau + s$ is the earliest feasible starting date concludes. *Q.E.D.*

Proof of Proposition 2. The results in Proposition 1 imply that (7) can be written as

$$h \cdot A(\tau + s) \cdot V\left(\{m_j\}_{j=1}^{J}\right),$$

with

$$V\left(\{m_j\}_{j=k}^{J}\right) \equiv \sum_{j=k}^{J} W(m_j)e^{-(r+p-\gamma)\Sigma_{i=k}^{j-1}m_i}.$$

So the problem is to choose J and $\{m_j\}$ that maximize $V(\{m_j\}_{j=1}^{J})$. The objective can be written recursively as

$$V\left(\{m_j\}_{j=k}^{J}\right) = W(m_k) + e^{-(r+p-\gamma)m_k}V\left(\{m_j\}_{j=k+1}^{J}\right),$$

for $k = 1, 2, \ldots, J$. Let $V_k \equiv \max V(\{m_j\}_{j=k}^{J})$. Since the value $V(\{m_j\}_{j=k}^{J})$ only depends on the sequence of tenures, the solution must be the same value $V = V_k$ for all k. Hence the problem is equivalent to solving (8). Expression (9) follows from the definitions. *Q.E.D.*

Proof of Proposition 3. To simplify notation, let $D(m) \equiv e^{-(r+p-\gamma)m}$. At points $m \neq \mu$ the derivative of the objective in (8) is given by $W'(m) + D'(m)(W(m)/(1 - D(m)))$. By (8), at the optimum, $V = W(m)/(1 - D(m))$ and this derivative is given by the left-hand side of (10) denoted by $\Delta(m)$.

It is helpful to write $\Delta(m)$ as

$$W'(m) + D'(m)\frac{W(m)}{1 - D(m)} = e^{-(r+p)m}\Gamma(m),$$

where

$$\Gamma(m) \equiv \begin{cases} 1 - e^{\gamma m}\left[\dfrac{r+p-\gamma}{r+p}\left(1 - (r+p)\dfrac{\pi}{h}\right) + e^{-(r+p)m}\dfrac{\gamma}{r+p}\right] \\ \quad \text{if } m < \mu \\ \delta - e^{\gamma m}\left[\dfrac{r+p-\gamma}{r+p}\left(1 + (\delta - 1)e^{-(r+p)\mu} - (r+p)\dfrac{\pi}{h}\right)\right. \\ \quad \left. + \, \delta e^{-(r+p)m}\dfrac{\gamma}{r+p}\right] \quad \text{if } m > \mu \end{cases}$$

Thus I can analyze the shape of $\Delta(m)$ in (10) by analyzing $\Gamma(m)$.

a. By C1 and C2 it follows that $\Gamma(+\infty) < 0$ which rules out the optimality of $m = +\infty$.
By C2, $\Gamma(0) > 0$ which rules out $m = 0$. C2 also implies $\lim_{\mu^+} \Gamma(m) > \lim_{\mu^-} \Gamma(m)$ which rules out optimality of $m = \mu$ where $\Gamma(\cdot)$ is non differentiable. Any local extremum must then satisfy $\Gamma(m) = 0$.

b. A little algebra shows that in either region, $m < \mu$ or $m > \mu$, there can be at most one root of $\Gamma(m) = 0$. This follows from the fact that if, for some m, $\Gamma(m)$ is decreasing, then it will be so for any larger m in either region.
It is useful to state this as a lemma.

LEMMA 1. *Let m and m' belong to either* $(0, \mu)$ *or* (μ, ∞). *Assume $m' > m$. If* $\Gamma'(m) < 0$, *then* $\Gamma'(m') < 0$. *If* $\Gamma(m) = 0$, *then m is a local maximum.*

Now I will take the different cases in turn.

—Case $\lim_{\mu^-} \Gamma(m) > 0$. By C2 one finds $\Gamma(0) > 0$ and $\Gamma'(0) > 0$. Then, by Lemma 1, $\lim_{\mu^-} \Gamma(m) > 0$ implies that $\Gamma(m) > 0$ all $m < \mu$. Thus there is no $m < \mu$ such that $\Gamma(m) = 0$.
By C2 it follows that $\lim_{\mu^-} \Gamma(m) < \lim_{\mu^+} \Gamma(m)$. Then $\lim_{\mu^-} \Gamma(m) > 0$ implies $\lim_{\mu^+} \Gamma(m) > 0$. Now C1 and C2 imply $\Gamma(\infty) < 0$ and then there exists an $m > \mu$ such that $\Gamma(m) = 0$. Lemma 1 concludes, establishing this is unique and the maximum.

—Case $\lim_{\mu^-} \Gamma(m) < 0$ and $\lim_{\mu^+} \Gamma(m) < 0$. C2 implies $\Gamma(0) > 0$. Then $\lim_{\mu^-} \Gamma(m) < 0$ and Lemma 1 implies that there is a single $m < \mu$ such that $\Gamma(m) = 0$ which is a local maximum.
It is a fact that $\lim_{\mu^-} \Gamma'(m) = \lim_{\mu^+} \Gamma'(m)$. Since Lemma 1 implies $\lim_{\mu^-} \Gamma'(m) < 0$, it follows that $\lim_{\mu^+} \Gamma'(m) < 0$. Now the assumption that $\lim_{\mu^+} \Gamma(m) < 0$ implies that $\Gamma(m) < 0$ all $m > \mu$. Part a of Proposition 3 then concludes that the maximum is the $m < \mu$ such that $\Gamma(m) = 0$.

—Case $\lim_{\mu^-} \Gamma(m) < 0$ and $\lim_{\mu^+} \Gamma(m) > 0$.
As in the previous case there exists one $m < \mu$ such that $\Gamma(m) = 0$. Similarly, using that C1 and C2 imply $\Gamma(+\infty) < 0$, there is also a local maximum $m > \mu$. By (8), at the solution $V = W(m)/(1 - D(m))$. This concludes the proof. *Q.E.D.*

REFERENCES

Bahk, B., and Gort, M. (1993). "Decomposing Learning by Doing in New Plants," *Journal of Political Economy* **101**, 561–583.

Barro, R. J., and Lee, J-W. (1993). "International Comparisons of Educational Attainment," *Journal of Monetary Economics* **32**, 363–394.

Barro, R. J., and Lee, J-W. (1994). "Sources of Economic Growth," *Carnegie-Rochester Conference Series on Public Policy* **40**, 1–46.

Barro, R. J., and Lee, J-W. (1997). "Schooling Quality in a Cross-Section of Countries," NBER Working Paper No. 6198.

Bartel, A. P., and Litchtenberg, F. R. (1987). "The Comparative Advantage of Educated Workers in Implementing New Technology," *The Review of Economics and Statistics* **69**, 1–11.

Bessen, J. (1997). "Productivity Adjustments and Learning-by-Doing as Human Capital," mimeo.

Bils, M., and Klenow, P. J. (1998). "Does Schooling Cause Growth or the Other Way Around?" NBER Working Paper No. 6393.

Chari, V., Kehoe, P., and McGrattan, E. (1997). "The Poverty of Nations: A Quantitative Investigation," Fed Minneapolis Research Department Staff Report 204.

Ciccone, A. (1996). "Rapid Catch-Up, Fast Convergence and Persistent Underdevelopment," mimeo.

Cooley, T. F., Greenwood, J., and Yorukoglu, M. (1997). "The Replacement Problem," *Journal of Monetary Economics* **40**, 457–499.

Foster, A. D., and Rosenzweig, M. R. (1996). "Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review* **86**, 931–953.

Greenwood, J., and Yorukoglu, M. (1996). "1974," Rochester Center for Economic Research Working Paper No. 429.

Grossman, G. M., and Helpman, E. (1991). *Innovation and Growth in the Global Economy*, MIT Press.

Hendricks, L. (1997). "Equipment Investment and Growth in Developing Countries," mimeo.

Jones, C. I. (1994). "Economic Growth and the Relative Price of Capital," *Journal of Monetary Economics* **34**, 359–382.

Jovanovic, B. (1998). "Vintage Capital and Inequality," *Review of Economic Dynamics* **1**, 497–530.

Jovanovic, B., and Rob, R. (1998). "Solow vs Solow," mimeo.

Jovanovic, B., and Nyarko, Y. (1995). "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves," *Brooking Papers of Economic Activity*, Microeconomics, 245–300.

Jovanovic, B., and Nyarko, Y. (1996). "Learning-by-Doing and the Choice of Technology," *Econometrica* **64**, 1299–1310.

Keller, W. (1996). "Absorptive Capacity: On the Creation and Acquisition of Technology in Development," *Journal of Economic Development* **49**, 199–227.

Klenow, P. J. (1998). "Learning Curves and the Cyclical Behavior of Manufacturing Industries," *Review of Economic Dynamics* **1**, 531–550.

Lucas, R. E. (1988). "On the Mechanics of Economic Development," *Journal of Monetary Economics* **22**, 3–42.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* **107**, 407–437.

Nelson, R., and Phelps, E. (1966). "Investments in Humans, Technological Diffusion, and Economic Growth," *American Economic Review*, *Papers and Proceedings* **61**, 69–75.

Parente, S. L. (1994). "Technology Adoption, Learning-by-Doing and Economic Growth," *Journal of Economic Theory* **63**, 346–369.

Parente, S. L. (2000). "Learning-by-Using and the Switch to Better Machines," *Review of Economic Dynamics* **3**, 675–703.

Parente, S. L., and Prescott, E. C. (1994). "Barriers to Technology Adoption and Development," *Journal of Political Economy* **102**, 298–321.

Rapping, L. A. (1965). “Learning and World War II Production Functions,” *Review of Economics and Statistics* **47**, 81–86.

Restuccia, D. (1997). “Technology Adoption and Schooling,” mimeo.

Rosenzweig, M. R. (1995). “Why Are There Returns to Schooling?,” *American Economic Review*, *Papers and Proceedings* **85**, 151–158.

Schultz, T. W. (1975). “The Value of the Ability to Deal with Disequilibria,” *Journal of Economic Literature* **13**, 827–846.

Stokey, N. L. (1991). “Human Capital, Product Quality, and Growth,” *Quarterly Journal of Economics* **106,** 587–616.

Summers, R., and Heston, A. (1991). “The Penn World Table (Mark 5): An Expanded Set of International Comparisons,” *Quarterly Journal of Economics* **106**, 327–368.

Welch, F. (1970). “Education in Production,” *Journal of Political Economy* **78**, 35–59.

Wolff, E. (1991). “Capital Formation and Productivity Convergence over the Long-Term,” *American Economic Review* **81**, 565–579.