The Social Value of Education and Human Capital

Fabian Lange **and** **Robert Topel**

Yale University **University of Chicago**

Revised

September, 2004

Abstract

We review and extend the empirical literature that seeks evidence of a wedge between the private and social returns to human capital, specifically education. This literature has two main strands. First, much of modern growth theory puts human capital at center-stage, building on older notions of human capital externalities as an engine of economic growth. Empirical support for these ideas, based on both the comparative growth of national outputs and on the geographic dispersion of wages within countries, is meager. There is a strong association between average earnings and average education across nations and regions in the US that exceeds the private returns to education. However, problems of omitted variables and endogeneity inherent in spatial equilibrium models mean that this association can hardly be understood as evidence for social returns, There is no evidence from this literature that social returns are *smaller* than private ones, yet neither is there much to suggest that they are *larger*. We then turn to the literature on job market signaling, which implies that social returns to education are smaller than private returns. Consistent with our earlier conclusions, we find scant evidence of this. We construct a model of the speed of employer learning about workers' unobserved talents, and we estimate the model using panel data on young workers. We find that employer learning about productivity occurs fairly quickly after labor market entry, implying that the signaling effects of schooling are small.

Economists (and others) have generally had little success in estimating the social effects of different investments, and, unfortunately, education is no exception.

Becker, *Human Capital* (1975)

1. Introduction

This paper reviews and extends the literature on the social returns to accumulation of human capital, with particular emphasis on the social returns to education. Here and in what follows we define "social returns" to be the sum of the private and external marginal benefits of a unit of human capital. In other words, we study the problem of human capital externalities—does an individual's private decision to accumulate human capital confer external benefits or costs on others?

There are three main strands to the literature on human capital externalities, each of which touches on externalities created by the accumulation of education. First, in a formalization of ideas that go back at least as far as Marshall (1890)[1] recent theories of economic growth emphasize human capital accumulation as an engine of growth. Following Lucas (1988), who built on earlier work by Uzawa (1965) and others, growth theorists have emphasized interactions amongst agents that may cause the social returns to human capital to exceed the private ones. Persons with greater skill may raise the productivity of others with whom they interact, so accumulation of human capital may increase total factor productivity in an economy. We provide an overview of growth models that generate such externalities, and we critique the empirical literature that seeks to identify these effects from both aggregate and micro data. We then extend this literature, modeling the contribution of education to growth in total factor productivity in states and regions of the U.S. since 1940. Looking ahead, our conclusion is that empirical evidence for important human capital externalities is, at best, weak.

In contrast to the growth literature—where education is alleged to produce positive externalities—models of the signaling value of education raise the possibility that some

[1] Marshall emphasized the social benefits of valuable ideas, which are public goods and, he implies, are more likely to be produced by the highly educated. "..[F]or one new idea, such as Bessemer's chief invention, adds as much to England's productive power as the labour of a hundred thousand men. …All that is spent …[in educating the masses] would be well paid for if it called out one more Newton or Darwin, Shakespeare, or Beethoven." *Principles of Economics, 8th Edition,* (1920).

component of schooling is a social waste. In the extreme form first formulated by Spence (1974) schooling acts as a signal of private information about individual productivities, for which employers are willing to pay, though it does not raise anyone's productivity. In the model's equilibrium it is privately optimal to invest in schooling—education has a private return because it transfers wealth from less to more skilled individuals—but the social return is negative since schooling does not raise individuals' productivities and it reduces social output by using valuable resources.[2] We review and critique the existing empirical literature on signaling, which we find provides little convincing evidence for an important role of Job Market Signaling. We extend that literature with an empirically tractable model of employer learning about individuals' talents. Applying our model to panel data from the National Longitudinal Study of Youth (NLSY) we find that learning about individual talents occurs fairly quickly. This allows us to put an empirical upper bound on the contribution of signaling to the private returns to schooling. We conclude that the impact of signaling on private returns is small—no more than about a tenth of the private return—so most of the returns to schooling reflect a positive impact of education on productivity.

A third strand of literature emphasizes possible external benefits of education that do not apply directly to the production process. They are not reflected in factor payments,[3] and so they are often less amenable to empirical research. Such external benefits might arise because education reduces criminal behavior (Moretti 2004), because education enables individuals to participate more efficiently in the political process (Friedman 1963), or because education carries direct consumption externalities. If knowledge of Shakespeare or Astronomy makes one more interesting, then investment in education raises the welfare of others through a form of network externality borne of social interactions. (Study of say, accounting, might have the opposite effect). This raises welfare without any discernable impact on wages or productivity. With rare exceptions—crime is the only one that we can think of—these putative social benefits of education are unmeasured. So our empirical review touches them only in passing.

[2] It could be that signaling improves the allocation of individuals to tasks and thus generates social returns. Butt in its most basic form we would expect that the social returns are lower than the private ones.

[3] By definition externalities are not reflected in the payments to factors generating the external effects. However, they might be reflected in payments to those benefiting from the external effects. This forms the basis of most existing empirical studies of external effects of education.

The remainder of the paper is organized in 2 major sections. Section 2 considers the evidence concerning external effects of education in production. Section 3 examines the empirical evidence on Job Market Signaling. Section 4 concludes.

2. Growth and Production Externalities of Human Capital

This section reviews and extends the empirical literature on production externalities of human capital, as measured (mainly) by education. For such externalities to exist, an individual's human capital must cause an unpriced increase in the productivity of others. Typically these effects are thought to occur through proximity and human interaction, though when productive interactions occur within firms they are merely complementarities that will be internalized and priced. This has led to an emphasis—both in theories and in applied work—on social interactions in cities, where ideas are sort of 'in the air.' Then the external benefits of human capital are localized, which has empirical implications for wages and land prices that are taken up in Section 2B. But the idea of productive externalities from human capital is not limited by geography, as Marshall's (1920) emphasized with his example of Pasteur's contribution to public health. Members of the "charmed circle" that produce such ideas "have probably earned for the world a hundred times or more as much as they have earned for themselves."[4] Such global externalities will not show up in geographic variation of factor prices and are thus outside of the scope of existing empirical attempts (including ours) to identify the social benefits from education.

We begin in 2A with a review of the place of human capital in modern theories of economic growth, and the state of empirical evidence derived from growth of national outputs. Section 2B outlines a model for evaluating the impact of human capital on local productivities, and reviews empirical work that seeks to identify externalities from geographic differences in wages and educational attainment. This research is extended in 2C, where we apply the model to data on the growth of U.S. states from 1940 to 2000.

---

[4] Marshall (1920), p598.

2.A. Human Capital, Education and Economic Growth

Recent interest among macroeconomists in the possibility of human capital externalities follows the revival of growth theory, which is built on the idea that human capital is central to growth. Following Lucas (1988), neoclassical models of growth treat human capital as a produced input to a standard constant returns technology, so that growth of human capital and growth of output are nearly synonymous.

To appreciate the special place of human capital in modern growth models, we begin with two key facts. First, as noted by Kaldor (1961), most countries have experienced sustained growth over very long periods of time—for example, growth in U.S. per-capita income averaged 1.75 percent per year over the 20th century. Second, the capital/output ratio is remarkably stable across countries, both rich and poor.[5] To accommodate these facts, Solow's (1956) original formulation of a growth model introduced an exogenous rate of labor-augmenting technical change to generate sustained growth in the face of diminishing returns to physical capital. To fix ideas let aggregate production be Cobb-Douglas with constant returns to capital and labor, with zero labor force growth:

(1) $$Y_t = K_t^{\alpha}(A_t L)^{1-\alpha}$$

Here A_t denotes the state of labor-augmenting technical progress, which grows at exogenous rate $\dot{a}_t = d\log(A_t)/dt$. Let $\dot{y}_t$ and $\dot{k}_t$ denote the growth rates of output per worker and capital per worker, so (1) implies

(2) $$\dot{y}_t = \alpha\dot{k}_t + (1-\alpha)\dot{a}_t$$

With a constant savings rate, output, capital and consumption grow at the common rate $\dot{a}$. This also means that the capital/output ratio is fixed in the steady state.

Human capital entered the picture when Theodore Schultz (1963) and other development economists interpreted the Solow residual $\dot{a}$ as growth in human capital.

[5] See Young (1992) or Figure 1 in Topel (1999).

This was formalized by Uzawa (1965) and Lucas (1988) who interpreted A_t as the average stock of human capital per worker, so $H=AL$ is the human capital stock. In Lucas' formulation aggregate human capital is an input to its own production, much as private human capital is in Ben-Porath's (1967) model of human capital investment for individuals:

$$Y_t = K_t^{\alpha}(uH_t)^{1-\alpha}$$

(3)

$$\frac{dH_t}{dt} = BH_t(1-u) - \delta H_t$$

where 1-u is the proportion of time devoted to production of new human capital. Here workers embody skills that accumulate through wealth maximizing investment decisions—schooling, training and learning by doing. As above, in the steady state the economy's stocks of physical and human capital grow at a common endogenous rate, which sustains economic growth. As specified the model admits no distortions between private and social values, so growth and investment in human capital are socially efficient. There is no efficiency argument for government participation in human capital production.

Yet publicly financed education is near-universal, at least in the lower grades in many countries but all the way through college in others. An efficiency rationale for government participation in education rests on the possibility of positive externalities: do individual decisions to acquire human capital create external benefits for others? For example, it is plausible that one individual's human capital is more productive when other members of society are more skilled. As we noted above, the benefits of such complementarities will be internalized when they occur within firms, but not if they are produced by social and other interactions that are external to firms. There are many ways to model these interactions (see Acemoglu (1996), Jovanovic and Rob (1989), or Glaeser (1999), for examples) which need not concern us here. The possibility of such external effects of human capital motivated Lucas (1988) to study a reduced-form extension of (3) in which output of each firm depends on the human capital of its workers and also on the average human capital of workers in the economy as a whole.

The hypothesis of human capital externalities is not easy to test—it requires evidence that the social return to a "unit" of human capital is different from the private return. If we take schooling as our prototypical measure of a human capital component, then Mincerian estimates of the private return to schooling investments abound. Further, the consensus of surveys of this literature indicates that the "true" return to schooling is not much different than what is found from an OLS regression of individual log earnings on years of completed schooling, which would put the private return to an additional year of schooling on the order of 5-8 percent, though recent returns in the U.S. are slightly higher.[6] Given this base for the private return, the question of interest here is whether the social return to an additional year of average schooling—the "Macro-Mincerian" return—substantially exceeds the "Micro-Mincerian" private return that is estimated from the wage and schooling outcomes of individuals. And of course the opposite is possible: the Macro-Mincer return might be smaller than the micro return—a negative externality—as in signaling models (See Section 3, below).

To put some structure on this question in an empirical model of growth let the stock of human capital be $H = hL$ where h is human capital per worker while A is the state of labor-augmenting technology. With Cobb-Douglas production output per worker in country j follows

$$\ln y_{jt} = \alpha_j \ln k_{jt} + (1-\alpha_j)\ln h_{jt} + (1-\alpha_j)\ln A_{jt} \tag{4}$$

where lower case letters denote per-worker quantities. Now exploit the observation that the capital/output ratio is approximately constant, as would occur with a perfectly elastic supply of capital. Then $\ln y_{jt} - \ln k_{jt} = \tilde{\kappa}_j$ and appropriate substitution yields

$$\ln y_{jt} = \kappa_j + \ln h_{jt} + \ln A_{jt} \tag{5}$$

where $\kappa_j = \dfrac{\alpha_j}{1-\alpha_j}\tilde{\kappa}_j$ is a country-specific constant. According to (5) output per worker grows in proportion to human capital h and labor-augmenting technical knowledge, A. Consistent with the form of human capital earnings functions since Mincer (1962), let the

[6] See Card (1999) for a comprehensive survey of earlier literature.

human capital of person *i* in country *j* and in period t satisfy

(6)
$$\begin{aligned}\ln h_{ijt} &= X_{ijt}\beta^P + u_{ijt} = X_{jt}\beta^P + (X_{ijt} - X_{jt})\beta^P + u_{ijt} \\ &= X_{jt}\beta^P + z_{ijt}\end{aligned}$$

where X is a vector of human capital determinants such as schooling and experience and X_{jt} is the mean of *X* in *j* at *t*. The parameters β^P measure the private returns to a unit increase in *X* on an *individual's* stock of human capital—the Micro-Mincer return in reference to schooling. Aggregating (6) over the labor force, log human capital per workers is

(7)
$$\ln h_{jt} = X_{jt}\beta^P + \ln\sum_i \exp(z_{ijt}) = X_{jt}\beta^P + \xi_{jt}$$

Notice that ξ_{jt} depends on the distribution of human capital in the workforce. Since z_{ij} is the deviation of an individual's human capital from the economy-wide mean, and ξ is convex in z, a mean-preserving spread of the distribution of human capital increases ξ and hence $\ln h$.[7]

To complete the model we need to incorporate Lucas' (1988) notion of human capital externalities. One form of this hypothesis is that greater amounts of measurable human capital—say schooling or experience—raise total factor productivity. So let

(8)
$$\ln A_{jt} = X_{jt}\beta^E + a_{jt}$$

where β^E measures the extent of human capital externalities. Inserting (7) and (8) in (5) yields

(9)
$$\ln y_{jt} = \kappa_j + X_{jt}\beta^S + \xi_{jt} + a_{jt}$$

[7] This raises obvious econometric issues. When we apply the model to the returns to schooling, the effect of schooling on productivity will be biased down (up) if an increase in average schooling reduces (increases) overall inequality of human capital.

where $\beta^S = \beta^P + \beta^E$ is the social impact of human capital measures on output per worker. Then the empirical question of (positive) human capital externalities comes down to whether $\beta^S > \beta^P$: e.g. does a unit increase in *average* years of schooling raise aggregate productivity by more than the private return?

Topel (1999) estimates various forms of (9) using an unbalanced panel of 111 countries at 5 year intervals between 1960 and 1990.[8] Tables 1 and 2 summarize his estimates, in which the only measure of human capital per worker is average years of schooling. Table 1 applies a fixed effects estimator to (9), so the estimates of the impact of a year of additional average schooling are generated by within-country variation in productivity and educational attainment of the workforce. In the estimates that contain year effects (column 3)—inclusion of which seems appropriate—the estimated social return is 0.10 per year of schooling. This is somewhat higher than the typical estimate of private returns, but given the quality of the data and the lack of other controls we are reluctant to interpret this as firm evidence in favor of $\beta^S > \beta^P$. On the other hand, Table 1 provides little comfort to those who would argue that social returns are smaller than private ones, as in signaling models.

[8] Output and productivity data were from the Summers-Heston Mark 5.6 (1995) files, while information on educational attainment of the labor force was collected by Barro and Lee (1993).

Table 1 **The Effects of Education on Labor Productivity** **Fixed Country Effects, 1960-1990 (*N*=719)**				
	(1)	(2)	(3)	(4)
Avg. Years of Schooling	0.23 (0.010)		0.10 (0.016)	
Avg. Years of Primary Schooling		0.20 (0.019)		0.06 (0.029)
Avg. Years of Secondary Schooling		0.28 (0.037)		0.14 (0.024)
Country Effects	Yes	Yes	Yes	Yes
Year Effects	No	No	Yes	Yes
R^2	.46	.46	.58	.59

Notes: standard errors in parentheses. Based on Summers-Heston Mark 5.6 and Barro-Lee (1993) data.

The fixed effects (deviations from means) estimates in Table 1 are not an explicit model of economic growth. Take first differences in (5) and (7)

(10) $$\Delta \ln y_{jt} = \Delta X_{jt}\beta^{P} + \Delta\xi_{jt} + \Delta \ln A_{jt}$$

The last term in (10) is growth in total factor productivity. In Lucas' formulation of externalities the level of productivity depends on the average level of human capital per worker, so a surge in investment in human capital would lead to a one-time surge in productivity. But it is equally plausible that the *level* of human capital affects growth, as suggested by Nelson and Phelps (1965). In their model skilled workers are more likely to innovate new technologies and more capable of adopting existing technologies to local production. Further, as noted by Barro and Sala-Martin (1997) the opportunities to grow may be greater for economies that are inside the technological frontier, which they term "convergence". We represent these ideas as

(11) $$\Delta \ln A_{jt} = \Delta X_{jt}\beta^{E} + X_{jt}\delta_{X} + \ln y_{jt}\delta_{y} + X_{jt} \ln y_{jt}\delta_{Xy} + \Delta a_{jt} \text{ where } \delta_{X} > 0$$

if the level of human capital is a boon to growth and $\delta_{Xy} < 0$ if the impact of human capital on growth is greater in less advanced countries.

Table 2
The Effects of Education on Productivity and Growth
First-Difference Estimator at Various Growth Intervals
(dependent variable: Δy_{jt})

	5-year growth (N=608)	10-year growth (N=290)	15-year growth (N=186)	20-year growth (N=101)	5-year growth, fixed effects (N=604)	10-year growth, fixed effects (N=290)
Δ Education: ΔX_{jt}	0.115 (0.022)	0.115 (0.022)	0.155 (0.030)	0.246 (0.043)	0.022 (0.017)	0.086 (0.030)
Years of schooling: X_{jt}	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)	0.004 (0.001)	0.004 (0.003)	0.009 (0.004)
Ln output/worker: $\ln y_{jt}$	-0.004 (0.003)	-0.004 (0.003)	-0.005 (0.003)	-0.009 (0.004)	-0.043 (0.007)	-0.047 (0.008)
$\Delta X_{jt} \times \ln y_{jt}$	-0.060 (0.022)	-0.060 (0.022)	-0.041 (0.032)	-0.025 (0.044)	-0.020 (0.016)	-0.049 (0.025)
R^2	.332	.332	.391	.399	.287	.493

Notes: standard errors in parentheses. Based on Summers-Heston Mark 5.6 and Barro-Lee (1993) data. All models include year effects. Effects of ΔX_{jt} are evaluated at the mean level of $\ln y_{jt}$.

Estimates based on (11) are shown in Table 2, again taken from Topel (1999), who estimates the effects of education on growth at various growth intervals. With measurement error and serial correlation in schooling there is an econometric tradeoff: short growth intervals increase sample size, but admit larger downward bias in estimated returns due to measurement error. Topel's estimates are consistent with this: at long growth intervals (e.g. 15 or 20 years) the estimated impact of education on productivity is much larger than at short intervals (5 years). At a 20 year growth interval the estimated impact of one year of average schooling on productivity is .246, which is vastly larger than the typical private returns to schooling estimated from micro data. And unless human capital externalities are truly grand, it also implausibly large as an estimate of the social return to schooling, β^S.

These estimates do not control for other elements of human capital than schooling, and it is reasonable to assume that growth in other forms of human capital is correlated with growth in schooling. That is, if investment in education is worthwhile, then investments in other forms of skill acquisition are also likely to be profitable. Then the impact of education on growth will be biased up for the usual omitted variables reasons. Given the quality of the data, measurement of these variables is infeasible. An alternative is to assume that unmeasured elements of human capital evolve at a constant rate within each country, which adds a fixed country effect λ_j to (11). With this assumption the estimates are unaffected by correlation between innovations to education and unmeasured factors in λ_j. A limitation is that the estimator can only be applied to fairly short growth intervals—5 or 10 years in the available data—which increases the importance of measurement error in recorded schooling. With this limitation in mind, the last two columns of Table 2 show the results of applying this “diffs-in-diffs” methodology to growth data. At a 10-year interval the main effect of schooling on productivity is 0.086 per year, which is near the top of the range of estimates of private returns typically estimated from micro data.

Two recent studies (Krueger and Lindahl (2002) and de la Fuente and Donenech (2000)) focus on the importance of measurement error in aggregate education measures for estimating Macro-Mincer returns to schooling. Table 3 reports the correlations between the two most prominent measures of education in the literature, produced by Barro and Lee (1991) and by Kyriacou (1991). They report average years of schooling for 68 countries in 1965 and 1985, based on differing methodologies. Both data-sets pick-up the large differences in education between less and more highly developed countries, reflected in the high correlations between the contemporaneous measures of education in 1965 and also in 1985. But for specifications like (11) that rely on first-differences the correlations in *growth* of education are decisive. The correlation between growth in average years of schooling between 1965 and 1985 is fairly low (.34). For shorter time-horizons the correlation is likely to be even smaller. This clearly underlines the importance of measurement error in the education data and motivates Krueger and Lindahl (2000) to instrument the Barro-Lee measure of schooling with Kyriacou’s measure of the same thing. The resulting point estimate[9] of the Macro-Mincer return to schooling (0.069) implies a private return to schooling of 11.5% if we assume that human capital has a share in GDP of 60%. The standard error for this estimate is large, however, and Krueger and Lindahl are unable to reject the hypothesis of zero returns at conventional levels significance levels.

Table 3
Correlation between Barro-Lee's (1991) and Kyriacou's (1991) measures of schooling as reported by Krueger and Lindahl (2000)

	Barro-Lee (1965)	Barro-Lee (1985)	Kyriacou (1965)	Kyriacou (1985)	ΔBarro-Lee	ΔKyriacou
Barro-Lee (1965)	1.00					
Barro-Lee (1985)	0.97	1.00				
Kyriacou (1965)	0.91	0.92	1.00			
Kyriacou (1985)	0.81	0.86	0.88	1.00		
ΔBarro-Lee	0.23	0.46	0.36	0.51	1.00	
ΔKyriacou	-0.12	-0.03	-0.17	0.33	0.34	1.00

Table A1 from Krueger and Lindahl (2000).

So what do we learn from macroeconomic data on growth and average educational attainment? Our reading is that the evidence is inconclusive. There is no plausible evidence that the social returns to education are *smaller* than the private returns, which one might take as evidence inconsistent with important signaling effects in the returns to schooling (see Section 3). The Macro-Mincer estimates tend to be at the upper end of the usually reported 6-9% range for private returns to education, and are often substantially higher. Yet the macroeconomic evidence for positive educational externalities is at best weak (see also the chapter by Lant Pritchett in this *Handbook*). In part this weakness is due to the limitations of growth accounting data. Measures of educational attainment are not typically comparable across countries, educational

[9] See Krueger and Lindahl (2000), Table 5, column (5).

attainment may be measured with substantial error, and other forms of human capital remain unmeasured, among other shortcomings. This has led some researchers to seek evidence of excess social returns in more traditional (for labor economists) sources of micro data in the U.S. We now turn to that approach.

2.B. Evidence from Local Data: States and Cities

Several studies (Rauch 1993; Acemoglu and Angrist, 1999; Ciccone and Peri 2002; Moretti 2003, 2004) have sought evidence of human capital externalities from the spatial distribution of wages in the United States. These studies examine the effect of variation in aggregate measures of education at a local level on wages. The presumption is that production externalities of education increase individuals' marginal product and by extension their wages. A prototype empirical specification in these studies is

$$w_{li} = X_{li}B + S_{li}\beta^P + \overline{S}_l\beta^E + \varepsilon_{li} \tag{12}$$

where w_{li} is the log wage of individual i in local market (e.g. state or city) l, X_{il} is a vector of controls, and S_{li} is person i's years of completed schooling. Then β^P is the Micro-Mincer return to schooling. The twist introduced by the cited studies is to include $\overline{S}_l$, the average years of schooling in market l, in the regression. $\overline{S}_l$ is intended to pick up the effect of human capital externalities—a more educated local labor force increases local productivity. This impact on productivity raises the price of land as firms choose where to locate, and it raises wages because mobile labor must be indifferent among locales. This literature attempts to control for unobserved factors that may be correlated with $\overline{S}_l$ in a variety of ways, mainly involving the use of instrumental variables, which we discuss below.

In what follows we outline a spatial equilibrium model of local wage determination, based on Lange and Topel (2004). This model serves two purposes. First, it guides our review of the empirical studies of educational externalities based on micro data, cited above. Each of these studies appeals to the theoretical framework formulated by Roback (1982) to justify the empirical model (12). Looking ahead, our view is that these studies give insufficient weight to endogeneity issues implied by a spatial equilibrium, so that the econometric methods they apply are unlikely to be valid. Second, the model guides our own attempts to identify the determinants of long term growth in American states, and the role of education in that process. We take up these issues in Section 2.D.

2B.1 A Model of Spatial Equilibrium in Labor Markets

In Roback's (1982) spatial model land is both a consumption good and a productive input. Firms' and individuals' location decisions are made conditional on rental prices of land as well as wages in different places. In equilibrium prices leave individuals and firms indifferent between locales, so that local externalities are reflected both in rental prices and wages. Production externalities increase both. The effect on land prices is usually ignored in the empirical literature[10], citing lack of reliable and comparable data across areas.

We consider a labor and product market equilibrium defined over a large number of locales, *l*. Output in *l* is tradable across locales for a competitive price and, consistent with the growth models outline above, it is produced by capital, land (*M*) and human capital (*H)* with constant returns to scale. We assume that capital is in perfectly elastic supply, so it can be ignored in what follows. Local output Y_l is then

$$(13) \qquad Y_l = M_l^{\alpha}\left(\prod_{s=1}^{S}\left(A_{ls}H_{ls}N_{ls}\right)^{\delta_s}\right)^{1-\alpha}$$

where $\sum_s \delta_s = 1$, N_{ls} is the number of workers of skill-type s who live and work in *l,* and H_{ls} is the average amount of human capital that workers of type-s in locale l bring to the task and M_l is land used in production in *l.* The nested Cobb-Douglas specification implies that δ_s is the share of labor income accruing to type-s workers. The restriction that δ_s is fixed across locales is easily relaxed (δ_{ls}), but the Cobb-Douglas assumption of

[10] A notable exception is Rauch (1993).

constant shares is a bit more serious. With this technology an increase in the productivity of type-*j* labor ($dA_{lj} > 0$) leaves its income share unchanged because all inter-group elasticities of substitution are unitary ($\sigma_{jk} = 1$), so employment of type-*j* labor is unchanged. Relaxing this adds some complexity without much corresponding benefit in terms of additional insight, so we maintain the Cobb-Douglas assumption in our exposition.

It will be convenient to represent the production side of the market in terms of its dual, the unit cost function for locale *l*'s output. Denoting this unit (marginal) cost by C_l and normalizing the price of the tradable good to unity, product market equilibrium requires

(14) $$C_l = \left(\frac{R_l}{\alpha}\right)^{\alpha}\left(\prod_{s=1}^{S}\left(\frac{W_{ls}}{(1-\alpha)\delta_s A_{ls}}\right)^{\delta_s}\right)^{1-\alpha} = 1$$

In (14), R_l is the rental price of land at location *l* and W_{ls} is the rental price (wage) of one unit of skill type *s* in that market.

Workers of type *s* are mobile across locales. We assume that utility depends on consumption of local amenities, units of the tradable good and land. Let the utility of individual *i* of type *s* in locale *l* take the form:

(15) $$U_{lsi} = V_{lsi} Z_{lsi}^{\theta} M_{lsi}^{1-\theta}$$

where V indexes utility from local amenities, *Z* is consumption of the composite tradable good, and *M* is land. The budget constraint is

(16) $$Z_{lsi} = H_{si} W_{ls} - R_l M_{lsi}$$

where H_{si} is the type-*s* human capital of person *i*, whose observed wage is $W_{lsi} = H_{si} W_{ls}$.

We can think of H_{si} as unobserved talent or the quality of the individual's human capital—for example, if *s* indexes labor in different schooling and experience cells, then H_{si} is human capital that is not directly measured by the index. With this definition the indirect utility of person *i* in location *l* is:

$$U(V_{lsi}, H_{si}, R_{l,} W_{ls}) = \theta^{\theta}(1-\theta)^{1-\theta} H_{si} W_{ls} R_l^{\theta-1} \tag{17}$$

In spatial equilibrium the marginal worker of type *s* must be indifferent between living and working in *l* and in the best alternative locale. We treat this reservation value as constant for skill type-*s*, so in equilibrium:

$$V_{ls} W_{ls} R_l^{\theta-1} = U_s \tag{18}$$

where V_{ls} is the value of local amenities for the *marginal* worker of type-*s*. In writing this condition we have made use of the fact that the "quality" of person *i*'s human capital, H_{si}, raises utility by the same proportion in all locations, so *H* does not appear in (18).

If all individuals value local amenities identically, then V_{ls} is a constant. Then the supply of type-s skills to a locale is perfectly elastic. Heterogeneous tastes for living in a locale will generate a rising supply price of skills. In this case the marginal V_{ls} will vary with local conditions—for example, a surge in demand for type-*s* skills will reduce V_{ls}. This point plays an important role in our subsequent discussion of empirical evidence—most instruments for education that have been proposed in the literature will also affect V_{ls} in a systematic and predictable way.

The cost and utility conditions (14) and (18) are sufficient to characterize the spatial distribution of skill prices and land rents. Adopt the notational convention that lower-case letters represent natural logarithms. Then taking logs in (14), spatial competition in production of the tradable good yields an "indifference" relation for unit cost:

$$w_l = \lambda + a_l - \frac{\alpha}{1-\alpha} r_l \tag{19}$$

where $w_l = \sum_s \delta_s w_{ls}$ is the income-share weighted average of log skill prices and $a_l = \sum_s \delta_s a_{ls}$ is total factor productivity (TFP) in market *l*. Equation (19) must be satisfied for employers to operate in market *l*—wages in market *l* can be higher the greater is TFP, the greater is worker quality, or the lower the price of land.

Taking logs in (18) yields a family of indifference relations between skill prices and land rents for workers of each type:

$$w_{ls} = \mu_s + (1-\theta) r_l - v_{ls} \tag{20}$$

To retain workers in *l*, a higher price of land in *l* must be offset by greater wages, while more valuable local amenities reduce wages, holding constant the price of land. Now form $w_l = \sum_s \delta_s w_{ls}$ in (20) and solve for the (log) price of land:

$$r_l = \frac{\lambda - \mu + a_l + v_l}{1-\theta+\phi} \tag{21}$$

where $\phi = \alpha / 1-\alpha$ is the ratio of cost shares of land (α) and human capital ($1-\alpha$) and $v_l = \sum_s \delta_s v_{ls}$ is the average valuation of location *l* amenities by the marginal worker of each type. According to (21), the price of land in *l* will be higher (a) the greater is total factor productivity in *l*, a_l or (b) the greater the value of local amenities in *l*, v_l.

Inserting (21) into (20) yields a solution for the price of type-*s* skills in *l*:

$$w_{ls} = \mu_s + \gamma\left(a_l + v_l\right) - v_{ls} \tag{22}$$

where we have absorbed constants into μ_s and

(23) $$\gamma = \frac{1-\theta}{1-\theta+\phi} \le 1$$

where the inequality is strict when land is a factor of production ($\phi > 0$). According to (23), the log price of type-*s* skills consists of (a) a skill-specific component μ_s that is common to all locales; (b) a locale-specific effect reflecting the impact of total factor productivity (a_l) and the average valuation of local amenities (v_l) on the price of land; and (c) a "supply" shifter (-v_{ls}) that reflects the marginal cost of attracting and retaining type-*s* labor.

The fact that $\gamma < 1$ is important. In country studies, where the dependent variable of interest is per-capita income or the average product of labor, a unit increase in TFP raises log productivity by exactly one unit, by definition. But an increase in local TFP (a_l) raises local wages less than proportionately because greater local TFP also drives up the price of land, and workers are mobile. How big is γ? As a reasonable calibration, let the cost share of human capital be α_H=.60, based on national income accounts. Then physical capital and land together account for $\alpha_K + \alpha_L$=.40 of cost. If land accounts for a quarter of this, then $\phi = \frac{\alpha_L}{\alpha_H}$=1/6. If housing accounts for 1/3 of a typical household's expenditures[11], and land is half of that, then $\gamma = \frac{1/6}{1/6+1/6}$=.50. So an event that raises local productivity by one percent would raise wages by half that amount. Put the other way around, if econometric evidence indicates that a locale-specific productivity shifter raises wages by *x* percent, then it must raise local TFP by 2*x*. This will prove important in interpreting econometric evidence on the magnitude of human capital externalities.

Equations (21) and (22) characterize a spatial equilibrium of land and labor markets. We can express the observed log wage of individual *i* as $w_{lsi} = w_{ls} + h_{si}$. Then

[11] Lucas and Rossi (2002) calibrate the share of land in consumption and production based on evidence by Casselli and Coleman (2001) and Roback (1982) to equal 0.1 and 0.05 respectively. These parameter values imply $\gamma \approx 0.66$, slightly higher than the value used in our discussion here.

using (22):

(24) $$w_{lsi} = \mu_s + \gamma a_l + (\gamma - 1)v_l + h_l + u_{lsi}$$

where v_l and h_l are the within-locale means of v_{ls} and h_i, respectively.

Equation (24) is in the form of (12), and it is the foundation for our interpretation of the results of studies that attempt to estimate educational externalities from cross-sectional survey data. To put this in familiar form let μ_s represent the systematic component of a standard human-capital earnings model—including controls for an individual's years of schooling, experience and the like—the parameters and form of which need not concern us here. The issue at hand is how *market-wide* measures of human capital affect total factor productivity in *l*. To this end, let E_l denote human capital measures (average years of schooling, for example) for location *l* that may generate social returns through an impact on total factor productivity. To represent this, let TFP in *l* be

(25) $$a_l = E_l \pi^E + a_l^0$$

where π^E represents the parameters of interest—the magnitude of human capital externalities that are revealed through differences in land prices—and a_l^0 represents unobserved components of local TFP differences. Note that a_l^0 will not, in general, be orthogonal to E_l; for example, with skill-biased technological change areas with greater TFP may demand greater amounts of skills. Using (25), individual log wages follow the regression model:

(26) $$\begin{aligned} w_{lsi} &= \mu_s + E_l \beta^E + \varepsilon_l + u_{lsi} \\ \varepsilon_l &\equiv \gamma a_l^0 + (\gamma - 1)v_l + h_l \end{aligned}$$

As noted above $\beta^E = \gamma \pi^E$: the effect of E_l on *wages* is strictly smaller than its effect on *productivity*.

What can one learn about human capital externalities from econometric estimates of β^E? As (26) is essentially a comparison of wages across areas, the empirical question is whether areas with greater levels of E_l also have higher average wages and, if so, why? The unfortunate fact is that equilibrium outcomes in labor markets rarely provide clean "natural experiments", and the situation here is worse than usual. An area can be "human capital intensive" because of either supply or demand factors, and in either case the conditions for a consistent estimator $\hat{\beta}^E$ of β^E are unlikely to be satisfied. To be more precise, let Z_l denote instrumental variables that can be used to impute E_l. These instruments could represent either demand (Z_l^D) or supply (Z_l^S) forces. In the case of ordinary least squares $Z_l \equiv E_l$, but whatever the estimation method consistency requires $p\lim n^{-1}(Z'\varepsilon) = 0$ where *n* is the number of locales in the data. Applying this condition to the individual components of ε_l, the relevant orthogonality conditions are:

$$
(27)\qquad
\begin{aligned}
&(a) \quad p\lim\left(\frac{1}{n} Z_l' a_l^0\right) = 0 \\
&(b) \quad p\lim\left(\frac{1}{n} Z_l' h_l\right) = 0 \\
&(c) \quad p\lim\left(\frac{1}{n} Z_l' v_l\right) = 0
\end{aligned}
$$

The issue is whether conditions (a)-(c) can be plausibly satisfied by some instrument Z_l that predicts E_l.

Condition (a) requires that instruments for local human capital be orthogonal to productivity differences across locales. It isn't hard to see how this condition would fail. For example, suppose there are two skill groups—workers with a high school education (*s=1)* and those with a college education (*s=2)*—and that E_l is average years of schooling of workers in *l*. Assume that high school graduates are equally productive everywhere ($a_{l1} = a_1$) but college graduates are more productive in some locales than in others. If the elasticity of substitution between college and high school labor exceeds 1.0, as most studies suggest[12], then areas with greater productivity of college graduates will have greater TFP (a_l^0) *and* a larger share of college graduates in the local labor force, so E_l is higher. Then $p\lim n^{-1}E_l'a_l^0 > 0$: differences in relative factor demands as represented by a_l^0 are correlated with differences in average schooling levels, so $\hat{\beta}^E_{OLS}$ is biased up. Stated more broadly, if local human capital measures are correlated with local demands for human capital, then estimators for human capital externalities will be biased up—instruments Z_l must be uncorrelated with differences across locales in the *demands* for skill. Obvious candidates that would satisfy *this* condition are things that affect the supply of human capital to an area, Z_l^S , though we argue next that such "supply shifters" imply biases of their own.

Condition (b) requires that *observable* measures (instruments) of local human capital be orthogonal to *unobservable* local human capital. For example, if E_l is average years of schooling the condition will fail if areas with more educated workers also have higher levels of other dimensions of skill, such as higher quality schooling or higher average ability of workers. It is difficult to think of instruments that would get around this—an instrument Z_l that is correlated with measured human capital in an area is highly likely to be correlated with unmeasured human capital too, for *both* supply and demand reasons. For example, if E_l in an area is high for demand (productivity) reasons, it is plausible that the demand for unmeasured components of skill will also be high. Conversely, if E_l is large because of supply factors—say area-specific investments in schooling or a climate that is unusually attractive to educated labor—those supply factors are likely to produce workers with more h_l as well. In both of these cases the estimator $\hat{\beta}^E$ is biased up, and the magnitude of local externalities is exaggerated. In section 2.C. we will show that unmeasured components of human capital are indeed important in generating a large positive (statistical) relation between aggregate wages and education.

Condition (c) relates measured local human capital to the valuation of local amenities for marginal workers, v_l . If the supply of skills to a locale is perfectly elastic

[12] See, e.g., Katz and Murphy (1992) who estimate an elasticity of substitution between college and high school labor of about 1.4. Hamermesh (1986) provides a useful survey of such estimates.

then we can think of $v_l \equiv \sum \delta_s v_{ls}$ as an area-specific constant. In the more general case of rising supply price of skills to an area, factors Z_l^D that affect the demand for skill will be negatively related to v_l (demand shifts pull in workers who put lower value on local amenities), while factors Z_l^S that affect the supply of skill will be positively related to v_l (supply shifts reduce the cost of retaining the marginal worker). In either case condition (c) is unlikely to be satisfied—if E_l varies across locales for demand reasons then an estimator of human capital externalities is biased up, and if E_l varies for supply reasons the estimator is biased down. And, sadly, it is hard to think of an instrument that represents neither demand nor supply differences, but which affects the observed stock of human capital in an area.

In the introduction we referred to a strand of the literature that emphasizes consumption externalities in education. Such externalities will also cause condition (c) to fail and result in biases in the estimation of production externalities of education. If education produces positive consumption externalities, then individuals will be willing to pay for living in cities with high levels of education. Firms can only maintain their unit costs and remain competitive if wages decline. In equilibrium we therefore predict a positive relation between rents and average education and a negative relation between wages and average education due to consumption externalities in education. This represents a fundamental identification problem for disentangling external effects in production and consumption—a problem that is not addresses in the existing literature, nor in our contribution[13].

2.B.2: *Empirical Studies of Schooling Externalities*

With this framework as a guide, we turn to existing estimates of human capital externalities. As noted above, these have the common form of (12), which we repeat here:

---

[13] The spatial equilibrium model presented here however suggests, that in the presence of consumption externalities education should raise housing prices and lower wages. This lack of good data on land prices makes tesing this hypothesis difficult. The findings of Rauch (1993) that we review next are consistent with this hypothesis.

(12) $$w_{li} = X_{li}B + S_{li}\beta^P + \overline{S}_l\beta^E + \varepsilon_{li}$$

The issue is whether the evidence from these studies provides evidence of $\beta^E > 0$.

Rauch (1993)

Rauch's (1993) study is the first attempt to identify human capital externalities in cross-sectional data. An attractive feature of his study is that he estimates effects of productivity shifters on both wages and "land" rents, as implied by a spatial model. Using data on wages and housing rents (imputed for home owners) for individuals in 237 SMSAs, taken from the 1980 U.S. Census Public Use files, he estimates models of the form:

(28) $$\begin{aligned} w_{li} &= X_i^w\beta^w + E_l\beta^{wE} + u_l^w + e_{li}^w \\ r_{li} &= X_i^r\beta^r + E_l\beta^{rE} + u_l^r + e_{li}^r \end{aligned}$$

where the left-side variables are the log wage of individual *i* in location *l* (w_{li}) and log monthly housing expenditures (r_{li}). His measure of E_l is average years of schooling for individuals in SMSA *l*, while the *X*s are standard controls—including an individual's education, experience and so on—that are incidental to our discussion.

Rauch finds that an additional year of SMSA-average education raises wages (β^{wE}) by from 2.8 (se=.016) to 5.1 percent (se=.013). An additional year of average schooling raises "rents" by about 13 percent. An additional year of SMSA-average experience has a small positive impact on wages, but raises rents by about 1.5-2.0 percent. Rauch interprets these results as being consistent with an environment in which higher average human capital raises overall productivity, which is reflected in both wages and land values.

If there are human capital externalities, then the magnitudes of Rauch's wage estimates are not implausible. Yet the econometric conditions necessary to identify such externalities are unlikely to be satisfied by his regression procedure, which fails to ask *why* some SMSAs have more educated workers than others. Specifically, these estimates are likely to overstate the size of local externalities, especially if human capital differences among locales are demand-driven.

Acemoglu and Angrist (2000)

Acemoglu and Angrist's (A-A) implementation of the spatial equilibrium model uses Census data from 1950 to 1990. They define a "locale" as the state a person resides in, and they estimate models of the form:

$$w_{lsti} = \mu_{st} + \delta_l + \delta_t + E_{lt}\beta^E + u_{lt} + \varepsilon_i \tag{29}$$

Similar to Rauch (1993) they measure local human capital by average years of schooling in the state-*l* labor force, so β^E measures the return to average schooling over and above the return to individual schooling, which is embedded in μ_{st}. For their main results they focus on white men between the ages of 40 and 49, and Census cross sections from 1960-1980.

A-A are concerned with two main sources of bias in estimating (29). First, demand-side changes that spur growth and raise wages may also raise the demand for schooling. Then *E* is positively correlated with the residual in (27). Second, "labor productivity and taste for schooling may change at the same time", which they assert may also generate an upward bias. To deal with these sources of bias, they instrument *E* with compulsory schooling laws that were in effect in an individual's state of birth at age 14. As it turns out, roughly 2/3 of persons schooled in a state stay there, so compulsory schooling laws (CSL)—which mandate that all individuals must complete a minimum number of years of schooling—raise average completed schooling in a state. Note that CSLs will raise average schooling by impacting the lower end of the schooling distribution—more students will be required to complete the 10th grade, for example, so this component of the stock of human capital must be an important source of externalities in order for the experiment to make sense. A-A argue that this instrument is "unlikely to be correlated with state-specific shocks since they are derived from laws passed 30 years before education and wages are recorded." CSLs then generate "exogenous variation", as they put it, so that conditions (27) are satisfied. Notice also that their model contains fixed effects for state of residence, δ_l, so identification comes from within-state time series variation in (imputed) average schooling and average wages.

It is true that the component of E that is predicted by CSLs is unlikely to be correlated with current "shocks", but it is not true that this component is econometrically exogenous in the sense of conditions (27). To see this, suppose that supplies of skill types were perfectly elastic, so v_l defined above is fixed. Then differences in CSLs would not predict E_l in a spatial model: if all workers are indifferent among areas and are freely mobile, then the place where human capital was *produced* bears no relation to where it *works*. If an area produced more educated workers then, absent mobility, the returns to education would fall. With mobility, educated labor would migrate elsewhere to equate the returns across areas, and with perfectly elastic supply a "shock" to the local number of, say, high school graduates would not affect the number of high school graduates who reside in the locale in the long run. But A-A find that CSLs *do* predict average education levels 30 years down the road, which means that labor supply is not perfectly elastic. Areas with more stringent CSLs have higher values of v_l (lower costs of retaining workers). In the notation used above, CSLs are supply side instruments, Z_l^S, that reduce the cost of local human capital. So the instrumental variables estimator $\hat{\beta}_{IV}^E$ is biased down because $p\lim(n^{-1}Z_l'v_l) > 0$.[14] Bluntly put, if CSLs "work" as instruments, they must be invalid[15].

Viewed in this light, A-A's estimator might be viewed as a lower bound on human capital externalities. This lower bound is not that informative, however, as their

[14] There can also be an upward bias. Condition (b) requires that instruments be orthogonal to the average quality of human capital. If areas that invest in schooling be requiring more years of schooling also improve school qualities, then this condition may fail.

[15] In a recent paper Lochner and Moretti (2004) exploit compulsory schooling laws to estimate the effect of education on crime rates. They provide evidence that compulsory schooling laws "work" in raising education of different states by reporting F-values of close to 50 for whites (see Lochner and Moretti (2004). While this suggests that indeed compulsory schooling laws predict schooling, closer inspection also shows that the variation predicted is small. The degrees of freedom for the F-test are 3 and approximately 3,000,000 which implies that compulsory schooling laws explain at most 1/200 of a percent of the residual variance in schoolings in the sample. This in turn implies for the study by Angrist and Acemoglu that even small violations of conditions (a)-(c) will result in large biases. We were able to confirm the finding that the variation in schooling explained by compulsory schooling and attendance laws is small using the census data that forms the basis of the empirical results in Section 2.C.

IV estimates of β^E are very close to zero.[16] A-A estimate $\hat{\beta}^E_{IV}$=.004 in their most complete specification, compared to an OLS estimate of .073. A-A interpret their results as indicating that virtually all of the returns to education are private. This may be too pessimistic, for two reasons. First, their findings may indicate that increases in average education that are produced by CSLs reduce the cost of retaining an educated workforce—presumably part of the purpose of such laws—which will cause a downward bias in estimated externalities. Second, as noted above, CSLs increase average schooling by raising completion rates at lower schooling levels. If these generate small or no externalities—say because high school graduates are not the source of new ideas that drive growth—then the effects will be small in any case.

Moretti (2003)

Moretti's (2003) analysis of spatial wage differentials is more ambitious than earlier studies. Using data from the 1980 and 1990 Censuses , he considers a variety of conceptual experiments that typically produce evidence of higher wages in locales with greater aggregate schooling, controlling for the private returns to schooling. He interprets this as evidence in favor of positive human capital externalities. His measure of externality-producing educational attainment is the percentage of the local labor force with college degrees, where "local" means a Metropolitan Statistical Area (MSA). Concerned about the correlation of this index with local demand shocks, he uses two instrumental variables: the age structure of cities calculated from Census data (younger cohorts are more educated) and an indicator for the presence of a land-grant college in the MSA.

Whether these instruments are plausibly orthogonal to unobserved components of productivity and labor force quality is largely a matter of faith, but in our view their ability to satisfy conditions (27) are not apparent.[17] And the magnitudes of "externalities"

---

[16] A two standard error band for their preferred estimate is [-.053,.061], which includes some substantial positive effects, but also admits a substantial range of negative social returns.

[17] All three of (27a-c) may fail. For example, the age structure is unlikely to be orthogonal to productivity differences across locales, especially as younger families are more geographically mobile. Then serial correlation in productivity growth will cause (27a) to fail even in differenced data. Further, areas with more young people, who tend to be more educated, may also have higher *h* when more able quality workers congregate due to complementarities, or when there are cohort effects in the quality of human capital and

they produce strike us as implausibly large. For example, Moretti reports that an a *one percentage point* increase in the share of college graduates in an MSA raises average wages in that locale by about 1 percent, after controlling for the private returns to schooling and other factors. To put this in familiar units, think of its implications for the Macro-Mincer return to schooling. In 1990 the average share of college graduates in MSAs was about .20 (Moretti, 2003, Table 1). For the sake of argument, assume that average years of schooling in the other 80 percent of the workforce is 12 (they are high school graduates, on average). Then it takes a change in the college share of .25 to increase average years of schooling by 1 year. Further, we know that the effect of local human capital externalities on wages is smaller than the external impact on productivity—that is, $\gamma < 1$ in our previous notation. Our back-of-the envelope calculations suggested $\gamma \approx 1/2$, so Moretti's estimates imply a Macro-Mincer impact of a year of average schooling on average productivity on the order of .50. This is more than *five times* the Micro-Mincer private return to schooling. It would be nice if this were true—then education is surely the path to economic development. But we doubt it.

2.B.3: Conclusion: What Have We Learned from Micro-Data?

Estimates of productivity externalities based on augmented Micro-Mincer earnings regressions range from zero (Acemoglu and Angrist) to not-so-implausible (Rauch) to simply huge (Moretti). Combined with Macro-Mincer estimates from the growth literature, we think its fair say that there is little evidence in favor of *negative* external returns to education. This finding alone is useful, as it casts doubt on earlier studies (e.g. Pritchett, 2001; Benhabib and Spiegel, 1994) that argued for small or even zero aggregate returns to schooling. Yet the evidence for positive external returns is weak, at best, and founded on dubious identifying assumptions. The next section attempts to cast new light on these issues by combining the two approaches: we study the

---

schooling, causing (27b) to fail. Moretti attempts to deal with the latter effects by controlling for individual effects within cities in NLSY data, but these estimates do not employ the instruments mentioned in the text. Finally, as in Acemoglu and Angrist (2000), the fact that these instruments predict education at all implies that supplies are not perfectly elastic across locales, so (27c) is unlikely to hold. In a more recent paper Moretti (2004) attempts to directly estimate firms productivity linking data from the Census of Manufacturing with the Census of Population. However, any of the sources of bias summarized in equation (27c) survive since the spatial mobility of firms ensures that wage differences have to correpond to difference in productivity by firms.

growth of wages and productivity in U.S. states between 1940 and 2000, using the spatial model of Section 3.B.1 as a framework.

2.C. Human Capital and Growth Revisited: American States, 1940-2000

This section uses individual level data from the 1940-2000 U.S. Censuses to study wage and productivity growth in American states, summarizing results from Lange and Topel (2004). Our framework for this exercise is the model of spatial equilibrium outlined in Section 3.B.1. In differenced form, this model can be viewed as a standard model of economic growth augmented by mobility decisions that connect geographically disparate markets.

A rough look at the data suggests a (perhaps too) strong relationship between growth in educational attainment and growth in wages. Figures 1a-b graph 1940-2000 growth in state averages of log wages and years of schooling of employed individuals against the initial, 1940, values of these variables. There is no doubt that low-wage (mainly Southern) states led the way in terms of growth, and that states with growing educational attainment also experienced the greatest growth of wages (Figure 1-c). A simple regression of 60-year wage growth on growth in schooling yields a coefficient of 0.21 ($t=8.4$) per year of schooling. This is a big number. Either growth in average years of schooling is correlated with other determinants of productivity or the external benefits of schooling are very large.

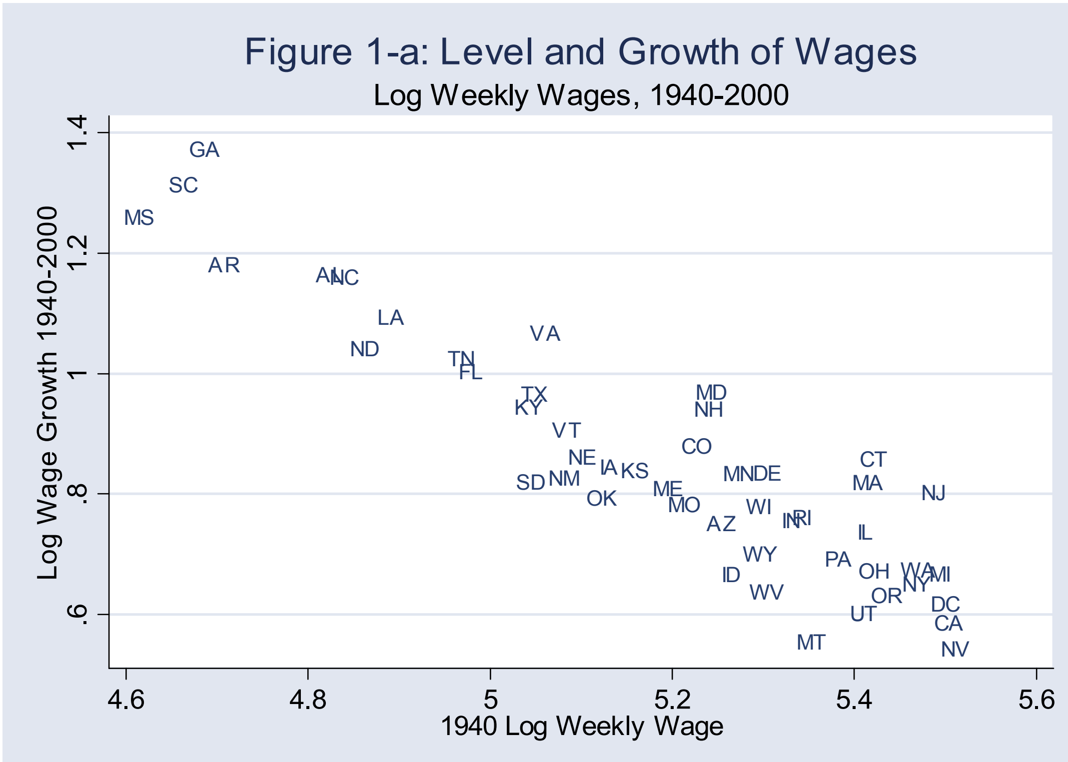


Figure 1-a: Level and Growth of Wages
Log Weekly Wages, 1940-2000
Log Wage Growth 1940-2000
1940 Log Weekly Wage

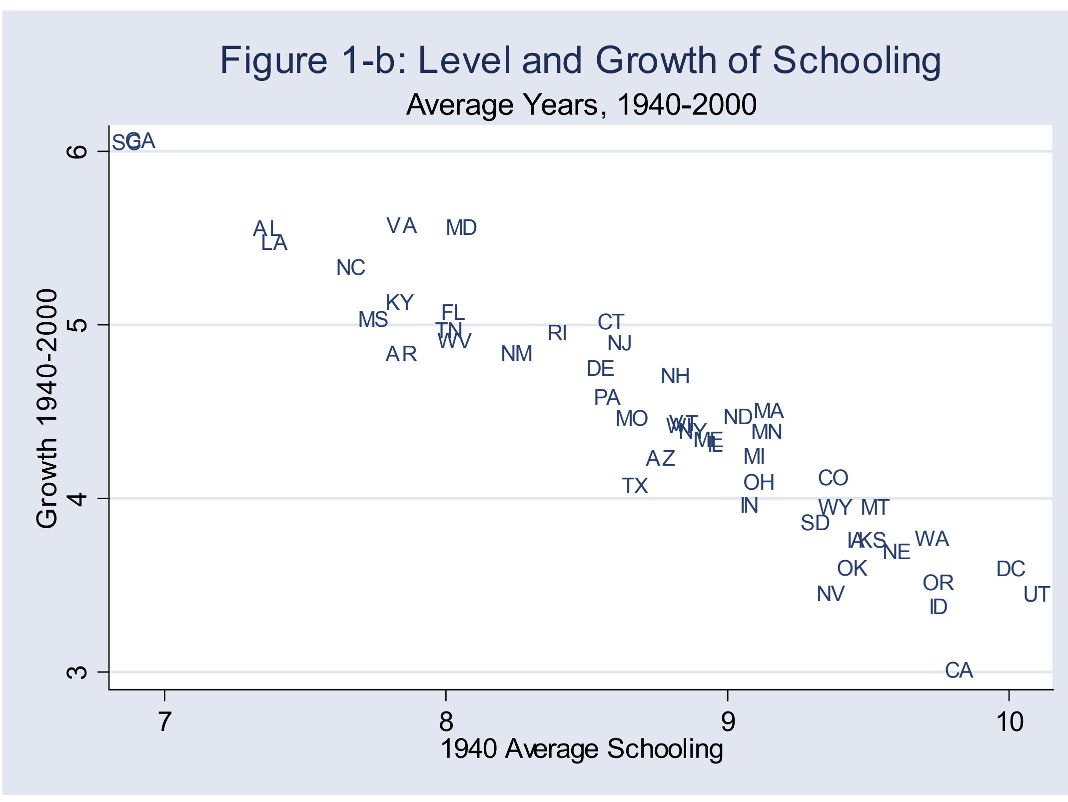


Figure 1-b: Level and Growth of Schooling
Average Years, 1940-2000
Growth 1940-2000
1940 Average Schooling

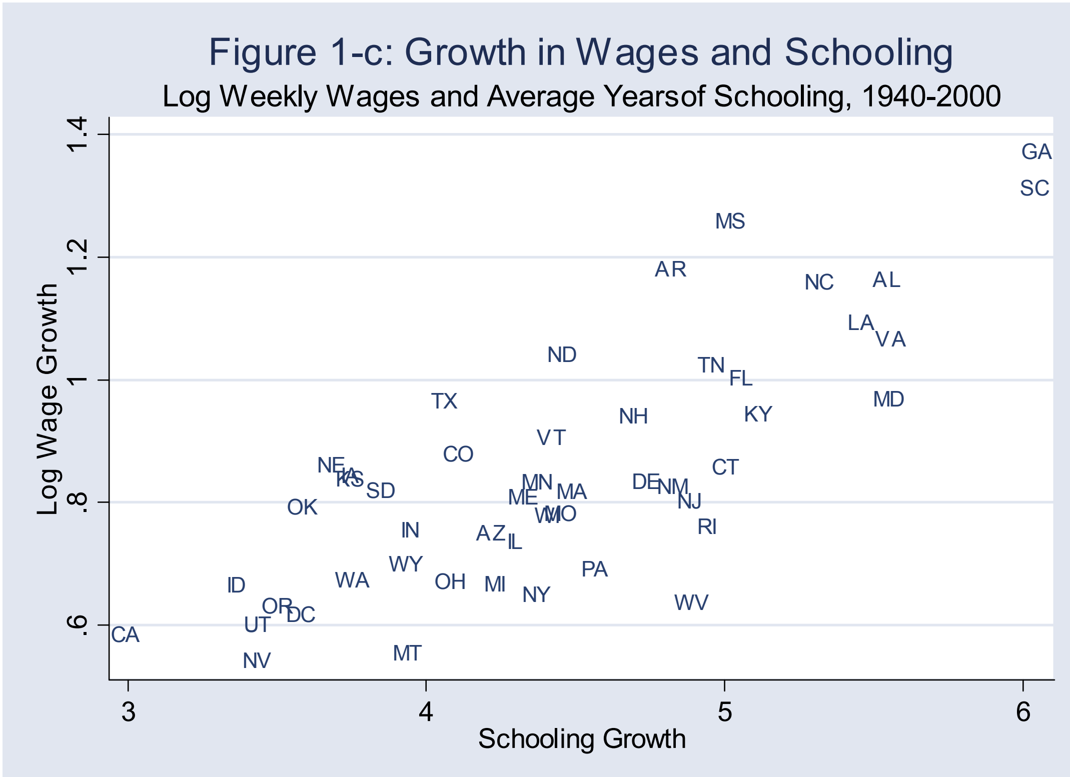


Figure 1-c: Growth in Wages and Schooling
Log Weekly Wages and Average Yearsof Schooling, 1940-2000

We use the model summarized by equation (24) to guide our analysis. In addition to the usual earnings and schooling measures recorded in micro-data, Census files record each respondent's state of birth. To the extent that area-specific differences in schooling quality and other environmental factors experienced while young affect adult productivity, this information provides additional leverage for assessing the average "quality" of human capital in local labor markets. Specifically, augmenting (26) we express the (log) wage of individual *i* in market *l* at time *t* as:

(30a) $$w_{it} = X_{it}\beta_t + T_{lt} + \delta_{bc} + u_{it}$$

(30b) $$T_{lt} = E_{lt}\beta^E + \gamma a_{lt}^0 + (\gamma - 1)v_{lt} + h_{lt}$$

Model (30a) expresses wages in terms of observable human capital controls X_{it} (education and potential experience), state-specific productivity T_{lt}, and a fixed-over-time effect of an individual's birth-state (*b*) and birth-cohort (*c*), δ_{bc}. In turn, state-specific productivity is determined by human capital externalities, unobserved demand (a_{lt}^0) and supply (v_{lt}) conditions and the unobserved quality of state-*l* workers, h_{lt}.

We estimate (30a) by pooling the data over Census years. We adopt an unrestrictive form for the component $X_{it}\beta_t$, with a set of indicators for completed schooling and 5-year intervals of potential experience. We estimate state-specific productivity with fixed effects for each of the 48 contiguous states, plus the District of Columbia, in each year. The birthstate/cohort "quality" effects δ_{bc} are estimated from birthstate-by-cohort indicators, where we define birth cohorts by 10 year intervals; e.g. persons born in Michigan between 1920 and 1930. As experience, year and cohort effects are not separately identified, we impose the restriction

(31) $$\sum_b \delta_{bc} = 0$$

so that the δ_{bc} measure the within-cohort *relative* "quality" of persons born in state *b*.

We estimate (30a) on 7 cross-sections (1940-2000) of Census data. This yields estimates of local productivity differences $\hat{T}_{lt}$ and birthstate-cohort effects $\hat{\delta}_{bc}$, along with the usual cross sectional returns to schooling and experience $\hat{\beta}_t$. Now take means by state and year, and linearize the aggregate returns to schooling and experience. Differencing the data between Census years yields a growth model:

(32) $$\Delta w_{lt} = A_{0t} + \Delta Educ_{lt} A_1 + \Delta Exp_{lt} A_2 + \Delta T_{lt} + \Delta \overline{\delta}_{lt} + e_{lt}$$

The parameter A_1 measures the aggregate return to schooling: conditional on other factors that determine local wages, by how much does an additional year of *average* schooling in a state raise the average log wage in that state? Note that there is no necessary relation between this growth-based estimate of the impact of schooling on average log wages and the "private" return that is typically estimated from cross-sectional data. For example, the extreme form of signaling models of education implies A_1=0—education does not raise aggregate productivity—even if education commands a positive cross-sectional return.

Estimates of model (32) for various growth intervals are shown in Table 4. The estimates in column (1) under each growth interval are analogous to the country-based growth models of Table 2 above, though they are based on average log wages rather than log income per capita. Columns (2) condition on estimates of ΔT_{lt} and $\Delta\bar{\delta}_{lt}$ from the cross-section model (30a). Given the setup of the model, we expect $A_3 = 1$ and $A_4 = 1$: growth in TFP or in the average quality of workers raises wages proportionally. As these regressions must pass through the means of the data, estimates of A_1 and A_2 reproduce the average cross sectional Micro-Mincer returns to schooling and experience. As in the country-based results the estimated social returns to schooling in columns (1) exceed the typical private return, and they are larger the longer is the growth interval used for estimation. For example, at 30-year intervals the Macro-Mincer return to an additional year of average schooling is .187, which is more than double the corresponding Micro-Mincer estimate.

Table 4

Wage Growth in U.S. States: 1940-2000[a]

$$\Delta w_{lt} = A_{0t} + \Delta Educ_{lt} A_1 + \Delta Exp_{lt} A_2 + \Delta T_{lt} A_3 + \Delta\bar{\delta}_{lt} A_4 + e_{lt}$$

	10 year growth			20 year growth[b]			30 year growth[c]			60 year growth	
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)
$\Delta Educ_{lt}$	.117 (.018)	.072 (.006)	.067 (.007)	.166 (.022)	.076 (.006)	.068 (.009)	.187 (.021)	.083 (.006)	.075 (.011)	.203 (.023)	.097 (.006)
ΔExp_{lt}	-.006 (.005)	.017 (.002)	.017 (.002)	.008 (.008)	.019 (.002)	.018 (.002)	.029 (.011)	.019 (.004)	.020 (0.004)	.077 (.021)	.022 (.005)
ΔT_{lt}		1.01 (.020)	1.01 (.022)		1.03 (.021)	1.04 (.028)		1.04 (.031)	1.10 (.052)		1.01 (.032)
$\Delta\bar{\delta}_{lt}$		1.15 (.141)	1.39 (.282)		1.10 (.134)	1.72 (.256)		1.02 (.135)	1.74 (.281)		.780 (.163)
State effects	No	No	Yes	No	No	Yes	No	No	Yes	No	No
R^2	.887	.990	.991	.941	.997	.998	.968	.998	.999	.693	.993

Notes:
a. Estimated using state averages for indicated variables from U.S. Census files, 1940-2000. All models contain year effects.
b. 20-year growth intervals are 1940-60, 1960-80, and 1980-2000.
c. 30-year growth intervals are 1940-70, 1970-2000
Standard errors are estimated taking into account the generated regressor problem arising because the independent variable Δδ is estimated in the first step regression. See Murphy and Topel (1985).

It isn't plausible that the estimated social returns to schooling in Table 4 are merely redistributive, as implied by the pure form of signaling models. The wage bill, and hence productivity, must rise with the average educational attainment of the workforce in order to generate these results. The evidence in Columns 1 shows that aggregate earnings increase in aggregate education. Comparison with Columns 2 and 3 shows that aggregate earnings increase by more with education than private returns. In other words, Table 4 indicates that additional education really does raise individual and aggregate productivities, quite apart from issues of externalities. We have more to say on the plausibility of signaling as an explanation for the private returns to schooling in Section 3.

The evidence in Table 4 indicates that education is an important contributor to economic growth, with social returns that are at least the equal of private ones. But they do not address the externality issue, which is that increases in aggregate schooling raise total factor productivity. In terms of the model, the issue is whether $\beta^E > 0$ in (30b).

Table 5
Regression Estimates of the Effect of Average State Years of Schooling on Total Factor Productivity, U.S. States, 1940-2000

	1940	1950	1960	1970	1980	1990	2000	Pooled 1940-2000	
1.	.071	.086	.094	.059	.071	.093	.093	.079[a]	.092[a]
Levels	(.026)	(.019)	(.022)	(.026)	(.029)	(.046)	(.049)	(.010)	(.010)
2. Growth[a]									
10 years	-.053	.084	.087	-.030	.161	.015	--	.046[b]	.029[b]
	(.048)	(.021)	(.036)	(0.042)	(.061)	(.030)		(.017)	(.021)
20 years	.041	.115	.086	.098	.086	--	--	.067[c]	.050[c]
	(.034)	(.028)	(.026)	(.023)	(.037)			(.019)	(.031)
30 years	.077	.092	.149	.059	--	--	--	.067[d]	.034[d]
	(.032)	(.020)	(.022)	(.017)				(.018)	(.033)
40 years	.057	.140	.107	--	--	--	--	--	--
	(.058)	(.020)	(.019)						
50 years	.100	.117	--	--	--	--	--	--	--
	(.021)	(.017)							
60 years	.081	--	--	--	--	--	--	--	--
	(.021)								
State fixed effects	no	no	no	no	no	no	no	no	yes

Notes:

a. All pooled models contain fixed year effects.
b. Pooled estimates of 10 year growth using 1940-50, 1950-60, ... , 1990-2000.
c. Pooled estimates of 20 year growth using 1940-60, 1960-80, 1980-2000.
d. Pooled estimates of 30 year growth using 1940-70, 1970-2000.

Table 5 shows least squares estimates of β^E based on (30b) from both levels and growth regressions. The cross-sectional estimates in row 1 are consistent with previous findings of Rauch (1993), Acemoglu and Angrist (2000) and Moretti (2003, 2004), all of whom find a positive impact of average schooling on local wages, after controlling for private returns. As $\beta^E = \gamma\pi^E$, the pooled estimate of 7.9 percent implies a huge impact of education on TFP: if $\gamma = .5$ then the external impact of an additional year of average schooling is to raise local productivity by about 16 percent. Estimates at various growth intervals also produce large effects; at 20 years the impact of an additional year of schooling on T_l is .068 and at 60 years it is .081, which also implies an external impact on TFP of about 16 percent.

As in the papers Acemoglu and Angrist (2000) and Moretti (2003, 2004) the issue is whether these effects are caused by externalities ($\beta^E > 0$) or by other omitted factors (a^0, v, and h in our notation) that are correlated with average years of schooling. Previous authors have sought consistent estimators of β^E via instrumental variables techniques, but our view (see above) is that the validity of any instrument that successfully predicts education is dubious at best. So we take an alternative approach. Instead of seeking exogenous variation via an instrument, we seek to measure factors that are omitted from the simple regression relationship between T_l and average years of schooling. We focus on labor force quality, h_l .

Our candidate is $\overline{\delta_l}$ which is a labor force weighted average of estimated birthstate-cohort effects, δ_{bc} , for individuals who work in state *l*. As equation (30a) includes state fixed effects for each Census year, estimates of δ_{bc} are formed from within-state comparisons of wages for people who were born in different states. These comparisons are uncontaminated by externalities, which (according to the theory) affect all wages in a locale by the same proportional amount. Then $\overline{\delta_l}$ is a measure of labor force quality in *l*, which we can take as a measure or correlate of h_l in (30b). In terms of growth, δ_{bc} is fixed over time for a given birthstate-cohort pair, so *changes* in $\overline{\delta_l}$ are generated by changes in the labor force shares of different cohorts.

Figures 2-a and 2-b show the relationships between long run (1940-2000) growth and 1940 levels for T_l and $\overline{\delta_l}$. As with wages and education, estimates of T_l and $\overline{\delta_l}$ converge in the sense that states with low values of T_l (quality) in 1940 experienced the greatest productivity (quality) growth.

Figure 2-a: Level and Growth of TFP

Estimated State Specific Productvity, 1940-2000

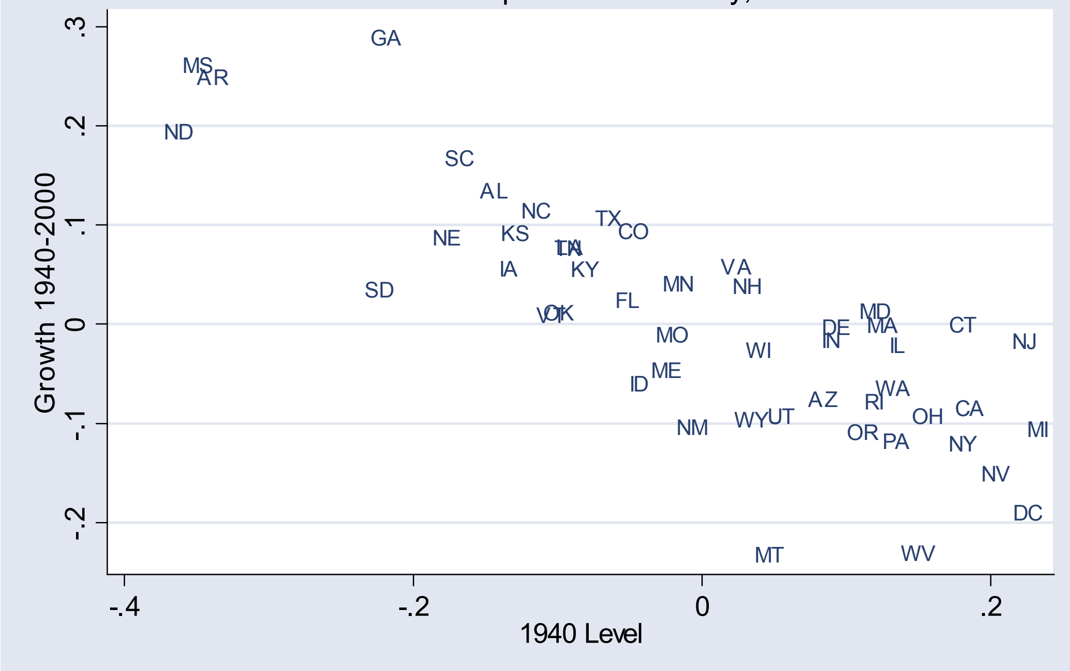
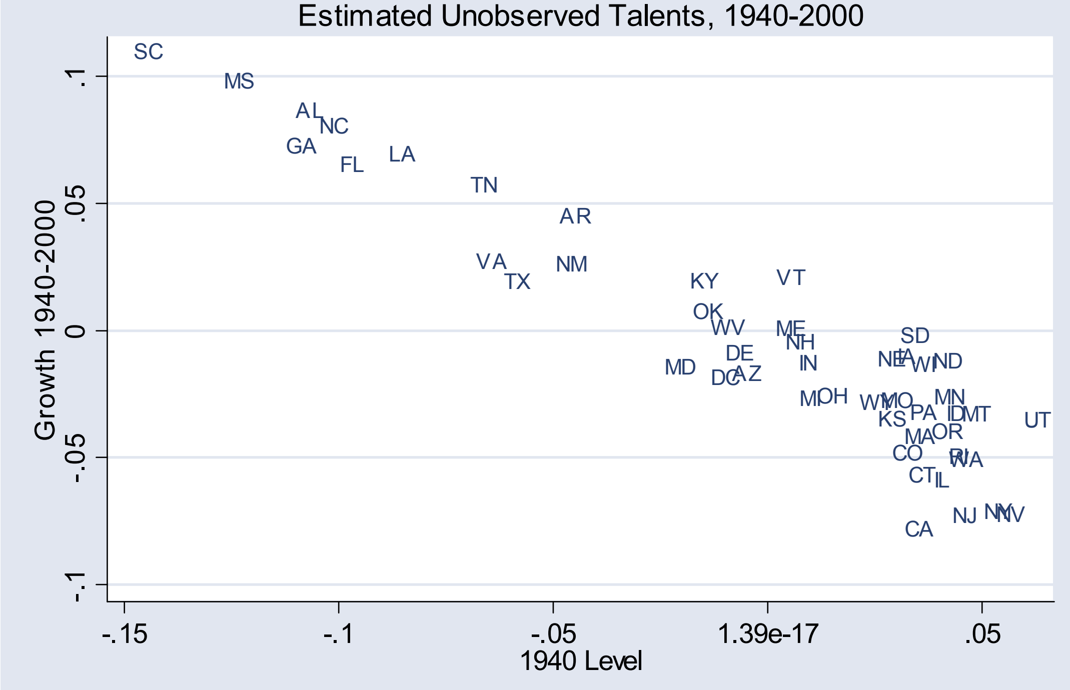


Figure 2-b: Level and Growth of Unobserved Skills

Estimated Unobserved Talents, 1940-2000



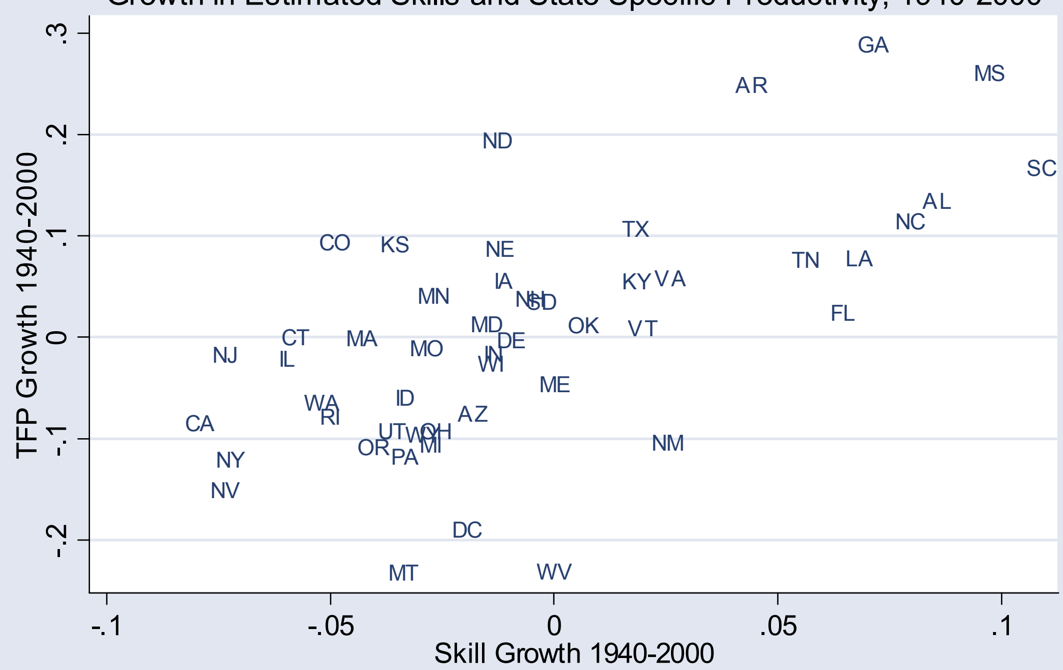
Inspection of the figures reveals that rapid growth on both fronts occurred in Southern states, where relative wages and schooling also grew. Figure 2-c shows that states with growing educational attainment also had greater growth in $\overline{\delta}_l$, and Figure 2-d demonstrates a similar relationship between productivity growth and growth of $\overline{\delta}_l$.

Figure 2-c: Growth in Skills and Schooling

Growth in Estimated Skills and Mean Years of Schooling, 1940-2000

Figure 2-d: Growth in Skills and Local TFP

Growth in Estimated Skills and State Specific Productivity, 1940-2000



These data indicate that states with rapidly growing educational attainment, which experienced growing productivity, also experienced an upgrading in the measured *quality* of their labor forces. As we indicated above, if education and labor force quality go hand-in-hand, a simple regression of changes in productivity on changes in education may "find" externalities where none exist.

How big might the bias be? Table 6 makes some headway on this question by simply adding $\Delta\bar{\delta}_l$ to the growth models of Table 5. For each growth interval we reproduce in Column (1) the corresponding growth estimate from Table 5. All of these estimates of β^E are numerically large, with the biggest effects for the longest growth intervals. For example, the 60 year estimate is .081, suggesting that an additional year of education raises T_l by 8.1 percent. Adding changes in "unobserved" labor force quality $\Delta\bar{\delta}_l$ in Columns (2) reduces the impact of education in each case. For the longest (60 year) interval, the point estimate falls from .081 to .021. None of the Column (2) estimates are significantly different from zero by conventional standards.

Table 6
Education and Productivity Growth: U.S. States 1940-2000[a]

$$\Delta T_{lt} = B_{0t} + \Delta Educ_{lt} B_1 + \Delta\bar{\delta}_{lt} B_2 + e_{lt}$$

	10 year growth		20 year growth[b]		30 year growth[c]		60 year growth	
$\Delta Educ$	.046	.026	.068	.040	.067	.036	.081	.023
	(.017)	(.018)	(.019)	(.021)	(.017)	(.020)	(.021)	(.024)
$\Delta\bar{\delta}_l$		1.23		1.08		1.01		1.35
		(0.44)		(0.40)		(0.34)		(0.38)
R^2	.894	.897	.952	.954	.978	.980	.248	.415

Notes:

a. See notes to Table 5.
b. 20-year growth intervals are 1940-60, 1960-90, and 1980-2000.
c. 30-year growth intervals are 1940-70 and 1970-2000.
d. Standard errors are estimated taking into account the generated regressor problem arising because the independent variable $\Delta\delta$ is estimated in the first step regression. See Murphy and Topel (1985).

The evidence in Table 6 does not demonstrate that externalities are unimportant. For example, the 20-year growth point estimate of .04 implies an external impact on productivity of about .08, which rivals conventional estimates of the private return to schooling. But this evidence surely raises doubts about the importance of externalities, estimates of which are surely overstated by least squares. The evidence is that states with growing productivity and educational attainment also attracted or produced "better" workers, and even a simple measure of labor force quality eliminates up to three-fourths of the alleged relation between education and TFP. In our view, the data do not provide a strong reason to believe in the importance of productive externalities from schooling.

3. Job Market Signaling and the Social Value of Schooling

In the standard Human Capital model (hereafter HC) without externalities the private returns to schooling equal the social returns. The Job Market Signaling (JMS) model provides a competing explanation for the observed private returns to schooling. JMS—in contrast to HC—generates a wedge between private returns and social returns to schooling, as individual returns from signaling do not fully translate into productivity increases at the aggregate level.

In this section we review the available empirical evidence on the JMS and HC models, mindful of the need not only to test, but also to quantify the relative importance of each. As Wolpin (1977) puts it: *"[T]he real issue concerns not the mere existence of one or the other effect, but the extent to which schooling performs each of these roles."* We come to the conclusion that the available empirical evidence is insufficient to achieve either goal. There are few convincing tests of JMS and even less evidence that allows us to quantify the contribution of HC relative to JMS. This leads us to provide some new evidence of our own in Section 3C.1.

In Section 2.C. we presented evidence of a large positive relation of aggregate wages and education between 1940-2000 from US Census data (Tables 4-6) and from cross-national studies (Table 1, 2 from Topel 1999). Most formulations of JMS generate private returns in excess of the social returns to schooling. The fact that observed Macro-Mincer returns are as large as or larger than the Micro-Mincer returns is prima facie evidence against JMS. The empirical findings reviewed in Section 2 therefore constitute important evidence against a signaling explanation of the cross-sectional relation between wages and education. At the very least it raises the standard of empirical proof required to take JMS seriously.

.

The JMS model traces its roots to Spence (1973), who developed an equilibrium model of the labor market with positive individual returns to schooling absent any productivity effects of schooling. Testing JMS against HC is complicated by the fact that both rely on similar assumptions regarding individuals' and firms' objectives. This means that JMS and HC are observationally equivalent with regard to data generated within a single labor market and time-period.

JMS relies on 3 crucial assumptions:

1. Individuals have private information on their productive types.
2. Costs of schooling and productivity of individuals are negatively related.
3. Contracts cannot be written conditional on information that is initially private.

Any model with incomplete information requires a restriction on how agents form expectations. Spence's asymmetric information setup owes its prominence in the literature to the fact that both firm and individual expectations and decisions are fully rational. Agents' expectations are based on the mathematical distributions that pertain in equilibrium (macro-economists would say that agents hold rational expectations). Firms' expectations of unobserved characteristics of individuals are consistent with the stochastic relation between observed and unobserved characteristics obtained in equilibrium. Many of the difficulties[18] of testing JMS can be traced to this consistency requirement.

Spence's original formulation allows schooling to be continuous. This raises the possibility of multiple separating equilibria, each with a different set of employer beliefs about the relation between schooling and unobserved productivity. We assume instead that there are only 2 possible levels of schooling: s=0 or s=1. In this case there is a unique separating equilibrium. The basic structure of the model is summarized in Table 7:

[18] *T he discussion of the difficulties of testing JMS against the Human Capital Model summarized in Tables 7-9 benefited tremendously from a conversation with Robert Willis*

Table 7 The Basic Structure of the Job Market Signaling Model

Individual Type	Marginal Product for schooling=0 and 1	Costs of Schooling Level=1
I	(1,1)	Y
II	(2,2)	y/2

In the pure signaling model schooling does not raise the productivity of either type. Individuals' productive types are not directly observed and the costs of schooling decline with productivity. Firms eventually learn the true productivity of individuals, but it is not possible to write enforceable contracts contingent on this information. We are interested in the separating equilibrium where individuals' types are revealed by their choice of schooling—i.e. where type I workers choose s=0 and type II workers choose s=1. Both types are paid their productivity, which is revealed by their choice of schooling. Condition (33) ensures existence of a separating equilibrium where higher ability individuals attend school and less able individuals do not.[19]

(33) $$1<y<2$$

Compare this signaling model to a human capital model with a similar cost condition. Assume again that there are 2 types of individuals and 2 levels of schooling. Assume that schooling raises productivity by *r* percent. The productivity of a type I with no schooling is 1 and that of type II with no schooling is 2/(1+r). Costs are the same as in Table 7. Table 8 lists the fundamentals of this model.

Table 8 A Human Capital Model

Individual Type	Marginal Product for schooling=0 and 1	Costs of Schooling Level=1
I	(1,1+r)	Y
II	(2/(1+r),2)	y/2

[19] This discussion is extremely simplified. See Riley (1998) for a more detailed discussion of the theoretical considerations, in particular the problem of the existence and the uniqueness of equilibria that arise in more complex models).

Schooling is sufficiently expensive to deter type I, but not expensive enough to also deter type II:

(34) $$r<y<4r/(1+r)$$

Table 9 shows the outcomes for individuals of different types in the JMS and HC models. The first entry corresponds to the signaling model and the second to the human capital model.

Table 9

Data Generated by the Signaling and the Human Capital Model

Type/School	0	1
I	(1,1)	(1,1+r)
II	$(2,\frac{2}{1+r})$	(2,2)

The observed (equilibrium) data are on the diagonal. To distinguish JMS from HC using data on schooling and wages requires access to the off-diagonal cells, which are not observed in equilibrium. Maximizing individuals do not consider the impact of schooling on productivity, but instead consider only the relation between wages and schooling itself. Likewise firms do not inquire about the mechanisms generating productivity differences across schooling levels, but instead are interested only in the relation between individual productivities and what must be paid for them. The decentralizing role of the price mechanism implies that equilibrium models of job market signaling and human capital are observationally equivalent. This simple problem makes testing JMS against the Human Capital model a difficult endeavor, a fact acknowledged by Lang and Kropp (1986): "*[M]any members of the profession maintain (at least privately) that these hypotheses cannot be tested against each other and that the debate must therefore by relegated to the realm of ideology*".

An early paper illustrates the difficulties of testing the JMS-model against the HC-model. Psarachopoulos and Layard (1974) mistakenly argue that JMS implies that the returns to schooling decline with experience as employers learn about individuals' characteristics. This is not a prediction of the JMS model. The assumption that agents' beliefs are rational implies that any wage-schooling gradient reflects real productivity differences. These productivity differences will persist across the life-cycle unless we make additional assumptions external to the JMS-model.

Psarachopoulos and Layard (1974) also propose that JMS predicts excess returns for years of schooling that correspond to degrees, such as 16 for college graduates—diplomas carry special weight in transferring information. They did not observe diploma effects and rejected JMS on that basis. Subsequent empirical work has consistently demonstrated the existence of diploma effects in a variety of countries and time-periods (Hungerfort and Solon (1987), Belman and Heywood (1991), Jaeger and Page (1996), Frazis (2002)), which some interpret as evidence against the HC model and in favor of JMS (Frazis (2002)). We disagree. Instead we interpret diploma effects as evidence that individuals resolve uncertainty about their individual returns to schooling while still in school. Whether these returns are generated by a HC or a JMS model is immaterial. This argument was first made (informally) by Chiswick (1973) and relies on modeling the drop-out decision directly. Then diploma effects are consistent both with both the HC and JMS models.

We next consider 2 papers (Lang and Kropp (1986), Bedard (2001) motivated by the difficulties in testing the JMS against the HC-model. These papers study differences in equilibrium outcomes across segmented labor markets that differ in the structure of their education system. They relate data on differences in regional education systems to data on the variation in the equilibrium distribution of schooling in these regions. The JMS and the HC model differ in their prediction of how a change in the cost structure of education affects the equilibrium distribution of schooling. Roughly speaking, the idea is that agents' decisions in the HC-model are affected by only their own costs of schooling. Variation in the costs of schooling faced by type-*j* individuals does not affect outcomes for type-*k* individuals. This contrasts with JMS where the returns to schooling are determined by the equilibrium distribution of ability types across schooling levels. Then the returns to schooling depend on the schooling decisions of other agents, and variation in the costs of schooling for some agents affect the schooling decisions of *all* agents. If we can identify such variation for some (but not all) schooling levels, then the HC-model predicts that only those agents directly affected will change their schooling decisions. The JMS-model instead predicts that such variation in the costs of schooling will affect the schooling decisions of all agents in the market in a predictable manner. Lang and Kropp (1986) and Bedard (2001) do provide some support for the JMS-model, though they cannot provide guidance on the magnitude of JMS and HC motives.

Finally we consider empirical approaches to testing JMS that exploit the assumption that firms ultimately learn about individuals' productive types. An early body of work (Wolpin (1974, 1977), Riley (1979), Albrecht (1981)) examines the implications of differences in the speed with which employers learn across occupations or industries. The evidence on JMS from this literature is mixed. Riley (1979) finds evidence in favor of the signaling model with employer learning, Wolpin (1974, 1979) and Albrecht (1981) are not supportive. We present new evidence from more recent data that rejects Riley's predictions.

More recently the literature on statistical discrimination and employer learning assumes that the data contain a measure of individual productivity that is hard to observe for firms (Foster and Rosenzweig (1994), Farber and Gibbons (1996), Altonji and Pierret (2001), Galindo-Rueda (2003), Lange (2004)). This assumption generates testable predictions for the interaction between schooling and ability with experience in earning regressions. The literature on employer learning can be understood as a test of the core assumption of JMS that firms statistically discriminate on the basis of schooling. Altonji and Pierret (1997) and Lange (2004) point out that the speed with which employers learn ultimately limits the contribution of JMS to the private returns to schooling. Lange (2004) estimates this speed to be rapid and therefore argues that JMS can only explain a small fraction of the private gains from schooling, even if we maintain the assumptions of the employer learning literature.[20]

3.A. Diploma Effects

The existence of diploma effects ranks among the most persistent empirical

[20] Lange (2004) also points out that the same empirical patterns emphasized in the Employer Learning model can be generated by post-schooling investment behavior and provides evidence in favor of this explanation.

findings in labor economics. Completion of degree years is associated with an increase in wages above that observed for other years, and the distribution of completed years of schooling in the population exhibits spikes at those years. Table 10 demonstrates these 2 facts using the 1998 wave of the NLSY for white males.

Table 10:
Diploma Effects in the Distribution of
and the Returns to Schooling[a]

Highest Grade Completed[b]	Frequency	Return to Grade[c]
≤ 8	2.63%	Omit
9	2.28%	-3.89% *(11.77%)*
10	2.28%	+14.51% *(11.78%)*
11	2.28%	7.39% *(11.80%)*
12	42.32%	+16.18% *(8.54%)*
13	7.47%	+20.11% *(9.59%)*
14	8,85%	-0.23% *(9.51%)*
15	3.25%	-2.07% *(11.16%)*
16	16.74%	+37.51% *(9.28%)*
17	3.67%	-19.15% *(11.06%)*
18+	8.23%	+34.61% *(11.62%)*
Observations	1,446	

Notes:

a. The sample consists of white males in NLSY (year=1999).
b. Schooling is bottom coded at 8 years and top-coded at 18 years.
c. Obtained from regression of log(wage) on schooling and cubic in experience.

The returns to high school (12th grade) and college (16th grade) graduation are 16% and 38% respectively. These 2 years also stand out in the distribution of schooling. 42.45% of the male population graduate from high school but do not continue further in their education. 16.74% terminate their education after 16 years. This compares with less than 9% of the population for any other year of schooling.

Diploma effects are often (Psarachopoulos and Layard 1974; Jaeger and Page 1996; Frazis 2002; Habermalz 2002) presented as evidence for screening theories of schooling. We disagree. Instead we view diploma effects as evidence that individuals face uncertainty about their individual returns to schooling and that this uncertainty is revealed as individuals acquire schooling. Those least capable to profit from schooling drop out before the completion of degree years. Those graduating exhibit larger returns than those who dropped out at lower levels of schooling. This reasoning was informally developed by Chiswick (1973)[21]. Since then, a number of authors (Altonji (1993), Heckman, Lochner and Todd (2003), Keane and Wolpin (1997) and Eckstein and Wolpin (1999)) examined different aspects of sequential schooling choice under uncertainty. We build a simple model in this spirit with the intention to show how individual's schooling decisions can generate (large) diploma effects if individuals learn about their returns while in school. We build a simple model based on this analysis that shows how individual's schooling decisions can generate (large) diploma effects if individuals learn about their returns to schooling while in school.

Suppose that individuals can choose 3 levels of schooling. They might decide not to enter into a degree program (S=0), to enroll but drop out after 1 period (S=1) or to complete the degree program, which takes 2 periods (S=2). The returns to schooling depend on ability, β, which is either high (β_H) or low (β_L). At t=0 an individual knows the probability *p* of being high ability. By attending school the agent discovers at the end

[21] Altonji (1993) analyzes the effect of uncertainty about completion of degrees on observed returns to schooling in a model incorporating choice of college major and learning about tastes and ability while in school. He as well as Heckman, Lochner and Todd (2003) show that uncertainty associated with the returns to schooling can generate option values for education that can substanstially affect the returns to schooling. Keane and Wolpin (1997) and Eckstein and Wolpin (1999) develop quite general models of sequential schooling under uncertainty.

of period 1 the true value of β. He then needs to decide whether or not to complete the program. Individual life-times are infinite and all individuals start with the same initial level of human capital H_0. All individuals also have the same discount rate r. The present discounted value of choosing S=0 is then

(35) $$V(S=0,t=1)=\frac{H_0}{r}$$

An individual who attends school for 1 period and has productivity β_i,>1 i=L,H, has a present discounted value (at S=1) of income given by

(36) $$V(S=1,\beta_i;t=1)=\frac{\beta_i H_0}{r}, \quad i\in\{L,H\}$$

If the agent attends schooling for 2 periods then

(37) $$V(S=2,\beta_i;t=2)=\frac{\beta_i^2 H_0}{r}, \quad i\in\{L,H\}$$

The agent's problem can be solved by backwards induction. At t=1 he chooses S=2 iff

(38) $$V(S=1,\beta_i;t=1)=\frac{\beta_i H_0}{r}<\frac{1}{1+r}\frac{\beta_i^2 H_0}{r}=V(S=2,\beta_i;t=2)$$
$$\Leftrightarrow 1<\frac{\beta_i}{1+r}$$

Imposing the restriction $\beta_H>\frac{1}{1+r}>\beta_L$ implies that all individuals who are of high ability complete their degree and those of low ability drop out. At t=0 individuals only know their individual probability of being of high ability and therefore choose to enroll in the degree program iff

$$\frac{H_0}{r} < E\left[Max\left\{ \frac{\beta_H{}^2 H_0}{r}\left(\frac{1}{1+r}\right)^2, \frac{\beta_L H_0}{r}\left(\frac{1}{1+r}\right) \right\} \mid p \right]$$

(39)

$$\Leftrightarrow 1 < p\left(\frac{\beta_H}{1+r}\right)^2 + (1-p)\frac{\beta_L}{1+r}$$

which defines an optimal cut-off point p* below which individuals do not enroll into the degree program. Observed returns are β_L for those completing one period and $\beta_H{}^2/\ \beta_L$ for those completing two periods. The true return to schooling for agents who complete their degree is β_H, less than the estimated return.

The observed returns for completion of the degree are higher than the returns observed during the initial year of the degree since those individuals that complete the degree earn larger returns to schooling during the entire duration of the degree program. The excess returns observed in the last year captures not just the difference in the return generated during the last year between high and low ability individuals. These observed returns also reflect the greater learning in the initial years of schooling for those individuals who continue schooling rather than drop out.

The process of learning about individual ability also generates the spikes in the distribution of completed schooling observed during degree years. Individuals with p<p* will attend only schooling=0. The fraction of individuals who drop out will be relatively small. This is because these are individuals who ex ante believe they have a high return and only find out ex-post that they actually have low returns to education. If individuals make fairly good predictions about their own ability then only a small portion will choose to drop out before the completion of their degree (making mistakes is a low probability event).

This argument implies that diploma effects do not provide evidence for or against JMS. That individuals are uncertain about their own ability to learn and that this uncertainty is (at least partially) resolved while in school naturally imply the existence of diploma effects as well as of bunching at degree years. We therefore dismiss diploma effects as evidence for or against JMS. We turn to other evidence that might inform us on the existence and size of JMS.

3.B. Testing JMS using segmented labor markets

The value of signaling through schooling depends on the difference in the average productivity by schooling level. In turn average productivity by schooling level depends on individuals' costs of acquiring schooling. JMS therefore implies that if costs vary for one part of the support of schooling, then this will affect the returns of schooling throughout the entire support. As individuals adjust to the changing costs along one part of the support of schooling they affect the returns for adjacent schooling levels. Then changes in the costs of schooling cause ripple effects in the distribution of schooling at all levels, even if these cost changes only apply to part of the schooling support. If one can identify variation in costs for only a subset of the schooling distribution, then one can test JMS by evaluating the changes in the distribution of schooling across the entire support. This idea lies at the heart of the empirical approach in Lang and Kropp (1986) and Bedard (2001).

Lang and Kropp exploit variation in compulsory attendance laws (CALs) across states. Imagine an increase in the compulsory schooling age from 12 to 14 years. HC predicts[22] that this regulatory change will only affect individuals previously constrained by compulsory attendance laws. JMS predicts that the change in regulation will affect the entire schooling distribution. Children who previously left school at age 12 will now remain in school until age 14. These children are less able than those who chose to remain in school until age 14 before the change in the CAL, so the average productivity of individuals who leave school at age 14 declines. This raises the return to schooling at age 14, so some individuals who otherwise would have left will remain in school longer. The JMS implies that the fraction of individuals who remain in school longer than age 14 will increase. By extension JMS also predicts an increase in the fraction of individuals who remain in school past age 15, and so on. Lang and Kropp (1986) find this "ripple effect" in response to increases in CALs across U.S. states enacted during the 1908-1968 period.[23]

---

[22] This relies on the assumption that there are no general equilibrium effects on the wage structure. Lang and Kropp (1986) rely on factor prize equalization to rule out such GE effects.
[23] Recent work on the impact of compulsory attendance laws on the distribution of educational attainment does not confirm this prediction (see Lochner and Moretti 2004, table 4)

An important criticism of this work is that an increase in the compulsory schooling age plausibly reflects an overall increase in the value of education. This would cause a rightward shift of the education distribution, which is consistent with the data. A recent study by Bedard (2001) is not subject to this critique. Bedard examines how the presence of post-secondary education institutions in local labor markets affects the high school drop-out decision. The assumption is that proximity reduces the costs of attending post-secondary education. JMS predicts that lower costs of post-secondary education lead the more capable high-school graduates to continue on, which reduces the average productivity of the remaining pool of high school graduates. This reduces the signaling return to graduating from high school., so the drop-out rate will rise. This is indeed the pattern that Bedard finds, which is robust to the criticism that the presence of a post-secondary institution might indicate a greater value attached to education in the local labor market. The papers by Lang and Kropp (1986) and Bedard (2001) therefore do provide some evidence for sorting in determining schooling choices and the distribution of schooling in the population.

3.C. Employer Learning Models

JMS assumes firms are initially unsure about individuals' productive types. They therefore use schooling to infer individuals' productivity. It seems reasonable to assume that firms learn about individuals productive types as time passes and productivity is observed. This assumption forms the basis to a stimulating and active branch of the empirical literature on JMS.

Two approaches have been used to exploit the idea of employer learning. A number of authors use presumed differences in the ease with which firms can learn about individual productivity between different industries (Wolpin (1977), Riley (1981)) or types of applicants (Albrecht (1979)). More recent contributions test for statistical discrimination and employer learning by assuming that the econometrician can observe a pre-market ability measure, based on test scores, that is not observed by employers. This literature (Foster and Rosenzweig (1994), Farber and Gibbons (1996), Altonji and Pierret (1997, 2001), Galindo-Ruedo (2003), Lange (2004) and others) evaluates how the estimated returns to schooling and ability evolve over time. We will examine this literature first.

We present here a recent formulation by Lange (2004), who extends the analysis of Altonji and Pierret (2001) to estimate the speed with which employers learn about individuals' abilities. The speed of employer learning is crucial for JMS since it limits the time interval during which schooling can be a useful signal of unobserved ability—the presumption is that repeated observations of productive outcomes "reveal" a person's true talents. Then wages reflect new information learned about workers' true talents as time passes. This framework then carries a testable hypothesis about learning and signaling: as labor market experience increases, pre-market signals such as schooling have declining influence on observed wages, while unobserved talents have ever increasing weight. The rate at which this shift occurs is determined by the speed of employer learning. Following Lange (2004) we demonstrate how to estimate the speed of employer learning from the set of OLS-coefficients on schooling and a pre-market ability measure, fully interacted with experience, in an earnings regression. And we show how this estimate and the first order condition for choosing schooling can be used to bound the contribution of signaling to the returns of schooling.

The key assumption for what follows is that a source of panel data on individuals records a pre-market ability measure z_i that is not directly observed by employers[24]. Assume further that the log of individual productivity x_i depends linearly on z_i, years of completed schooling s_i, and on information that is observed by employers but that is not recorded in the data, q_i. Adding an individual effect u_i, productivity follows:

(40) $$x_i = \alpha_0 + \alpha_1 s_i + \alpha_2 z_i + \alpha_3 q_i + u_i$$

It will be convenient to write $\tilde{x} = x - (\alpha_0 + \alpha_1 s + \alpha_3 q)$ and to suppress individual subscripts from now on. $\tilde{x}$ is the component of individual productivity that is not directly observed by employers. Following the JMS framework, we suppose this component to

[24] Altonji and Pierret used the AFQT-score, father's education and siblings' wages in the NLSY. Their main analysis proceeds with the AFQT-score. Test-scores have also been used in UK and German data (Galindo-Rueda (2003), Bauer and J. Haisken-DeNew (2001)). In an interesting paper Foster and Rosenzweig (1995) use earnings of rural workers observed in a period when these were paid piece rates.

be private information of individuals—there is asymmetric information.[25] Firms use s and q to make predictions about $\tilde{x}$. Assume that the conditional expectation of $\tilde{x}$ is linear:

(41) $$E\left[\tilde{x} \mid s,q\right] = \beta_1 s + \beta_2 q$$

Productivity x is then linear in s, q and an error e_1:

(42) $$\begin{aligned} x &= E\left[x \mid s,q\right] + e_1 = \alpha_0 + \alpha_1 s + \alpha_3 q + E\left[\tilde{x} \mid s,q\right] + e_1 \\ &= \alpha_0 + \left(\alpha_1 + \beta_1\right) s + \left(\alpha_3 + \beta_2\right) q + e_1 \\ &= b_o + b_1 s + b_2 q + e_1 \end{aligned}$$

We assume $e_1 \sim N(0,\sigma_1^2)$. Then the information (*s,q*) generates a noisy signal of log productivity that, by assumption, is available to firms. Firms also learn about worker productivity from repeated observations on productive outcomes, y_t, during each period an individual spends in the labor market. This learning is common to all firms.

(43) $$y_t = x + \varepsilon_t$$

where $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$. Firms update their expectations as new information becomes available. Denote by $y^t = \left[y_0, y_1, ..., y_{t-1}\right]$ the vector of measurements available for forming expectations at t and write $\bar{y}^t$ for the average of these measurements. Then[26] a firm's best guess of productivity x at experience level t is :

(44) $$E\left[x \mid s, y^t\right] = \left(1 - \theta(t)\right) E\left[x \mid s,q\right] + \theta(t)\bar{y}^t$$

The weight $\theta(t) = \dfrac{tK}{1 + \left(t-1\right)K}$ placed on new information tends to 1 as experience

[25] It may seem equally plausible that neither workers nor employers observe z_i and q_i, and that both learn about talent from repeated observations of market outcomes. Then schooling plays no signaling role because workers do not condition their choice of schooling on knowledge of their individual talents. We can in this case still estimate the speed of employer learning with the methodology described here.

[26] For a simple introduction to Kalman updating see appendix 21 in Ljungqvist and Sargent (2000)

grows. This weight on new information is greater the larger is $K = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\varepsilon^2}$, which Lange (2004) calls the "speed of employer learning" (SEL). The SEL is greater as subsequent measures of productivity y_t become more informative relative to initial estimates based on (*s*,*q*).

Now add a competitive spot market for labor services so workers are paid their expected productivity in each period. The normal structure of the problem implies that log wages follow:

(45) $$w(s, y^t) = (1 - \theta(t))E[x \mid s, q] + \theta(t)\bar{y}^t + \frac{1}{2}\sigma^2(t)$$

where $\sigma^2(t)$ denotes the variance of the prediction error in x conditional on the information (s,y^t). Equation (45) describes wages as a weighted average of the estimate based on the initial information (s,q) available to employers and subsequent observations of log productivity. The econometrician does not have the same information (s,q) available as the firm, but instead observes (s,z). To estimate K we need to understand what (45) implies for the relation between wages and the information (s,z) available to the researcher. At $t = 0$ the linear projection of wages on (s,z) is given by the projection of E[x|s,q] on (s,z). With increasing experience the coefficients on (s,z) in a wage regression converge to the coefficients obtained from the linear projection of $\bar{y}$ on (s,z). The speed of this convergence depends entirely on *K*, the speed of employer learning. This parameter can therefore be estimated using the pattern of convergence from the initial[27] to the final[28] cross-sectional coefficients on (s,z). To be precise: With the above structure the probability limit of the OLS coefficients for schooling and ability at each experience level will be given by

(46) $$p\lim(\hat{b}^j_{OLS}) = (1 - \theta(t))b^j_1 + \theta(t)b^j_T, \qquad \text{j}=s, z$$

where b^j_1 is the coefficient obtained from regressing wages on schooling and the

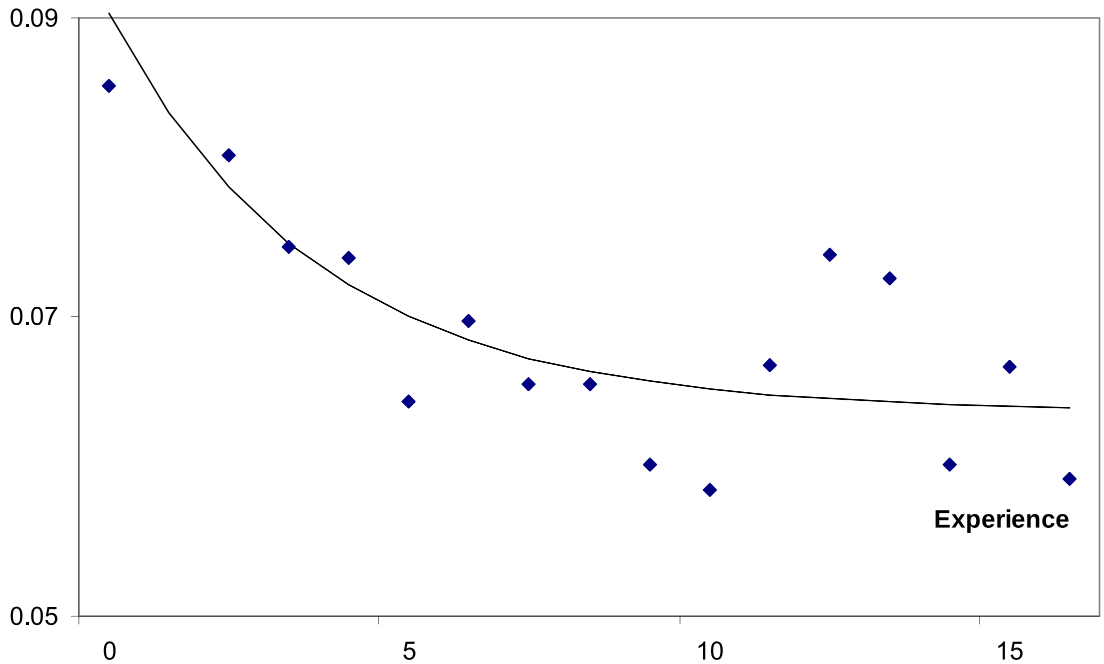
[27] experience=0
[28] experience→∞

productivity measure z before any learning takes place and b_T^j is the limit value of the coefficient.

We apply this framework to panel data from the National Longitudinal Study of Youth (NLSY). The NLSY records an individual's score on the Armed Forces Qualifying Test (AFQT), which was administered to each member of the panel while still in school. This plays the role of z_i in the above analysis—a measure of ability that is plausibly unobserved by employers. Following these individuals after labor market entry, we regressed log wages on indicators for each year of labor market experience, and interactions of these experience dummies with both years of completed schooling and AFQT score. The scatters in Figure 4, Panels A and B show the estimated returns to schooling and the AFQT at each experience level. These scatters confirm the findings that Altonji and Pierret (2001) report in a linear context. Consistent with employer learning about true productivity, returns to schooling decline with experience while returns to pre-market "ability" increase[29]. The scatter also reveals that returns to schooling and ability converge fairly quickly from their initial value to a more stable long-run level. This means that most learning occurs within the first few years of careers, so any role of education as a signal diminishes fairly quickly.

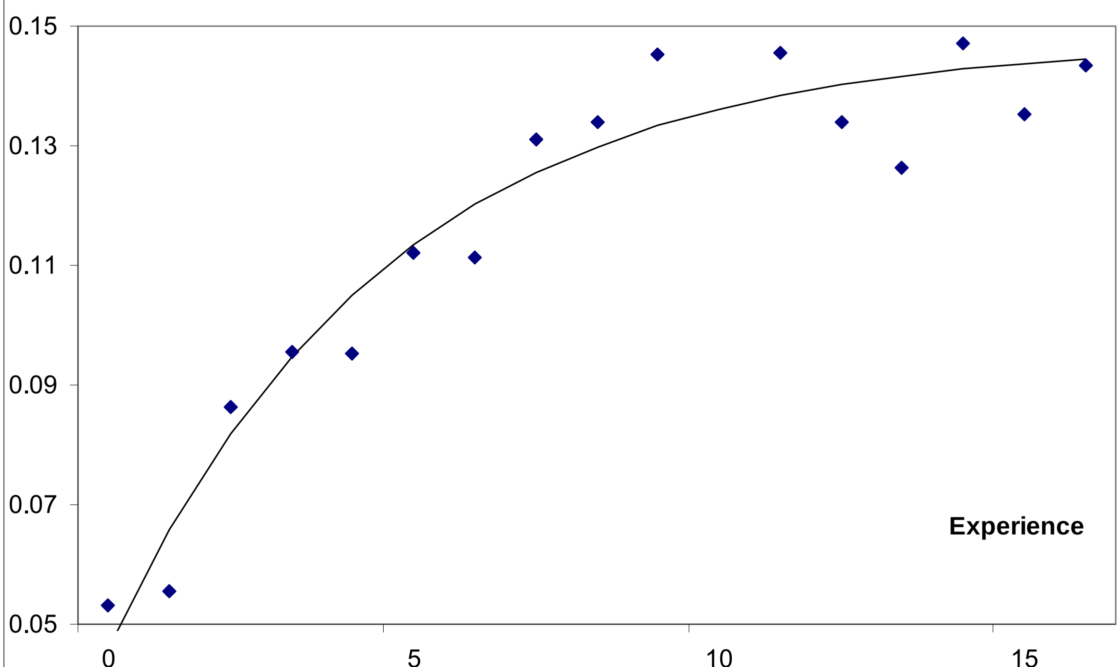
[29] Hause (1972) reports findings consistent with the findings of Altonji and Pierret (2001). Their findings refer to test scores from 4 different samples and a different time period than that analyzed by Altonji and Pierret (2001). See also Galindo-Rueda (2003) for similar findings from UK data.

Figure 3: The Returns to Schooling over the Life-Cycle



Shown are the estimated coefficients on schooling for each experience level as described in Section 3. The line shows the predicted returns to schooling over the life-cycle implied by the estimates in column 2, table 11.

Figure 4: The Returns to Ability over the Life-Cycle



Shown are the estimated return to a standard deviation in the AFQT-score for each experience level as described in Section 3. The line represents the predicted returns over the life-cycle corresponding to table 2, column 1.

The solid lines in each panel depict the predicted returns to schooling and AFQT-score at each experience level, obtained by fitting the non-linear function (46) to the estimated returns by choice of $\left(b_1^x, b_T^x, K\right)$. Estimated values of these parameters are provided in table 11. These estimates imply that employer learning is fast. One way to gauge the speed of learning is by asking how quickly the impact of an initial expectation error on wages will decline with experience. Using the definition $\theta(t) = \dfrac{tK}{1+\left(t-1\right)K}$ in the wage equation (46), a value of K=0.25 means that the weight on an initial expectation error will decline by half in the first three years of labor market experience, and by three-fourths in the first 9 years. To us, this speed of employer learning seems "fast" in comparison with the 40-45 year decision horizon relevant for schooling decisions.

Table 11) The Speed of Employer Learning[a,b]

	Schooling	AFQT-Score
The Speed of Learning K	0.2855***	0.2130***
	(0.1153)	*(0.0799)*
Initial Value b_1	0.1039***	0.0066
	(0.0122)	*(0.0259)*
Limit Value b_T	0.0568***	0.1788***
	(0.0061)	*(0.0186)*

Notes:

a. Parameters are estimated by non-linear least squares using the coefficient estimates on schooling and AFQT-score at different experience levels obtained from the NLSY as described in Lange (2004).
b. Standard errors are boot-strapped with 2,000 repetitions.

How much of the private return to schooling might be due to signaling? At each level of experience the estimates above allow us to gauge the impact of signaling on the wage, so the impact on wealth is the net present value of these effects, viewed from the start of a career. Lange (2004)[30] performs these calculations, based on wealth maximizing schooling choices. The basic idea is illustrated in Figure 5, which shows how the life-time gains from an additional year of schooling decompose into a

[30] A similar argument has been made previously by Altonji and Pierret (1997). Their work does not arrive at an estimate of the speed of employer learning. They demonstrate however that if the speed of learning is fast, then the contribution of JMS to the gains from schooling has to be low.

productivity component and a signaling component.[31] At labor market entry the effect of an additional year of schooling on the log wage is composed of a productivity increase α_1 and a signaling component β_1. The productivity component persists along the life-cycle, whereas the expected contribution from signaling declines as firms can be anticipated to learn about true productivity. For any *given* productivity effect we can decompose the total gains into a signaling and a productivity contribution. Yet the true productivity effect of schooling is unknown. The overall gains from schooling are monotonically increasing in the productivity contribution of schooling. There is therefore just one productivity effect of schooling that equates total gains to an estimate of the costs of schooling (not graphed). We can obtain such an estimate of schooling costs based on a discount rate, an estimate of tuition and the life-cycle patterns of earnings observed in the data. With this estimate of the costs of schooling we can identify a single productivity contribution that equates the gains to the costs of schooling. Based on this productivity contribution we decompose the gains from schooling into a signaling and a human capital contribution.

Fig. 5) Decomposing the Returns to Schooling into
a Signaling and Human Capital Component

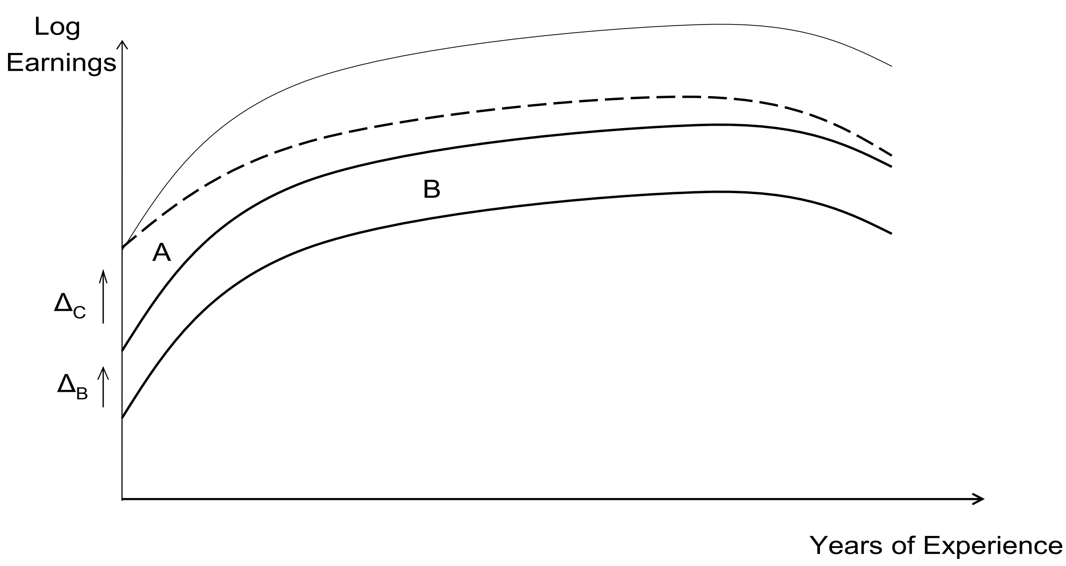


Table 12 shows the upper bounds for the contribution of signaling to the gains of schooling for various discount rates, as well as the implied productivity effects of

[31] The figure examines the particular case where firms do not have any additional information about individuals productivity (q is constant).

schooling. For a wide range of plausible discount rates the contribution of signaling is fairly small. For example, suppose the appropriate discount rate for returns on human capital investments is 6 percent—roughly the post-tax rate of return on physical capital observed in the US since World War II[32]. Then the upper bound on the contribution of signaling is smaller than 15%.

Table 12) The Contribution of Signaling to the Gains from Schooling With an Estimated Speed of Learning OF K(1) =0.25[a]

Interest Rate	Contribution of Signaling	Productivity Effects of Schooling
3%	31.65%	2.9%
4%	25.23%	3.6%
5%	19.30%	4.6%
6%	13.52%	5.6%
7%	7.90%	6.8%
8%	2.53%	8.1%
8.70%	-1.17%	9%

Notes:

a. The table shows the contribution of signaling and of the human capital model for different discount rates. This decomposition is arrived at in the manner proposed by Lange (2004) and described in more detail in that paper (see table 3, panel A)

Our conclusion from these estimates is that signaling is a minor contributor to the returns to schooling. This result is conditional on accepting the (strong) assumptions necessary to test for statistical discrimination using the method proposed by Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2004). But even with these restrictions the speed of employer learning is so fast that job market signaling is unlikely to be an important determinant of individual schooling decisions.

Differences in the Speed of Employer Learning across Industries and Occupation

The analysis above is based on the assumption that employers learn about individuals' productive types as labor market experience accumulates. An earlier literature tests the JMS-model by assuming that the speed with which employers learn varies across industries (Riley, 1979) or by other identifiable characteristics (Albrecht,

[32] see Mulligan (2002).

1981; Wolpin, 1977). For example, Riley's (1979) model predicts that average years of schooling should be lower in industries where employers learn more rapidly—the signaling returns to schooling conditional on unobserved ability $\tilde{x}$ are lower in those industries. This implies i) fewer years of schooling in industries with faster employer learning and ii) conditional on years of schooling, higher ability in industries with faster learning. Empirically, Riley's sorting model implies that, conditional on schooling, earnings should be higher in industries with lower average schooling.

Riley's prediction is contradicted by the data. Murphy and Topel (1990) estimated industry and occupational wage premiums, conditional on education, potential experience and other controls, using micro-data from the Current Population Surveys. Then they estimated "second stage" models that regressed the estimated industry and occupation premiums on education, experience and other typical controls. Their salient finding for present purposes was that the "effect" of years of schooling in the second stage was *positive*. That is, persons with more schooling also tend to work in industries and occupations that pay more for unobserved characteristics of workers—the opposite of what the signaling model would suggest. Table 13 implements this idea using more recent data for white males from the NLSY. Industry is defined using the first full-time job following completion of schooling. Again, the effect of schooling on estimated industry wage premiums is positive.

Table 13) The Relation between Industry Effects and Average Schooling [a]

Dependent Variable= Industry Specific Effects in Earnings Equation[a]		
Average Schooling[b]	0.0380***	0.0564***
	(0.0109)	*(0.0104)*
Experience	-0.0056	0.0165
	(0.0580)	*(0.0529)*
Experience squared	0.0014	-0.0002
	(0.0032)	*(0.0029)*
Fraction Black		-0.1396
		(0.1327)
Fraction Female		-0.3186***
		(0.0504)
R^2	0.0772	0.2473
Observations[c]	205	205

Notes:

a. The dependent variable is obtained from regressing log wages on schooling, experience, year dummies and demographics (black, female) as well as industry dummies using the 1979-1998 waves of NLSY. The estimated industry effects from the first stage serve as dependent variable in second stage. Regressions are weighted by number of observations by industry in first stage. *** denotes significance at the 1% level.
b. Average years of schooling are average of highest grade completed by individuals working in this industry.
c. 3-digit industries.

4. Conclusions: What Do We Know about the Social Value of Education?

We have reviewed and extended the literature on possible differences between the private and social values of education. In theory, such a “wedge” could be positive or negative. The possibility that social returns exceed private ones is typically ascribed to productive externalities—my productivity is raised by your human capital, in a way that markets do not take into account. The possibility that social returns are smaller than private ones typically views education as merely redistributive—able individuals use schooling to signal their ability, which may raise wages even if it has not impact on productivity.

The most striking finding from local aggregate data on education and wage growth in the US between 1940 and 2000 is that aggregate human capital measures are highly correlated with productivity, even after controlling for private returns. This finding is consistent with the evidence from cross-national data sets and so elemental that we believe that any account of income growth in the US must come to terms with it.

A number of authors have argued that education has substantial external benefits. This strand of the literature uses ordinary least squares (Rauch (1993)) or instrumental variable methods (Acemoglu and Angrist (2000), Moretti (2003, 2004)) to examine whether variation in education causes increases in aggregate wages after controlling for private returns. We believe that this type of evidence is inherently flawed, as it does not sufficiently account for endogeneity issues implied by a spatial equilibrium. We show that controlling for variation in skills by birthplace in aggregate wage regressions reduces the association between aggregate wages and schooling substantially. The fact that instrumental variable methods are incompatible with the restrictions imposed by a spatial equilibrium model also does not allow us to rule out skill-biased technological change as a cause for the association between schooling and aggregate wages.

The strong positive link between productivity and education is also problematic for adherents of the Job Market Signaling model. Job Market Signaling generates private returns to schooling without the corresponding social returns. This means that if one seriously entertains Job Market Signaling as an explanation for observed private returns to schooling, then the problem of explaining observed aggregate returns becomes even more severe. Our review of the available empirical evidence on Job Market Signaling leads us to conclude that there is little in the data that supports Job Market Signaling as an explanation for the observed returns to schooling.

References

Acemoglu, Daron "A Microfoundation for Social Increasing Returns in Human Capital Accumulation" *The Quarterly Journal of Economics,* 111 (3), (Aug. 1996) 779-804.

Acemoglu, D. and J. Angrist, "How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws," *NBER Macroannual* (2000) pp.9-59.

Albrecht, J. W. "A Procedure for Testing the Signaling Hypothesis," *Journal of Public Economics,* 15 (1981) 123-132.

Altonji, J., "The Demand for and Return to Education Outcomes are Uncertain," *Journal of Labor Economics,* (January 1993).

Altonji, J. and C. Pierret, "Employer Learnings and the Signaling Value of Education," *Industrial Relations, Incentives and Employment,* I. Ohashi and T. Tachibanaki, eds. (London; Macmillan Press Ltd., 1997).

Altonji, J. and C. Pierret, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics* (February 2001), 313-350.

Barro, Robert, "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics,* CVI (1991), 363-394.

Barro, R. and X. Sala-i-Martin, "Technological Diffusion, Convergence, and Growth," *Journal of Economic Growth,* March 1997.

Bauer T., and J. Haisken-DeNew, "Employer Learning and the Returns to Schooling," *Labor Economics, 8,* (2001) 161-180.

Becker, G., *Human Capital, (*New York: Columbia University Press for the National Bureau of Economics Research, 1964).

Bedard, Kelly, "Human Capital versus Signaling Models: University Access and High School Dropouts", *Journal of Political Economy,* Vol. 109, 4 (August 2001), 749-775.

Belman, D. and J. Heywood, "Sheepskin effects in the returns to education: an examination of women and minorities", *Review of Economics and Statistics,* 73 (1991), 720-724.

Benhabib, J. and M. Spiegel, "The Role of Human Capital in Economic Development, Evidence from Aggregate Cross-Country Data," *Journal of Monetary Economics* 34 (2), (1994) 143-173.

Bils, M. and P. Klenow, "Does Schooling Cause Growth," *American Economic Review,* 90 (2000), 1160-1193.

Borjas, George J., "Ethnic Capital and Intergenerational Mobility," *The Quarterly Journal of Economics*, CVII (1992), 123-150.

Borjas, George J., "Ethnicity, Neighborhoods, and Human Capital Externalities," *The American Economics Review,* Vol. 85, 3 (June 1995), 365-390.

Card, David, "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics,* Volume 3A, ed. by Orley Ashenfelter and David Card (North Holland, Amsterdam and New York 1999)

Conley, T. and C. Udry, "Learning About a New Technology: Pineapple in Ghana" *Economic Growth Center Working Paper* 817, (May 2004).

De la Fuente, A. and R. Domenech "Schooling Data, Technological Diffusion and the Neoclassical Model," *American Economic Review* 91 (2), (2001), 323-327.

Dee, Thomas, "Are There Civic Returns to Education?" Mimeo, Swathmore College (Febr. 2003).

Farber, H. and R. Gibbons, "Learning and Wage Dynamics", *Quarterly Journal of Economics*, (February 1996), 1007-1047.

Foster, A. and M. Rosenzweig, "Information, Learning and Wage Rates in Rural Labor Markets," *Journal of Human Resources,* 28, (1994), 759-790.

Foster, A. and M. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and the Technical Change in Agriculture," *Journal of Political Economy,* 103(6), (Dec. 1995), 1176-1209.

Frazis, Harley, “Human Capital, Signaling and the Patterns of Returns to Education,” *Oxford Economic Papers – New Series,* 54(2), (April 2002), 298-320.

Friedman, Milton, *“Captalism and Freedom”* (Chicago: University of Chicago Press 1962).

Galindo-Rueda, Fernando, “Employer Learning and Schooling-Related Statistical Discrimination in Britain”, *IZA Discussion Papers,* No. 777, (May 2003).

Glaeser, Edward, “Learning in Cities,” *Journal of Urban Economics* 46, 254-277 (1999)

Habermalz, S. “An Examination of Sheepskin Effects over Time,” Working Paper (2004).

Hall, R. and C. Jones, “Why Do Some Countries Produce So Much More Output Per Worker than Others?” *Quarterly Journal of Economics,* 114(1), 83-116.

Hamermesh, Daniel, “The Demand for Labor in the Long Run,” in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, (North-Holland Press, 1986).

Hause, John, “Earnings Profile: Ability and Schooling,” *The Journal of Political Economy,* Vol. 80, 3 (May-Jun. 1972) S108-S138.

Hause, John, “The Find Structure of Earnings and the On-the-Job Training Hypothesis”, *Econometrica,* Vol. 48, 4 (May, 1980), 1013-1029.

Heckman, J. “A Life-Cycle Model of Earnings, Learning and Consumption,” *Journal of Political Economy,* 84(2) (August 1976), S11-S44.

Heckman, J. and P. Klenow “Human Capital Policy,” *Policies to Promote Capital Formation,* ed. Michael Boskin (1998).

Heckman, J., Lochner, L. and P. Todd, “Fifty Years of Mincer Earnings Regressions,” *NBER Working Paper 9732* (2003).

Hungerford, T. and G. Solon, “Sheepskin Effects in the Returns to Education,” *Review of Economic Studies,* (Febr. 1987), 176-178.

Jaeger, D. and M. Page, “Degrees matter: new evidence on sheepskin effects in the returns to education”, *Review of Economics and Statistics,* 76 (1996), 733-739.

Jovanovic, B. and R. Rob., “The Growth and Diffusion of Knowledge,” *Review of Economic Studies* 56(4), (1989), 569-582.

Kaldor, N., “Capital Accumulation and Economic Growth”, in F. Lutz and D.C. Hague (eds.), *The Theory of Capital*, (New York: St. Martin’s Press 1961), pp. 177-222.

Katz, L. and K. Murphy, “Changes in Relative Wages, 1963-1987: Supply and Demand Factors”, *The Quarterly Journal of Economics,* Vol. 107, No. 1 (Feb., 1992) pp. 35-78.

Krueger, A. and M. Lindahl “Education for Growth: Why and for Whom?” *National Bureau of Economic Research* 7591 (2000).

Lang, K. and D. Kropp, “Human Capital vs. Sorting: The Effects of Compulsory Attendance Laws,” *Quarterly Journal of Economics,* 101, (August 1986), 609-624.

Layard, R. and G. Psacharopoulos, “The Screening Hypothesis and the Returns to Education,” *Journal of Political Economy,* 82, 985-998.

Lochner, L. and E. Moretti, “The Effect of Education on Criminal Activity: Evidence from Prison Inmates, Arrests and Self-Reports”, *American Economic Review,* 94(1), (2004).

Lucas, Robert E., “On the Mechanics of Economics Development,” *Journal of Monetary Economics”,* XXII (1988), 3-42.

Lucas, Robert E., “Why Doesn’t Capital Flow from Rich to Poor Countries?” *American Economcis Review Papers and Proceedings,* LXXX, (1990) 92-96.

Lungqvist, L. and T. Sargent, “Recursive Macroeconomic Theory”, (MIT Press 2000).

Mankiw, G., D. Romer and D. Weil, “A Contribution to the Empirics of Economic Growth” *Quarterly Journal of Economics,* 107(2), (1992) 407-438.

Marshall, Alfred, *Principles of Economics,* Macmillan, London 1890, 8th Edition published 1920.

Miligan, K., E. Moretti and P. Oreopoulos, “Does Education Improve Citizenship? Evidence from the U.S. and the U.K.,” *NBER Working Paper* 9584 (2003).

Moretti, Enrico “Estimating the Social Return to Higher Education: Evidence from Longitudinal and Cross-Section Data,” *Journal of Econometrics* (2003)

Moretti, Enrico “Workers’ Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions”, *American Economic Review,* 94(3), (2004).

Mulligan, Casey “Capital, Interest, and Aggregate Intertemporal Substitution” *NBER Working Paper 9373* (2002)

Murphy, K.M. and R. H Topel, “Estimation and Inference in Two-Step Econometric Models”, *Journal of Business & Economic Statistics,* October 1985, Vol. 3, No. 4.

Murphy, K.M. and R. H. Topel, “Efficiency Wages Reconsidered: Theory and Evidence”, in *Advances in the Theory and Measurement of Unemployment,* ed. Yoram Weiss & Gideon Fishelson, London: Macmillan, 1990

Nelson, R. and E. Phelps, “Investment in Humans, Technological Diffusion and Economic Growth,” Cowles Foundation Discussion Paper (1965).

Park, Jin Heum “Returns to Schooling: a Peculiar Deviation From Linearity”, Working Paper no. 335 (Industrial Relations Section, Princeton University).

Pritchett, Lant “Where Has All The Education Gone?” *World Bank Economic Review.* (December, 2001).

Rauch, James “Productivity Gains from Geographic Concentration in Cities,” *Journal of Urban Economics* 34, (1993) 380-400.

Roback, Jennifer “Wages, Rents, and the Quality of Life,” *Journal of Political Economy* 90, (1982) 1257-1278.Riley, John, “Testing the Educational Screening Hypothesis,” *Journal of Political Economy,* (Oct. 1979), S227-S252.

Riley, John, “ Silver Signal – Twenty Five Years of Screening and Signalling,” *Journal* *of Economics Literature,* 39 (2001), 432-478.

Sacerdote, B., "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics,* (May 2001) 681-704.

Schultz, Theodore, "The Economic Value of Education," (Columbia University Press, New York 1963)

Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics,* (August 1973), 355-379.

Stiglitz, Joseph, "The Theory of "Screening", Education and the Distribution of Income," *American Economic Review,* Vol. 65, 3 (June 1975), 283-300.

Temple, Jonathan "Growth Effects of Education and Social Capital in the OECD Countries," *OECD Economic Studies,* 33, (2002) 57-101.

Topel, Robert "Labor Markets and Economic Growth," *Handbook of Labor Economics,* ed. Orley Ashenfelter and David Card (Amsterdam: Elsevier Science 1999) 2943-2984.

Uzawa, Hirofumi "Optimum technical change in an aggregative model of economic growth," *International Economic Review* 6, (1965), 18-31.

Weiss, Andrew, "A Sorting-cum-Learning Model of Education" *The Journal of Political Economy,* Vol. 91, 3 (June 1983), 420-442.

Weiss, Andrew, "Human Capital and Sorting Models," *Journal of Economic Perspectives, IX (1995), 133-154.*

Wolpin, Kenneth, "Education and Screening," *American Economics Review,* 67 (Dec. 1977), 949-958.

Data Appendix

The sample consists of US males aged 18-65 from the Census. We used the standard 1940-2000 1-% microdata sample provided by the IPUMS (www.ipums.org). For 1940, 1950 and 1960 we use the general 1%-sample, for 1970 the "form 2 state" sample, for 1980 the 1% Metro (B sample), for 1990 the 1% Metro sample and for 2000 the general 1% sample. We base our analysis on employees with valid wage observations who worked 40 weeks or more. Employees can not be fully consistently defined over the sample period. In 1940 we defined employees as those for whom no susbtantial non-wage income is reported. From 1950 onwards employees were those without income from businesses or farms. We drop residents of Hawaii and Alaska since these states are not sampled in 1940 and 1950. Individuals with top-coded wages were excluded.
The coding of education in the census underwent some changes in 1990. From 1940-1980 the variable *HIGRADE* coded education as the highest year of schooling or college completed. In 1990 and 2000 the variable *EDUC99* codes the highest grade of schooling completed through 11^{th} Grade and classifies high school graduates by their degree completed. The schooling variable we used EDUCREC. The census provides this variable to make education variables comparable across census years. It classifies individuals in 9 categories of completed schooling (0 and Kindergarten, grades 1-4, 5-8, 9, 10, 11, 1-3 years of college, 4+ years of college). In the first stage regressions we use EDUCREC as a categorical variable. We calculated mean education by state by assigning to the categories provided by EDUCREC the vector of years of schooling (0, 2.5, 6.5, 9, 10, 11, 12, 14, 16).
The dependent variable in the first state is the log weekly wage (in 1990 dollars). In the first stage we define the skill categories using both the schooling variable and experience, where experience is choosen to be age minus cohort. We categorize experience into 10-year cells. Thus we have 9 education categories and 6 experience categories (18-20, 21-30, …, 61-65) and therefore arrive at 54 skill groups. We interact the skill effects with each census year and therefore estimate in total 54*7= 378 skill effects. We define the cohorts by decades and have 11 cohorts (1870-79, 1880-1889, … , 1980-1989). We therefore have 49*11=539 birthstate* cohort categories. Finally we estimate state-effects for each census year and thus estimate a total of 49*7=343 state effects. These state effects will provide the basis of the 2^{nd} step estimation.
Table A1 presents summary statistics for the Census Extracts used.

TABLE A1 DESCRIPTIVE STATISTICS FOR CENSUS IPUMS

	1940	1950	1960	1970	1980	1990	2000
Education	8.74	9.71	10.30	11.27	12.33	12.98	13.13
	(3.64)	(3.58)	(3.53)	(3.27)	(2.88)	(2.62)	(2.60)
Experience	22.31	22.23	23.01	21.92	18.94	18.75	19.98
	(13.08)	(13.66)	(13.51)	(13.82)	(13.35)	(12.05)	(11.69)
Age	37.05	37.93	39.31	39.19	37.27	37.73	39.10
	(11.69)	(12.31)	(12.13)	(12.68)	(12.58)	(11.64)	(11.54)
Log Weekly	5.28	5.50	5.95	6.16	6.16	6.12	6.07
Wage	(0.69)	(0.61)	(0.60)	(0.67)	(0.70)	(0.67)	(0.67)
N	164,307	97,688	298,153	358,355	399,746	437,905	512,599