

Empirical Evidence on Human Capital Spillovers

Jeremy Rudd^a
Federal Reserve Board

Abstract: This paper examines whether the average level of human capital in a region affects the earnings of an individual residing in that region in a manner that is external to the individual's own human capital. I find little evidence of an external effect of human capital, which suggests that human capital spillovers of the form postulated by the new growth literature are unlikely to matter much in practice.

JEL classification codes: D62, I20, J24, J34, O15, O18

Keywords: Human capital externalities, endogenous growth

^a Address: Board of Governors of the Federal Reserve System, Washington, DC 20551-0001. E-mail: jeremy.b.rudd@frb.gov; telephone: (202) 452-3780; fax: (202) 872-4927.

I thank Christina Paxson and Angus Deaton for numerous valuable comments and much helpful advice, and Alan Blinder, Anne Case, Ken Chay, Michael Greenstone, and particularly David Card for various useful conversations (the usual caveat applies). Funding received from the Lynde and Harry Bradley Foundation during the course of this research is gratefully acknowledged. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1. Introduction

This paper examines whether the average level of human capital in a region affects the earnings of an individual residing in that region in a manner that is external to the individual's own human capital. Specifically, I consider whether a higher overall level of educational attainment in a state acts to raise the earnings of an individual residing in the state even after controlling for the individual's own educational attainment. I label such a phenomenon a "human capital spillover."¹

The motivation for considering this question is twofold. First, if the average level of education in a region acts to augment the productivity of all residents, there will be a gap between the private and social returns to schooling. This might serve as a justification for subsidizing education in order to align the optimal schooling choice of private agents with the level they would obtain at the command optimum. Second, an important strand of the "new growth" literature invokes the presence of external productive effects to human capital; we might have more confidence in such a modelling strategy if such effects could be seen empirically.²

In observational terms, a true spillover would imply the following. Say we take two identical individuals and put each of them in one of two regions. The regions are also identical--at least in terms of productive characteristics--except that the level of human capital (*e.g.*, the level of schooling) is higher in one of the two regions. If human capital spillovers are present, then the individual in the high-schooling region will be more productive than his or her twin in the low-schooling region--which, under competitive conditions, will translate to a higher wage for the agent in the high-schooling region.

To test for the presence of human capital spillovers, I estimate a regression model of the form:

$$\ln y_{is} = \beta_1 \cdot X_{is} + \beta_2 \cdot E_{is} + \delta \cdot Z_s + \theta \cdot H_s + u_{is} . \quad (1)$$

Here, a measure of log earnings y for individual i in region s is projected on individual is 's own characteristics X_{is} , together with individual is 's years of schooling E_{is} (u_{is} is the regression error). This

¹ Here and elsewhere in the paper, I identify "education" or "educational attainment" with human capital.

² An exemplary paper is Lucas (1988), where the average level of human capital in the economy acts as a Hicks-neutral shift term in the production function. The presence of an "average human capital" term in the production function is neither necessary nor sufficient for endogenous growth; rather, it is the assumption of a linear accumulation constraint for human capital which generates endogenous growth in a Lucas-style model. (Solow, 1994 provides a discussion of this point.) Nevertheless, a positive elasticity of output with respect to average human capital raises the welfare gain from bringing private human capital accumulation in line with the social optimum, because the result is an economy whose growth rate--as opposed to level--of output is increased. See Lucas (1988), p.23.

yields a typical Mincerian earnings equation, to which I then add a vector of region-specific characteristics; in particular, I include a measure of statewide educational attainment H_s along with a portfolio of other state-specific correlates Z_s in the regression. The coefficient θ then captures the effect of statewide education on individual earnings.

Fitting equation (1) to single years of cross-sectional data reveals that the coefficient θ is typically positive and statistically significant. While this fact appears to have been part of the folk wisdom of labor economics, it received little attention until James Rauch examined it explicitly in a 1993 paper. Rauch fit a regression similar to (1) to city-level data, measuring H_s as the average number of years of schooling and potential labor market experience of workers in a given SMSA (potential experience was defined as age *minus* education *minus* six). Rauch employed a single cross section from the 1/1000 public use sample of the 1980 Census of Population and Housing in his analysis; as the Census contains information on housing characteristics and housing expenditures, he was also able to estimate an hedonic rent regression, which allowed him to assess how housing rents are affected by different levels of SMSA-specific human capital.³

An innovative aspect of Rauch's work was his use of Roback's (1982) locational equilibrium model to interpret his results. In a Roback framework, a productive site-specific characteristic that has no inherent amenity value should raise wages and land rents in a region (the effect on wages is ambiguous if the site-specific characteristic is an amenity). Thus, a positive coefficient on the regional human capital term H_s in (1) is consistent with the presence of productive external effects to human capital, so long as average human capital is not also an amenity. In fact, Rauch did find a positive relationship between wage and rent levels and the average level of education in an SMSA; the effect of potential labor market experience is also positive, but less robust.⁴ Rauch argued that his results provide evidence for the existence of human capital externalities.

In what follows, I use data from fourteen years (1978-91) of the March Current Population Survey (CPS) to estimate a modified version of the earnings regression (1). My analysis differs from Rauch's in three important respects. First, instead of considering a single cross section from one year's

³ An hedonic rent regression relates housing rents to housing characteristics, much as an earnings equation relates a worker's earnings to his or her own characteristics.

⁴ The coefficient estimates from Rauch's earnings regression are slightly peculiar. First, his regression implies that the return to an additional year of schooling is only 4.8 log points. This is low relative to many other estimates; see, for example, Card and Krueger (1992). It is very likely that this is the result of Rauch's using occupation indicators in his earnings regressions. In addition, the measured return to SMSA-level education is rather high--3.3 log points per year of additional (average) schooling, or almost 70 percent of the estimated return to own schooling. *A priori*, this might suggest either that average education is proxying for some other SMSA-specific productive characteristic, or that the return to own schooling has been mismeasured.

worth of data, I fit a version of equation (1) using a "panel" of pooled cross sections; this allows me to control for region-specific fixed effects.⁵ Second, my analysis proceeds at the state level. Although this approach can be criticized on the grounds that an SMSA is more likely to represent a local labor market, that is, any spillover effects might well be diluted over a region as large as a state, I believe that this objection is overruled by the benefits that obtain from using state-level data. State geography is stable over time, which is a requirement for creating a series of pooled cross sections; moreover, a number of state-specific correlates that might plausibly be driving the observed correlation between average education and earnings are unavailable for SMSAs. Third, I consider a general identification issue related to the measurement of spillovers. If human capital spillovers act to raise worker productivity, then an increase in the average level of educational attainment in a region shifts out the demand curve for labor and, all else equal, raises wages. But the *supply* of highly educated workers is also larger in regions with higher average education levels, and this tends to lower wages. Unless we control for this supply effect, we will be unable to identify the productive effect of human capital spillovers by considering the coefficient from a wage regression like (1).

In marked contrast to Rauch, I conclude that there is essentially no evidence of human capital spillovers in state-level U.S. data. Observed educational spillovers do not reflect true externalities, but instead largely stem from productive state characteristics that are themselves correlated with state educational attainment. In addition, a portion of the observed correlation between average education and earnings appears to depend on how the return to schooling is measured at the individual level.⁶

The organization of the balance of the paper is as follows. In section two I outline the estimation method I use in order to fit a version of equation (1) to pooled cross sections from the CPS; I discuss the results from these pooled regressions in section three. In section four, I examine several factors which might cause us to mismeasure the effect of average education on individual earnings, including the supply effect described above. In section five I present several extensions to the basic analysis that attempt to determine whether the observed relationship between state average education and individual earnings reflects true human capital spillovers. In particular, I apply to my sample of

⁵ The absence of housing data in the CPS makes it impossible to fit the sort of hedonic rent equation required for a fully implemented Roback model. I do not consider this to be too problematic, however, since a positive relationship between individual wages and average regional education is fully consistent with average education's being a site-specific productive characteristic (see above).

⁶ Some time after the research reported here was completed, I became aware of work by Acemoglu and Angrist (2000) that examines this topic using a somewhat different technique. Acemoglu and Angrist attempt to control for the possibility that other, unmeasured factors could induce a correlation between average educational levels and earnings by using an instrument--variation in compulsory schooling laws--that permits them to identify the true external return to education. They reach a conclusion that is broadly similar to the one reported here, specifically, that there is little evidence of human capital externalities in U.S. microdata.

U.S. states a framework used by Bils and Klenow (1996) in order to test for the presence of schooling-driven endogenous growth in cross-country data. Section six concludes.

2. Estimation method

There are two advantages to using pooled cross-sectional data in the analysis.⁷ The first advantage, noted above, is that combining cross-sections from several different years allows the inclusion of state fixed effects in the regressions. The second, subtler advantage is that pooling the data increases the number of degrees of freedom underlying the estimates of the coefficients on the state-specific characteristics Z_s and human capital term H_s .⁸

A straightforward method for estimating equation (1) over pooled data involves "stacking" the data from all years of the CPS. This suffers from at least two drawbacks. First, such a method requires estimating a large number of interaction terms--ideally, we would want to allow the coefficients on the individual characteristics in the earnings regressions to change over time. Second, such a procedure would be computationally onerous.⁹ I therefore take a slightly different, two-step approach. First, I estimate the earnings equation (1) in each year over a given sample of workers, but include a state fixed effect f_s in lieu of the state-specific characteristics H_s and Z_s .¹⁰ I then recover the various state-year effects f_{st} . Because my data cover 49 states over 14 years, there are 686 of these effects ($49 \times 14 = 686$)--a separate value of f_s from each of the fourteen regressions. Second, I treat the resulting panel of f_{st} terms as data, and estimate longitudinal regressions with the following specification:

$$f_{st} = \phi_s + d_t + \delta Z_{st} + \theta H_{st} + e_{st}, \quad (2)$$

⁷ I have also considered a large number of cross-sectional regressions of the form (1). The only result from these regressions worth noting is that the estimated coefficients on the H_s and Z_s terms vary considerably depending on the year of estimation.

⁸ Because they are constant across all individuals in a given state, we effectively only have one observation on these variables per state per year.

⁹ Besides involving a large number of observations, the stacked regressions include terms such as statewide educational attainment which are constant over clusters (states). As is now well known, OLS grossly understates the standard errors of the coefficients on such regressors--see Moulton (1986) and Deaton (1994) for a discussion. Correcting the standard errors for large-sample regressions such as these is not difficult in principle, but is extremely time-consuming in practice.

¹⁰ Note that these fixed effects will also pick up any state-specific component of the regression error; such a component might be present if there are additional state characteristics that are not captured by the included H_s and Z_s terms.

where ϕ_s denotes a time-invariant state-specific intercept, d_t is a year effect, e_{st} denotes the regression error term, and H_{st} and Z_{st} denote the human capital measure and vector of additional characteristics for state s at time t . Estimation is by weighted least squares, where the weights used are the cell sizes from the relevant cross-sectional regression, *i.e.*, the number of observations for state s used in fitting the least-squares dummy-variable version of regression (1).¹¹

The dependent variable used in the earnings regressions is the natural logarithm of average weekly wages; note that the March CPS provides figures for income received and weeks worked during the *previous* calendar year. The earnings regressions include a number of individual characteristics, namely, own years of completed schooling, potential labor market experience and its square (where potential labor market experience is defined as age *minus* years of education *minus* six), sex, race, marital status, and fourteen industry indicators. Sex is also interacted with potential experience, experience squared, marital status, and race. Individuals are included in the regressions if they meet certain criteria for minimum and maximum earnings, average weekly wage, and age; in addition, the regressions only include residents of the 48 contiguous U.S. states and the District of Columbia. Depending on the year, the sample sizes for the cross-sectional regressions range from 46,000 to 58,000 observations. The number of observations for a given state ranges from 400 to over 5,000 in most years; the smallest state has more than 400 observations in all but two years.¹²

I require a suitable measure of state-specific human capital H_s . In the baseline regressions, I use statewide average education, that is, the average number of years of completed education of state residents calculated over the sample of individuals who meet the criteria for inclusion in the wage regressions. For reasons detailed in section B of the Appendix, I use the value of this measure from the previous year's CPS.¹³

Finally, I include a number of state-specific characteristics Z_s in the pooled regressions. These terms are added in order to control for third factors which represent true sources of state wage

¹¹ I examine the conditions that are required in order for this method to be valid in section A.1 of the Appendix. It is readily apparent, though, that obtaining a consistent estimate of the f_{st} terms is critical.

¹² I included individuals in the wage regressions if their wage income was neither topcoded nor imputed, they reported an industry, had nonnegative potential labor market experience, were between 16 and 65 years of age, and had wage income greater than a minimum threshold and average weekly wages within a given range. The threshold I used for minimum wage income was 101 dollars in calendar year 1979; average weekly wages were required to be greater than or equal to 36 dollars and less than or equal to 2,500 dollars in 1979 (these thresholds are the same ones used in Card and Krueger, 1992). I scaled the thresholds used for other years according to changes in the mean average weekly wage and mean wage income, relative to their 1979 sample means.

¹³ The average education measure is calculated using the sampling weights provided in the March CPS files; however, I do not further weight individual observations by hours worked or some other measure of labor force participation. Summary statistics for the average education measure are given in Appendix Table B.1.

differentials, but which are incidentally correlated with state average education. In section three (below), I consider the effect of including these third-factor proxies in the regressions.

3. Results from the pooled regressions

I present the results obtained from fitting the baseline version of regression (2) in Table 3.1; note that the specifications reported in this table do *not* include state fixed effects.

Column 1(a) of the table presents the basic result that motivated this study in the first place. Even after we have controlled for own education and individual characteristics, a measure of state average education enters the earnings regression with a positive and statistically significant coefficient. If we take the regression result at face value, it implies that increasing a state's average educational attainment level by one year raises the wages of the state's residents by almost 11 log points. To put this figure in perspective, note that the return to *own* education over this period ranges from seven to nine log points per year of schooling.

As I have noted above, however, the presence of a positive coefficient on the statewide educational attainment variable in regression (2) is not sufficient proof of the existence of spillovers to education. It is easy to think of a host of other state characteristics that might be the true source of a region's productivity but which are themselves correlated with the region's average education level, or that might generate the observed correlation between average education and own earnings without implying the existence of productive spillovers. What are some of these "third factors," and what effect does controlling for them have on the statewide education coefficient?

Alternate sources of the correlation between average education and earnings

Perhaps the most basic interpretation of the correlation between state average education and earnings is simply that the causality runs in the opposite direction. If education is a normal consumption good, or if optimal schooling choice is an increasing function of own or family resources, then average education could merely be a proxy measure for whether a state is a high-wage state. In other words, if we were to fit these regressions and include a state's luxury car ownership rate in lieu of the average education term, we might perhaps obtain a very similar result.¹⁴ I therefore require some sort of control

¹⁴ The proposition that education is an increasing function of resources can be formally demonstrated using Becker's 1967 model of schooling choice--see Card (1994) for a discussion. The point that earnings or wealth might cause higher schooling, rather than *vice-versa*, has been noted in an intergenerational context by Birdsall and Behrman (1984). Bils and Klenow (1996) examine a

for state wealth or income in the regression.

It is worth emphasizing that by "control" I do not mean to imply that there is a contemporaneous relationship between state earnings and average education that would have to be corrected for using instrumental variables techniques. Instead, the problem is one of omitted variables bias--the estimated coefficient on state education is capturing the effect of an omitted factor (here, whether a state is a high-wage state) that is itself correlated with state average education. This is because the level of wealth or resources that is relevant to an individual's schooling decision is the level that prevails when the individual makes his or her schooling choice, which is not necessarily equal to the *current* level of resources in a state.

However, directly including a measure of per-capita income--for instance, from the state income and product accounts--is improper on two counts. First, it comes close to regressing a dependent variable--the wage term in the earnings regression--on itself. Second, if the average level of education in a region actually does augment labor productivity and labor income, then this effect will be washed out by the inclusion of a measure of state earnings as a regressor, since any common covariance between the statewide education and income terms will be thrown out by the regression.

As a compromise, therefore, I include state nonwage (unearned) income per capita in the pooled earnings regressions, as measured by the state income and product accounts.¹⁵ Although this variable does not measure the level of resources that is relevant for the schooling decision of older cohorts, it does provide a reasonable estimate of how well-off a state is relative to other states, and probably suffers less from being correlated with productive human capital spillovers.¹⁶ In addition, a measure of overall state well-being is likely to help control for other factors--such as relative price dispersion across states (see below)--that might be the source of the observed correlation between average education and earnings. Adding the income measure to the regression does in fact reduce the value of the average education coefficient--but not by much--as the result in column 2(a) of Table 3.1 indicates.

similar point; they argue that the correlation between schooling and growth that is documented by the empirical growth literature is likely the result of economic growth's raising the demand for education.

¹⁵ I provide details on the construction of and sources for all auxiliary variables in section B.3 of the Appendix; Appendix Table B.2 provides summary statistics for the various measures.

¹⁶ Technically, this will be true only inasmuch as human capital spillovers act to raise the productivity of labor exclusively--which is not likely to be the case. If the spillovers also increase nonwage income, say through their effect on physical capital productivity, then the measured relationship between average education and earnings might still be attenuated (see section A.4 of the Appendix for details). This should be kept in mind when assessing the impact that including nonwage income has on the average education coefficient.

One might also argue that a portion of the observed relationship between average education and earnings reflects agglomeration economies. If, for example, the extent of the market places a limit on labor's ability to profitably specialize, and if specialization raises the return to human capital, then densely settled regions should have a larger proportion of highly educated individuals. The coefficient on educational attainment in the modified earnings regression will therefore be pushed upward if the density of economic activity itself augments production. In addition, this sort of story suggests that we might plausibly expect to find a positive interaction between agglomeration and average human capital.¹⁷

I address this by adding a measure of agglomeration (a weighted index of population density) to the pooled regressions. This can be criticized on the grounds that population is endogenous in a locational equilibrium context--*e.g.*, higher wages induce migration into a region--although it is unclear whether this criticism also applies to population *density*.¹⁸ In his own work, Rauch finds that including population in his wage and rent regressions does indeed appear to affect the average schooling coefficient; in particular, the coefficient on average education in his wage regressions falls slightly. Note, though, that population does not enter his wage regressions significantly, and the instrument Rauch uses in order to control for population's endogeneity (SMSA land area) has poor first-stage fit. Including the log agglomeration index in my regression, however, reduces the magnitude of the average schooling coefficient by about a quarter (see column 3(a) of Table 3.1). The agglomeration index also enters the regression with the expected positive sign.

A third interpretation of the positive correlation between average statewide educational attainment and earnings invokes the existence of a "wage curve"--a negative relationship between the

¹⁷ A theoretical model which captures this idea is Becker and Murphy (1992), although strictly speaking Becker and Murphy argue that agglomeration operates not so much through extent of the market as by reducing the costs of coordinating and combining specialized inputs. Search models of skill acquisition (see Jovanovic and Rob (1989) and the references therein) would appear to support the view that skill acquisition--and thus productivity--depends on both the overall level of human capital in a region and how spatially concentrated this human capital is; this in turn suggests that there will be a positive interaction between agglomeration and average human capital. In section A.2 of the Appendix, I examine whether an agglomeration-education interaction is evident in my data.

For empirical evidence that agglomeration increases the *rate* at which workers acquire skills, see Glaeser and Maré (1994); Ciccone and Hall (1996) present evidence that agglomeration is itself inherently productive.

¹⁸ For example, it is possible to envision a model which implies an ambiguous relationship between site characteristics and population density. Consider an increase in a productive site characteristic with no amenity value; while this tends to induce immigration and thus raises population in a region, population *density* might decline if wages also rise and land is a normal good. In practice, however, this theoretical distinction is probably irrelevant, since any variation in population density in a regression that controls for fixed effects will reflect variation in population (because state land area is constant over time).

level of earnings in a region and the region's unemployment rate.¹⁹ If it is true that better-educated individuals have lower unemployment rates on average (this is mildly evident in my data), then we might argue that statewide educational attainment is proxying, in part at least, for the unemployment rate. When I include the log state unemployment rate in the regression model (see column 4(a) of the table), however, the coefficient on state education remains essentially unchanged.²⁰

The remaining columns of Table 3.1 report the results that obtain from including various combinations of these terms. Note that including both log nonwage income per capita and the log agglomeration index more than halves the size of the average education term; this is true even if the log unemployment rate is included (see columns 5(a) and 8(a) of the table). However, the coefficient on average education remains statistically and economically significant.²¹

Controlling for other state characteristics

There are several other state-specific variables that might be driving the observed correlation between average education and earnings; I consider these separately because I do not have enough data to include them in the pooled regressions, and can only assess their importance using one or more cross-sectional regressions.

The presence of capital-skill complementarities might result in there being a higher level of physical capital per worker in states with better-educated workers. While this does not mean that average education is unproductive--it is, in particular, raising the productivity of physical capital--it does imply that a portion of physical capital's contribution to state productivity will be erroneously captured by the average education term. I therefore added estimates of nonfarm capital per worker developed by Munnell (1990) to the cross-sectional regressions (1); invariably, the coefficient on the physical capital variable was insignificant and did not affect the average education term.²²

¹⁹ See Blanchflower and Oswald (1994) for an exhaustive examination of the wage curve. See also Card (1995) and Card and Hyslop (1997) for an alternative interpretation of the wage curve as a misspecified Phillips curve.

²⁰ Moreover, the coefficient on the unemployment rate is positive, not negative, and statistically insignificant. This is actually a standard result for U.S. data--state unemployment is not negatively correlated with earnings unless fixed effects are included in the specification. See Card (1995), p. 789, for a discussion. Using the log of the unemployment rate, as opposed to its level, in the regressions follows Blanchflower and Oswald (1994, ch. 4).

²¹ For reference, note that over the $14 \times 49 = 686$ state-year observations the average education measure is highly positively correlated with log nonwage income (the simple correlation coefficient r is 0.60), negatively correlated with log unemployment ($r = -0.24$), and weakly positively correlated with the agglomeration index ($r = 0.16$). Surprisingly, given Ciccone and Hall's (1996) recent work, the agglomeration index and nonwage income per capita are weakly *negatively* correlated, with $r = -0.09$.

²² Standard errors in these regressions were corrected for clustering with STATA's Huber routine.

School quality is another candidate third factor for which it might be necessary to control. Say that ability is evenly distributed across states, but that school quality positively affects both educational attainment and individual productivity; this would be the case, for instance, if school quality raised the marginal return to schooling--see Card and Krueger (1996) for a discussion. As a result, average education would enter the modified earnings regression as a proxy for school quality. I assess the importance of this by adding a measure of school quality to the cross-sectional regression (1). I have one year of data on school quality, namely, the pupil-teacher ratio that corresponds to the 1980 CPS sample. Including the school quality measure, however, has essentially no effect on the average education term.²³

Finally, I note that regional price dispersion might be driving the relationship between nominal wages and average education--in other words, states are "high-wage" states because their residents face a high cost of living. The only reason that I can think of for there being a correlation between price dispersion and statewide educational attainment is that states with better-educated populations tend to be richer, which acts to bid up the cost of living by raising the price of nontraded goods. Part of this effect will be captured by including nonwage income in the earnings regressions, but I cannot control directly for regional price dispersion because there is almost no information on the spatial variation of the cost of living in the U.S. What evidence there is, however, suggests that the nonhousing component of such variation is small, at least in urban areas (see Beeson and Eberts, 1989).

Results from including state fixed effects

The preceding results indicate that third factors do in fact play a large role in inducing the observed correlation between earnings and state average education, and suggest that controlling for unobserved time-invariant state characteristics (fixed effects) is likely to have an important effect as well. When I do so, the results change dramatically.

Table 3.2 reports the coefficient estimates that obtain from regressions which allow for state-specific intercept terms (the ϕ_s terms in equation (2), above). From column one of the table, it is apparent that including state fixed effects causes the coefficient on average education to vanish. The coefficient's point estimate remains more-or-less unchanged when other correlates are added to the

²³ In fact, including the pupil-teacher ratio usually acts to raise the value of the coefficient on average education, though the difference is not large. In addition, the coefficient on the pupil-teacher ratio is positive, which implies that higher school quality lowers average weekly wages, all else equal. These results are consistent with the story that school quality has amenity value, and that average education is an increasing function of school quality, which is in turn consistent with my data on state average education levels. Note that the school quality measure is cohort-specific, so a separate regression is run for each cohort; the cohort bands I use are ten years wide for all but the oldest cohort.

regression; it rises slightly when the state unemployment rate is included (columns four and seven), but falls when state nonwage income per capita is added. The coefficient is statistically significant at the five percent level in two specifications, but the point estimate is small--about a fifth to a quarter as large as the value that obtains when fixed effects are excluded. Importantly, it is the average education coefficient's point estimate--not its standard error--that changes when fixed effects are included in the pooled regressions.²⁴

Overall, the measured effects of the state characteristics (the Z_{st} terms) on earnings become stronger when we control for state fixed effects. In particular, the coefficient on state unemployment implies an unemployment elasticity of weekly wages of about -0.10 ; this is essentially the value found by Blanchflower and Oswald in their wage curve research. Note, however, that this elasticity drops sharply when nonwage income is included (columns six and eight).²⁵ The implication is that transitory fluctuations in the Z_{st} terms have larger effects on wages than permanent movements; this might reflect the fact that migration flows, and thus local labor supply, respond only slowly to interregional wage differentials.²⁶

Region-specific education returns and unrestricted education-earnings profiles

In estimating the least-squares dummy-variable (LSDV) versions of (1), I specified log individual earnings as a linear function of education, with the coefficient on own education constrained to be constant across states. There are at least two potential problems with this. First, it is known that the return to education varies by region (Chiswick (1974) provides an early study of this phenomenon). Second, Card and Krueger (1992) note that education-earnings profiles at the state level tend to be loglinear only after a certain minimum level of education is reached; moreover, this threshold varies according to the average level of educational attainment in the state.²⁷ Since recovering an accurate

²⁴ While we cannot identify just what it is that the fixed effects are controlling for in these regressions, one possibility is that the fixed effects are absorbing the persistent component of state well-being that influences residents' educational attainment decisions.

²⁵ Note also that Blanchflower and Oswald use a different measure of compensation (annual earnings) in their U.S.-based regressions. I have no strong intuition as to why including nonwage income should affect the "wage curve" elasticity to such an extent, but submit that it might be a further indication that the wage curve is truly a misspecified Phillips curve, with the income term serving as a summary statistic for statewide nominal aggregate demand.

²⁶ Blanchard and Katz (1992) provide evidence that interstate migration flows are indeed largely driven by factors other than wage differentials.

²⁷ There are also two "jumps" in the earnings-education profile at 12 and particularly 16 years of education. These "sheepskin effects" reflect the premium to receiving a high school or college diploma, and cause a departure from linearity that might be responsible for some portion of the measured relationship between state average education and individual wages. For

estimate of the return to own schooling is a critical step in estimating the pooled regressions (2), it is important to assess the effect of imposing a linear return to education that is constrained to be constant across regions.

I therefore examine the results that obtain from using three alternate specifications of the own education term. First, I reestimate the least-squares dummy-variable versions of regression (1) using a loglinear specification of the education-earnings relationship that allows the slope coefficient on education to vary over the four Census regions. The estimates of state-year fixed effects f_{st} thus obtained are then used to reestimate the pooled regression (2); Table 3.3 presents the results from these regressions. Incorporating a region-specific return to own education in the estimation procedure acts to raise the coefficient on average education; in some cases, the coefficient doubles in size.²⁸ In addition, the coefficients on the other state characteristics also tend to rise slightly.

Next, I consider reestimating the f_{st} terms using regressions that include a set of education indicator variables in lieu of the linear years-of-education term; Table 3.4 presents the estimated coefficients from the pooled regressions which employ these f_{st} values. The effect of including unrestricted education indicators is to increase the point estimates of the average education coefficients very slightly, at least in the specifications that control for state fixed effects. As a result, some of the coefficients become statistically significant; the remainder are borderline significant. The sharp decline that results from controlling for state fixed effects remains, however, as do the qualitative effects of including additional state-specific correlates in the regressions.

Finally, I estimate pooled regressions that use f_{st} terms from LSDV regressions which interact the education indicators with indicator variables for the four Census regions; Table 3.5 summarizes the results from these regressions. Here, allowing for regional differences in the return to education causes the point estimate of the coefficient on average education to fall in the specifications that include state fixed effects. None of the average education terms is statistically significant. In part, this is because the coefficients' standard errors are sharply higher in these regressions; however, the coefficients' reduced point estimates imply that they would not be statistically significant even if their standard errors had remained at the lower levels reported in Table 3.4.

Somewhat surprisingly, it turns out that the point estimates of the average education coefficients that obtain from the specifications that allow unrestricted education-earnings profiles with region interactions are virtually identical to those obtained from the specifications that constrain the

documentation of sheepskin effects at the aggregate level, see Park (1994).

²⁸ I provide an explanation of this result in section A.3 of the Appendix.

return to education to be constant over regions and loglinear. As noted in section A.3 of the Appendix, however, the data strongly reject the restriction that the return to education is constant across regions for both the loglinear and unrestricted specifications of the own education term. Furthermore, plots of unrestricted education-earnings profiles (not reported) indicate deviations from linearity at both the national and regional levels. I therefore think that allowing for an unrestricted education-earnings profile (with region interactions) is preferable, but am comforted by the fact that the relationship between average schooling and earnings changes little on net.²⁹

4. Potential sources of bias

There are at least two reasons to expect that the measured effect of state average education on individual earnings (the coefficient θ in regression (2), above) understates the magnitude of whatever true spillover effects might exist.

First, the presence of unobserved individual ability can in principle generate a negative correlation between average state education and own earnings; I demonstrate this point in section A.1 of the Appendix. Unfortunately, there is really no way to correct for unobserved ability, nor is there any way to get an idea of the size of the bias that it induces in the estimates of the state average education coefficient.

The second source of downward bias is related to a basic question of identification. An increase in the number of highly educated workers in a state raises that state's average education level. If spillovers in fact exist, this will raise the productivity and wages of workers in the state, all else equal. But one of the things which is *not* equal in this case is the supply of highly educated workers, which has risen. This higher supply acts to depress the wages of better-educated workers, offsetting the measured effect that higher statewide educational attainment has on individual earnings.³⁰

The point is that we can only identify human capital spillovers by controlling for this supply effect. A natural way to go about this involves measuring the effect that an increase in the number of highly educated workers in a state has on the wages of the less-educated. I do this by fitting LSDV versions of (1) to subsamples of less-educated workers--for example, workers with 12 years of education or less--and then using the state-year fixed effects from these regressions as the dependent

²⁹ This is not the case, however, for the regressions reported in the next section.

³⁰ See section A.4 of the Appendix for the requisite formalization.

variables f_{st} in regression (2). For the human capital measure H_{st} , I use the share of workers in a given state with a certain level of education or better--for example, the fraction of workers in a state with postcollege education.³¹ I also examine whether this variable has a smaller effect on the wages of the group over which it is defined, as would be the case were the measured human capital spillovers attenuated by supply effects.

As a representative example, consider the effect that workers with postcollege education have on the wages of those with less than 12 years of education, as summarized in Table 4.1.³² (Note that the earnings regressions (1) used to derive the f_{st} terms control for own education with a set of unrestricted indicator variables.) Taken at face value, these results imply that there is a positive relationship between individual wages and the fraction of workers in a state with postcollege education; this effect remains statistically significant even if time-invariant state fixed effects and other state characteristics are included in the regressions. These effects are not trivial--even the smallest point estimate of the education term (see column 2 of the table) implies that a one standard-deviation increase in the proportion of postcollege workers in a state raises the wages of workers with less than 12 years of education by a little over one log point.³³

This result is rather fragile, however; in particular, it is sensitive to the method used to control for own education in the earnings regressions (1). When I calculate the f_{st} terms using LSDV earnings regressions that allow the return to education to vary over regions, the effect that the fraction of highly educated workers has on the wages of less-educated workers vanishes (see Table 4.2). It therefore appears that the observed "spillovers" reflect a misspecification of the relationship between own education and earnings in the individual-level regressions.

³¹ Note that the existence of an intermediate "leave-out" group implies that the share of highly educated workers will *not* be perfectly correlated with the share of less-educated workers.

³² I have considered a number of other cases as well, for example, the effect of postcollege workers and college-educated (or better) workers on the wages of those with 12 years of education or less. The results are broadly similar to those reported here.

One problem with the underlying LSDV regressions used to construct the f_{st} terms is that the number of observations for smaller states in some years tends to be very small. This should be kept in mind when assessing these estimates.

³³ In addition, the change in the wage of workers with postcollege education that occurs when the fraction of workers with postcollege education rises (not reported) is also positive, but the effect is smaller and statistically insignificant. This is consistent with the view that the human capital spillovers induced by a greater number of highly educated workers are biased downward by supply effects for highly educated workers themselves.

5. Further extensions to the basic analysis

For the sake of completeness, I report the results from three other tests of whether the observed effect of state average education on individual earnings reflects actual human capital spillovers.

First, note that any reasonable story about human capital spillovers should predict that only the human capital of individuals *in the labor force* will have an effect on other workers' earnings. A natural test, then, is to construct measures of state average education that are calculated over two populations: workers, and individuals who are not in the labor force. We can then determine whether the average education of those not in the labor force has any effect on individual earnings.

For each state and year in my sample, I calculate the average education of individuals aged 25-65 who report zero weeks worked and zero wage and salary income in the previous calendar year. I then use this variable as the human capital measure H_{st} in the regressions (2); I also include it together with the original average education measure (the measure calculated over the sample of workers used to fit the earnings regressions). The results, which are summarized in Table 5.1, are not terribly interesting. By itself, the average education of nonworkers has a small, positive, and statistically significant effect on individual wages; the term's coefficient is about one-fourth as large as the coefficient on workers' average education.³⁴ This would argue against interpreting the effect of average education on earnings as reflecting human capital spillovers. However, this conclusion is tempered by the results from a horse race that includes both measures in the regression together. In those regressions, the coefficient on worker average education tends to rise slightly when nonworker average education is included. In addition, nonworker average education often enters the regressions with a small, negative, and statistically significant coefficient.³⁵ These results are probably largely driven by the fact that both average education measures are highly correlated with each other (the simple correlation coefficient r is 0.76), although they do suggest that average *worker* education is in fact the preferable measure.

A second extension involves estimating a series of cross-sectional earnings regressions on a subsample of individuals who have recently migrated to the state. If the effect of average state education on individual wages truly reflects human capital spillovers, then a new migrant to a state with

³⁴ The standard deviation of the not-in-labor-force average education measure is about 1.5 times as large as the standard deviation of the corresponding measure for workers. In addition, NILF education is less correlated than average education with log nonwage income per capita and the log agglomeration index, and more correlated with log state unemployment.

³⁵ Note that the f_{st} terms used in these regressions are obtained from individual-level earnings regressions in which the return to own education is constrained to be linear. They are therefore comparable to the results presented in Table 3.2.

a high average level of educational attainment should receive the same increase in earnings as a long-time resident. Note, however, that this argument can only be used to cast doubt on the existence of spillovers, since a positive effect of statewide human capital on migrants' wages still might merely reflect the presence of other productive state-specific characteristics. In addition, if human capital spillovers operate by allowing workers to acquire skills more quickly, as opposed to the case where spillovers merely induce a level shift in worker productivity, then we might argue that there is no reason to expect the wages of recent migrants to a human-capital rich region to manifest any increase.³⁶

Results from the migration regressions (not shown) are inconclusive, in part because the definition of the CPS's migration status variable changes several times over the period 1978-91 (this is what precludes my considering this issue using pooled cross-sectional regressions), and also because the subsample of migrants in a given year tends to be small. For surveys dated 1981-1984 and 1986-1991, the CPS asks whether a respondent has moved in the past year. For survey years 1978-1980, the CPS identifies migration status by asking about the respondent's residence in 1975. Finally, for the 1985 CPS, individuals are identified as migrants if they have moved in the past five years. As we would expect, the migration subsamples are much larger for the 1978-80 and 1985 surveys. In these years, as well in a few other survey years, the point estimates on the average education term are quite close to the corresponding estimates for the full sample (but the estimates are quite imprecise); in other years, the point estimate tends to be somewhat smaller, but is even less precisely estimated.

Finally, I note that the results presented in section three are somewhat reminiscent of the conclusions reached by *Bils and Klenow (1996)*. *Bils and Klenow* argue that the observed correlation between schooling and growth that appears in cross-country data is the result of faster economic growth's inducing greater demand for education. They base this conclusion in part on the results of two tests derived from the implications of a simple growth model that nests schooling-driven growth (endogenous growth) and growth-driven schooling (exogenous growth) as special cases.

First, under endogenous growth, a faster growth rate should act to reduce the observed return to experience sharply. *Bils and Klenow's* model implies that the offset should be one-for-one: one percentage point of faster growth should lead to a one percentage point reduction in the slope of the experience-earnings profile.³⁷ However, they find that there is essentially no relationship between the

³⁶ See *Glaeser and Maré (1994)* for suggestive evidence that a faster rate of skill acquisition in cities is the source of observed urban wage premia in cross-sectional data.

³⁷ The intuition is straightforward. If human capital accumulation drives growth, then young workers in countries with higher growth rates are entering the labor force with higher quality human capital stocks than older workers. This acts to reduce the difference in wages between old and young workers, implying a flatter experience-earnings profile in faster-growing countries. Of

slope of the experience-earnings profile and the growth rate of a country's real per-capita GDP over the preceding 20 years.

The second test that Bils and Klenow examine relates to the distinction between human capital stocks (average education levels) and flows (school enrollment rates). Say we consider the growth experience of a given country over two subperiods--for example, two decades. If growth drives schooling acquisition, say by lowering the effective discount rate for human capital investment, then the size of the schooling stock at the end of the first decade should be more highly correlated with economic growth over the first decade than with growth over the second decade. Likewise, if growth over the second decade is at least partially predictable, then faster growth should imply greater enrollment rates at the start of the second decade.

While a case can be made that the Bils-Klenow tests are inapplicable in this context--we are dealing with regions within a single country, as opposed to separate countries--I tried applying them anyway. I began by fitting separate earnings regressions for each state using data from the 1991 CPS; I allowed potential experience to enter as a quadratic. I then regressed the average annual growth rate of real gross state product per capita over the period 1970-90 on the derivative of the experience-earnings profile; I evaluated this derivative at several potential experience values. For white males, the slope of the experience-earnings profile in a state is essentially uncorrelated with the state's growth rate. For the full sample, the relationship between the slope of the experience-earnings profile is negative and statistically significant, but small--the largest negative coefficient is -0.25 , and most are below -0.15 . It would be difficult to argue that these results provide strong support for the proposition that human capital accumulation drives economic growth.

What does the stock-flow distinction tell us? The correlation between the growth rate of per capita gross state product over the period 1980-90 and state average education in 1980 is insignificant and negative ($r = -0.13$). However, the correlation between per capita GSP growth and state school enrollment rates is significant and positive ($r = 0.53$).³⁸ This result is consistent with the hypothesis that economic growth affects the schooling investment decision, *i.e.*, that growth causes schooling.

To what extent do these findings inform our interpretation of the observed correlation between average education and own earnings? Note first that Bils and Klenow's test of the endogenous growth

course, the one-for-one offset predicted by the Bils-Klenow model is an artifact of its specification, although the *sign* of the correlation between the experience-earnings profile and an economy's aggregate growth rate is a more general implication of an endogenous growth model in which schooling drives growth.

³⁸ Choice of periods was dictated by data availability. See section B.5 of the Appendix for a description of the sources and definitions of the measures used for these tests.

model is not intended to test for spillovers *per se*; rather, Bils and Klenow are interested in determining the nature of the accumulation constraint for human capital. However, if we interpret the state-level results above as providing evidence in favor of the hypothesis that faster growth causes greater schooling, we might be tempted to view the observed relation between state average education and own earnings as resulting from average education's being a proxy for the average level of (permanent) income or wealth in a region. I would be chary, though, of claiming too much for these results, especially because the periods considered here are probably dominated by cyclical (as opposed to trend) movements in GSP.³⁹

6. Summary and conclusions

This paper attempted to determine whether human capital spillovers of the form postulated by the new growth literature are observable in state-level U.S. data. I used a simple framework, which involved adding measures of statewide educational attainment to a standard Mincerian earnings regression. I then considered several tests in order to determine whether the measured positive relationship between state average education and individual earnings reflects actual external effects to human capital accumulation.

On balance, I believe that we can reject the hypothesis that human capital spillovers affect individual earnings. By controlling for time-invariant state fixed effects and a limited list of state characteristics, it is possible to cause almost all of the observed positive correlation between state average education and individual earnings to vanish. Altering the specification of the functional form relating own education to earnings in the individual-level wage regressions further erodes the correlation to the point where it is economically and statistically insignificant. I therefore conclude that much, if not all, of the observed correlation between state educational attainment and individual earnings is due to the fact that the average level of education in a state is a good proxy for other, truly productive factors.

³⁹ Moreover, it is not possible to duplicate the Bils-Klenow schooling-growth regressions on state-level data over the full period 1970-90, nor is it clear whether GSP growth rates for individual states are persistent (which is required in order to sensibly argue that residents base their schooling decisions on forecasted future GSP growth).

An interesting extension would involve using a longer period (or multiple decades) in the analysis. While real GSP data are unavailable before 1970, nominal state personal income data exist for much longer periods, and published Census tabulations could be used to calculate enrollment rates and average education statistics.

References

- Acemoglu, Daron, and Joshua Angrist, 2000, "How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws," Massachusetts Institute of Technology, processed. (March)
- Becker, Gary S., and Kevin M. Murphy, 1992, "The Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics*, **107**, 1137-60.
- Beeson, Patricia E., and Randall W. Eberts, 1989, "Identifying Productivity and Amenity Effects in Interurban Wage Differentials," *Review of Economics and Statistics*, **71**, 443-52.
- Bils, Mark, and Peter J. Klenow, 1996, "Does Schooling Cause Growth or the Other Way Around?" University of Rochester, processed. (January)
- Birdsall, Nancy, and Jere R. Behrman, 1984, "Does Geographical Aggregation Cause Overestimates of the Return to Schooling?" *Oxford Bulletin of Economics and Statistics*, **46**, 55-72.
- Blanchard, Olivier Jean, and Lawrence F. Katz, 1992, "Regional Evolutions," *Brookings Papers on Economic Activity*, **1**, 1-61.
- Blanchflower, David G., and Andrew J. Oswald, 1994, *The Wage Curve*. Cambridge, MA. MIT Press.
- Card, David, 1994, "Earnings, Schooling, and Ability Revisited," Princeton University, processed. (September)
- ___ 1995, "The Wage Curve: A Review," *Journal of Economic Literature*, **33**, 785-99.
- ___ and Dean Hyslop, 1997, "Does Inflation 'Grease the Wheels of the Labor Market'?" In Christina Romer and David Romer, eds., *Reducing Inflation: Motivation and Strategy*. Chicago, IL. University of Chicago Press.
- ___ and Alan B. Krueger, 1992, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, **100**, 1-40.
- ___ and ___ 1996, "Labor Market Effects of School Quality: Theory and Evidence," Princeton University, processed. (January)
- Chiswick, Barry R., 1974, *Income Inequality: Regional Analyses within a Human Capital Framework*. New York, NY. Columbia University Press for the National Bureau of Economic Research.
- Ciccone, Antonio, and Robert E. Hall, 1996, "Productivity and the Density of Economic Activity," *American Economic Review*, **86**, 54-70.
- Deaton, Angus S., 1997, *The Analysis of Household Surveys: Microeconometric Analysis for Development Policy*. Baltimore, MD. Johns Hopkins University Press.
- Glaeser, Edward L., and David C. Maré, 1994, "Cities and Skills," NBER Working Paper 4728, National Bureau of Economic Research, Cambridge, MA.

- Jovanovic, Boyan, and Rafael Rob, 1989, "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, **56**, 569-82.
- Lucas, Robert E., Jr., 1988, "On the Mechanics of Economic Development," *Journal of Monetary Economics*, **22**, 3-42.
- Moulton, Brent R., 1986, "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, **32**, 385-97.
- Munnell, Alicia H., 1990, "How Does Public Infrastructure Affect Regional Economic Performance?" Chapter in Munnell (ed.), *Is There a Shortfall in Public Capital Investment?* Boston, MA. Federal Reserve Bank of Boston.
- Park, Jin Heum, 1994, "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment," Princeton University, processed. (August)
- Rauch, James E., 1993, "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics*, **34**, 380-400.
- Roback, Jennifer, 1982, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, **90**, 1257-1278.
- Solow, Robert M., 1994, "Perspectives on Growth Theory," *Journal of Economic Perspectives*, **8**, 45-54.

Table 3.1**State-level pooled CPS regressions**

Full sample, constant own return-to-schooling coefficient
Lagged average education only (from first subsample)

Specifications *without* state fixed effects

<u>Variable</u>	<u>Specification</u>							
	1(a)	2(a)	3(a)	4(a)	5(a)	6(a)	7(a)	8(a)
ln nonwage (<i>Y/N</i>)		0.116** (0.015)			0.146** (0.011)	0.133** (0.016)		0.151** (0.012)
ln aggrind			0.034** (0.002)		0.036** (0.001)		0.034** (0.002)	0.035** (0.001)
ln unempl. rate				0.022 (0.013)		0.049** (0.012)	-0.015 (0.010)	0.014 (0.019)
Education measure	0.108** (0.009)	0.091** (0.009)	0.072** (0.007)	0.111** (0.009)	0.048** (0.007)	0.096** (0.009)	0.069** (0.008)	0.050** (0.007)
No. obs.	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged average education, first subsample							
Experience measure:	None							

Note: Specifications 1(a) through 8(a) correspond to specifications one through eight in later tables, except that the regressions reported here do *not* include indicator variables to capture state fixed effects (only indicator variables for year effects are included).

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 3.2

State-level pooled CPS regressions

Full sample, constant own return-to-schooling coefficient
Lagged average education only (from first subsample)

<u>Variable</u>	<u>Specification</u>								
	0	1	2	3	4	5	6	7	8
In nonwage (<i>Y/N</i>)			0.290** (0.013)			0.292** (0.013)	0.271** (0.017)		0.268** (0.017)
In aggrind				0.073* (0.030)		0.094** (0.023)		0.114** (0.027)	0.100** (0.023)
In unempl. rate					-0.108** (0.008)		-0.015 (0.009)	-0.111** (0.008)	-0.019* (0.009)
Education measure	0.108** (0.009)	0.017 (0.011)	0.010 (0.009)	0.020 (0.011)	0.020* (0.010)	0.014 (0.009)	0.011 (0.009)	0.025* (0.010)	0.015 (0.009)
No. obs.	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged average education, first subsample								
Experience measure:	None								

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 3.3**State-level pooled CPS regressions**

Full sample, region-specific own return-to-schooling coefficient
Lagged average education only (from first subsample)

<u>Variable</u>	<u>Specification</u>								
	0	1	2	3	4	5	6	7	8
In nonwage (<i>Y/N</i>)			0.354** (0.019)			0.358** (0.019)	0.334** (0.025)		0.330** (0.025)
In aggind				0.104* (0.041)		0.130** (0.033)		0.154** (0.037)	0.136** (0.033)
In unempl. rate					-0.130** (0.011)		-0.016 (0.013)	-0.135** (0.011)	-0.022 (0.013)
Education measure	0.169** (0.011)	0.037* (0.015)	0.028* (0.012)	0.041** (0.015)	0.040** (0.014)	0.034** (0.012)	0.029* (0.012)	0.047** (0.014)	0.035** (0.012)
No. obs.	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged average education, first subsample								
Experience measure:	None								

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 3.4**State-level pooled CPS regressions**

Full sample, unrestricted own return-to-schooling indicators
Lagged average education only (from first subsample)

<u>Variable</u>	<u>Specification</u>								
	0	1	2	3	4	5	6	7	8
In nonwage (<i>Y/N</i>)			0.283** (0.013)			0.286** (0.013)	0.264** (0.017)		0.261** (0.017)
In aggind				0.077** (0.030)		0.098** (0.022)		0.118** (0.026)	0.104** (0.022)
In unempl. rate					-0.106** (0.008)		-0.016 (0.009)	-0.110** (0.008)	-0.020* (0.009)
Education measure	0.105** (0.009)	0.021 (0.011)	0.015 (0.008)	0.025* (0.011)	0.024* (0.010)	0.018* (0.008)	0.015 (0.008)	0.029** (0.010)	0.020* (0.008)
No. obs.	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged average education, first subsample								
Experience measure:	None								

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 3.5

State-level pooled CPS regressions

Full sample, region-specific unrestricted own return-to-schooling indicators
Lagged average education only (from first subsample)

<u>Variable</u>	<u>Specification</u>								
	0	1	2	3	4	5	6	7	8
In nonwage (<i>Y/N</i>)			0.353** (0.057)			0.355** (0.057)	0.367** (0.076)		0.364** (0.076)
In aggind				0.071 (0.103)		0.097 (0.100)		0.115 (0.103)	0.095 (0.101)
In unempl. rate					-0.114** (0.030)		0.012 (0.040)	-0.118** (0.031)	0.007 (0.040)
Education measure	0.172** (0.018)	0.019 (0.039)	0.010 (0.037)	0.022 (0.039)	0.022 (0.038)	0.014 (0.038)	0.009 (0.038)	0.027 (0.038)	0.014 (0.038)
No. obs.	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged average education, first subsample								
Experience measure:	None								

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 4.1**State-level pooled CPS regressions**

Sample of workers with less than 12 years of education;
unrestricted own return-to-schooling indicators
Lagged fraction of workers with postcollege education (from first subsample)

<u>Variable</u>	<u>Specification</u>									
	0	1	2	3	4	5	6	7	8	
In nonwage (<i>Y/N</i>)			0.271** (0.020)			0.273** (0.020)	0.234** (0.027)			0.231** (0.027)
In aggrind				0.052 (0.039)		0.070* (0.034)		0.093** (0.036)		0.079* (0.034)
In unempl. rate					-0.109** (0.011)		-0.028* (0.014)	-0.112** (0.011)		-0.031* (0.014)
Education measure	1.940** (0.126)	0.597** (0.191)	0.492** (0.167)	0.640** (0.194)	0.585** (0.177)	0.549** (0.169)	0.503** (0.167)	0.660** (0.178)		0.568** (0.169)
No. obs.	686	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged fraction of workers with more than 16 years of education, first subsample									
Experience measure:	None									

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 4.2

State-level pooled CPS regressions

Sample of workers with less than 12 years of education;
 unrestricted region-specific own return-to-schooling indicators
 Lagged fraction of workers with postcollege education (from first subsample)

<u>Variable</u>	<u>Specification</u>								
	0	1	2	3	4	5	6	7	8
In nonwage (Y/N)			0.316** (0.054)			0.320** (0.054)	0.390** (0.074)		0.385** (0.074)
In aggind				0.138 (0.096)		0.159 (0.093)		0.169 (0.096)	0.145 (0.094)
In unempl. rate					-0.080** (0.028)		0.055 (0.038)	-0.085** (0.028)	0.048 (0.038)
Education measure	2.024** (0.253)	-0.108 (0.475)	-0.230 (0.463)	0.005 (0.481)	-0.117 (0.472)	-0.102 (0.468)	-0.253 (0.463)	0.021 (0.477)	-0.133 (0.469)
No. obs.	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91
Education measure:	Lagged fraction of workers with more than 16 years of education, first subsample								
Experience measure:	None								

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Table 5.1**State-level pooled CPS regressions**

Full sample

Average education of workers *versus*
average education of those not in the labor forceSpecification

<u>Variable</u>	0	1	2	3	4	5	6	7	8
Average educ, tot	0.108** (0.009)	0.017 (0.011)	0.010 (0.009)	0.020 (0.011)	0.020* (0.010)	0.014 (0.009)	0.011 (0.009)	0.025* (0.010)	0.015 (0.009)
Average educ, NILF	0.028** (0.006)	-0.011 (0.007)	-0.008 (0.005)	-0.011 (0.007)	-0.006 (0.006)	-0.009 (0.005)	-0.008 (0.005)	-0.007 (0.006)	-0.008 (0.005)
<i>Both measures included ("horse race"):</i>									
Average educ, tot	0.172** (0.014)	0.026* (0.012)	0.017 (0.009)	0.030* (0.012)	0.026* (0.011)	0.022* (0.009)	0.017 (0.009)	0.033** (0.011)	0.023* (0.009)
Average educ, NILF	-0.048** (0.008)	-0.016* (0.007)	-0.011* (0.005)	-0.017* (0.007)	-0.012 (0.006)	-0.013* (0.005)	-0.011* (0.005)	-0.014* (0.006)	-0.013* (0.005)
Z_{st} terms included ^a	---	---	1	2	3	1,2	1,3	2,3	1,2,3
No. obs.	686	686	686	686	686	686	686	686	686
CPS years	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91	78-91

^a Codes for state-specific characteristics (Z_{st} terms):

1. Log of nonwage income per capita
2. Log agglomeration index
3. Log unemployment rate

Note: Specifications one through eight include indicator variables for states and years. Specification zero includes indicator variables for years only.

Note: Standard errors in parentheses. ** denotes coefficient significant at the one percent level or better; * denotes coefficient significant at the five percent level or better.

Appendix

This Appendix is divided into three sections. The first collects several derivations and results that were too lengthy to include in the body of the text or as footnotes. The second part of the Appendix provides summary statistics and sources for the auxiliary data used in the earnings regressions; it also contains a discussion of some issues related to the human capital measure used in the regressions.

Section A: Supplemental derivations and results

1 *Estimation methodology.* Recall that the estimation procedure of section two required estimates of the state fixed effects f_{st} for each year; the manner in which I obtained these estimates involved fitting a separate least-squares dummy variable version of regression (1) in each year. A sufficient assumption in order for this methodology to be valid is simply that the coefficients on the individual characteristics X_{is} represent consistent estimators of the true coefficients. Assume the true model is

$$\ln y_{is} = \beta.X_{is} + \delta.Z_s + \theta.H_s + error , \quad (3)$$

where I have included the education term E_{is} in the vector of individual characteristics X_{is} in order to simplify the notation. The estimated least-squares dummy variable (LSDV) regression is

$$\ln y_{is} = b.X_{is} + f_s + error . \quad (4)$$

It is immediately evident that the estimated fixed effect f_s from the LSDV regression will be a consistent estimator of the true state-specific characteristics $\delta.Z_s + \theta.H_s$ so long as b in (4) is a consistent estimator of the true parameter β in (3).

The assumption that b converges in probability to β is likely to break down in an interesting way if the estimated return-to-schooling coefficient suffers from unobserved ability bias.⁴⁰ Assume, for simplicity, that the equation governing individual earnings takes the following form:

$$\ln y_{is} = \gamma_s + \beta.E_{is} + \alpha_{is} + \epsilon_{is} \quad (5)$$

where γ_s is a time-invariant state-specific fixed effect, E_{is} denotes worker is 's education, α_{is} denotes individual is 's ability (measured in units such that its coefficient in the earnings regression is unity), and ϵ_{is} is an error term. Ability is unobserved, so we run the regression

$$\ln y_{is} = g_s + b.E_{is} + e_{is} . \quad (6)$$

Note that the estimates of the state fixed effects, g_s , will be consistent estimates of γ_s so long as the true fixed effects are uncorrelated with the omitted ability term; for example, if the distribution of ability is random across states. However, the expected value of b is given by:

$$E(b) = \beta + 1 \cdot r_{\alpha E}$$

where $r_{\alpha E}$ is the (true) coefficient from a regression of ability on education. Note that if other correlates

⁴⁰ A similar point is made in a different context by Card and Krueger (1996).

are included in the regression, then $r_{\alpha E}$ is calculated conditioning on these additional variables.

Denote the cell (state) mean of a variable W as $W_{.s}$. Then, using the relations above, we have that the following is true:

$$\begin{aligned} E\{\ln y_{.s} - b.E_{.s}\} &= \gamma_s + \beta.E_{.s} + \alpha_{.s} - \beta.E_{.s} - r_{\alpha E}.E_{.s} \\ &= \gamma_s + \alpha_{.s} - r_{\alpha E}.E_{.s}, \end{aligned} \quad (7)$$

where $E\{\cdot\}$ denotes the expectations operator. Note that the term in curled brackets on the left-hand side of the first equality is the estimated state fixed effect g_s from (6). What this implies, then, is that a regression of the estimated fixed effect obtained from (6) on a state-specific fixed effect and average state education should yield a negative coefficient on average state education, assuming that ability and education are positively correlated.⁴¹ A corollary which has particular bearing on this paper's analysis is that the presence of unobserved ability bias should tend to attenuate the coefficient on average state education when it is used as the statewide human capital variable in equation (2), as can be shown by suitably modifying (5) to include average education as a state-specific determinant of individual earnings.

2 *Agglomeration-education interactions.* In section three, I pointed out that certain theoretical models of skill acquisition imply that human capital spillovers should be more pronounced in areas of dense economic activity. Intuitively, locating in a dense area raises the number of potentially profitable meetings with other skill- or knowledge-holders, while locating among an educated populace raises the probability of a given interaction's being productivity-enhancing. The observable implication is that there should be a positive interaction between agglomeration and education.

In the cross-sectional data, an interaction of the agglomeration and education variables suffers from being almost collinear with its components; to avoid this problem, I constructed an agglomeration indicator that is equal to one for states whose agglomeration index exceeds the median *national* value. I then included the agglomeration indicator and the agglomeration-education interaction term in the cross-sectional regressions. Unfortunately, the results from these regressions are inconclusive.

I also examined this issue using the longitudinal regressions (2). The results [not shown] from regressions that include an agglomeration-education interaction term, together with other state-specific correlates, are somewhat perverse inasmuch as both the education and agglomeration terms receive negative (and significant) coefficients. When I used the mean values of the variables to estimate the implied partial derivative of wages with respect to education or agglomeration, however, it became apparent that these regressions contain no new information; the effect of an additional year of statewide education or agglomeration is essentially equal to that implied by the estimates from the basic longitudinal regressions. In sum, then, I conclude that there is no evidence to support the proposition that the measured effects of the state education term are greater in dense areas.

3 *Justification for modelling region-specific returns to education.* The inclusion of a region-specific return to education has a profound effect on the results reported above. The effect of state average education on individual earnings becomes stronger when the loglinear return to own education is allowed to vary over regions (compare Tables 3.2 and 3.3). The opposite effect obtains, however, when unrestricted education-earnings profiles are interacted with region indicators: the average education term vanishes in these regressions (compare Tables 3.4 and 3.5).

There is strong statistical justification for using specifications that allow the return to schooling to vary over regions. *F*-tests of the hypothesis that returns to schooling are constant across regions are

⁴¹ Note that the fixed effects from this regression will absorb the true fixed effect γ_s , as well as average state ability $\alpha_{.s}$.

strongly rejected in all but one year for both the loglinear and unrestricted education-earnings profile regressions.

In the case of the loglinear regressions, it is also clear why the measured return to state average education should be attenuated when the return to schooling is constrained to be constant across regions. Education-earnings profiles for the West Census region tend to be flatter than those for the rest of the country (they also have a higher y-intercept). In addition, measured average educational attainment tends to be higher in the West. As a result, using a constant return-to-schooling coefficient in the LSDV regressions induces a negative correlation between average educational attainment and the estimated state fixed effects f_{st} --the values are too small for high-education regions like the West.

4 *Supply effects and the measurement of spillovers.* Section four argued that the measured effect of higher average education on individual earnings is likely to be attenuated by the presence of supply effects. This point can be made completely clear as follows. Assume a world in which there are two skill groups; denote the level of human capital (schooling) common to group i by S_i . Assume that group two is composed of "high-skill" workers, so $S_1 < S_2$. The total flow labor input of group i in efficiency unit terms is given by $\exp\{\beta_i \cdot S_i\} \cdot N_i$, where N_i is the number of workers in the group. Average schooling in a region, S^* , is therefore:

$$S^* = \omega_1 \cdot S_1 + \omega_2 \cdot S_2,$$

where ω_i is the labor-force weight of group i . I assume that an index of average education, $\exp\{\theta \cdot S^*\}$, acts as a Harrod-neutral shift term in the production function for total output, which is given as:

$$Y = F\left(e^{\theta S^*} \cdot e^{\beta_1 S_1} \cdot N_1, e^{\theta S^*} \cdot e^{\beta_2 S_2} \cdot N_2, K\right)$$

Assume labor supply is inelastic and exogenous. Then the marginal product of either group evaluated at the exogenously given supplies N_1^* and N_2^* determines the wages of each group:

$$MPL_i = F'_i(\cdot) \cdot e^{\theta S^*} \cdot e^{\beta_i S_i}$$

Note that the model has been specified in such a way as to yield loglinear earnings-education profiles.

From this it is immediately apparent that an increase in highly educated (type-two) workers can raise overall log wages by raising the average level of education S^* . The overall increase, however, does not equal the full spillover effect θ , since the rise in wages for type-two workers is diminished if the first partial derivative of the production function with respect to the second argument is negative (that is, if the marginal product of type-two workers falls as their number increases).

However, the effect of the additional number of high-skill workers on low-skill workers' wages *does* reflect the full spillover, with one (probably important) caveat--we have to assume that there are no substitution or complementarity effects between the two classes of workers (F_{12} must be zero). An equivalent assumption ($F_{13} = F_{23} = 0$) is required in order to unambiguously justify using nonwage income as a proxy for state well-being. None of these assumptions is particularly appealing on *a priori* grounds.

Section B: Data and measurement issues

1 *Measurement of average state education.* An issue that arises in constructing the average education measure used in the pooled regressions concerns timing. The March CPS asks respondents

to report their earnings from the previous calendar year; educational attainment, however, asks about completed years of schooling as of the date of the survey. Using an education measure calculated from the previous year's CPS, meanwhile, is complicated by the fact that the school year in the U.S. typically runs from September to June, although some college students graduate in January. On balance, I feel that using the previous year's human capital measure is best. Ideally, we want the human capital measure to reflect the human capital of those actively working during the period over which earnings are measured, and a lagged human capital measure likely has the most overlap with this period.⁴² In addition, using a lagged measure implies that the sample used to estimate the wage regressions and the sample used to calculate the human capital variables are different. Results from a number of cross-sectional regressions indicate that the estimated effect of average education on earnings is essentially unchanged no matter which measure is used.

2 *Use of average state potential experience.* In his 1993 paper, Rauch also uses average state potential experience as a measure of human capital. My experiments with average potential experience indicate that the effect which average potential experience appears to have on state wages results from its incorporating (by construction) the average age of workers in the state. Since including average worker age as a state-specific characteristic has little effect on the average education coefficient--average education and age are nearly orthogonal--I do not consider it further in the regressions above.

3 *Sources for and construction of auxiliary variables.*

Nonwage income per capita: Income measure from Bureau of Economic Analysis estimates of Gross State Product, *Survey of Current Business*, various issues. Population estimates by state from Bureau of Economic Analysis publication *Local Area Personal Income, 1969-92*.

State unemployment rate: Bureau of Labor Statistics LABSTAT database. I use the natural logarithm of the unemployment rate, following Blanchflower and Oswald (1994), chapter 4.

School quality: The pupil-teacher ratio was constructed from annual data kindly provided by Professor David Card, that were originally obtained from the *Biennial Survey of Education* and *Digest of Education Statistics*. Cohort- and state-specific measures were constructed following the procedure outlined in Card and Krueger (1992).

Nonfarm capital per worker: Capital stock figures are from Munnell (1990), and were kindly provided by Dr. John Greenlees, U.S. Department of the Treasury. Data on payroll employment are from the BLS *Handbook of Labor Statistics, 1988*. The capital stock measures were constructed by allocating the aggregate capital stock for the U.S. to individual states; in most cases, allocation is based on the states' shares of total capital at book value, as measured by the industrial censuses.

Agglomeration index: I define the agglomeration index as the weighted average of county population densities in a state, where the weights used are counties' shares of state population. (Simple population

⁴² An estimate from the previous CPS year will not include those who complete school in June of that year, which is reasonable since it is unlikely that those workers would enter employment immediately. An estimate from the current CPS, however, includes June graduates from the previous year as well as January graduates from the current year; most likely, both groups' size and influence on earnings in the previous year are small.

density is not a good measure of agglomeration; consider an example where all economic activity in a large state is concentrated in a single city. Measured population density would be low, even though the agglomeration in the state is high.) Population and population density estimates are from a database assembled by Professor Anne Case (I thank Professor Case for kindly allowing me to use these data), originally extracted from the *Consolidated City and County Databook*, and were extended using the *U.S.A. Counties* database. Population figures are intercensal estimates from the Bureau of the Census; 1980 county definitions were used throughout. The two sources of data were used to construct two indexes with overlap in 1980; the indexes were then ratio spliced. The simple cross-sectional correlation coefficient in the overlap year was in excess of 0.99.

The log agglomeration index used here is highly correlated with the index constructed by Ciccone and Hall (1996) for 1989, with a simple cross-sectional correlation coefficient in that year in excess of 0.9.

4 *Tables B.1 and B.2* provide summary statistics for the education measures and auxiliary variables used in the wage regressions.

5 *Enrollment rates, average state education, and gross state product (section five)*: School enrollment rates for 1970 and 1980 were calculated from published Census data and cover all individuals aged three to 34 years. In 1970, the enrollment rate itself is given by the Census tables. In 1980, the rate is calculated by dividing Census enrollment *levels* by an estimate of state population aged 3-34; the population figure is the Census estimate of the state population multiplied by the share of total state population aged 3-34 (as calculated from the CPS public-use tape).

State average education is defined as the average years of education of individuals aged 25 or older. In 1970, this measure is computed from categorical data published in the 1976 *Digest of Education Statistics*. When a category spanned more than one year, its midpoint was used. State average education in 1980 is calculated from the CPS public-use tape.

Real GSP growth rates were estimated using published and unpublished BEA data. (Note that these data were available for the 48 contiguous states only.)

Table B.1
Summary statistics for education measures

<u>CPS year</u>	Average educ. measure	Average educ., not in L.F.	Fraction of postcollege wkr.s
1978	12.352 (0.434)	10.928 (0.718)	0.070 (0.023)
1979	12.378 (0.463)	10.969 (0.694)	0.071 (0.025)
1980	12.516 (0.408)	11.034 (0.722)	0.076 (0.027)
1981	12.588 (0.405)	11.064 (0.695)	0.080 (0.031)
1982	12.635 (0.377)	11.133 (0.769)	0.078 (0.029)
1983	12.718 (0.419)	11.203 (0.692)	0.084 (0.025)
1984	12.814 (0.409)	11.216 (0.670)	0.088 (0.027)
1985	12.900 (0.380)	11.291 (0.600)	0.091 (0.029)
1986	12.916 (0.350)	11.312 (0.650)	0.090 (0.025)
1987	12.960 (0.341)	11.343 (0.678)	0.088 (0.023)
1988	13.028 (0.344)	11.433 (0.715)	0.092 (0.026)
1989	13.027 (0.353)	11.488 (0.651)	0.094 (0.028)
1990	13.044 (0.328)	11.445 (0.617)	0.093 (0.026)
1991	13.092 (0.282)	11.500 (0.642)	0.092 (0.027)
All years	12.783 (0.449)	11.240 (0.700)	0.085 (0.028)

Note: Figure given is the variable's mean across states (each state is treated as one observation); figure in parentheses is the variable's standard deviation.

Table B.2
Summary statistics for state-specific characteristics

<u>CPS year</u>	ln nonwage income per cap.	ln unempl. rate	ln agglom. index
1978	1.249 (0.208)	1.855 (0.262)	5.976 (1.573)
1979	1.373 (0.218)	1.692 (0.257)	5.979 (1.557)
1980	1.454 (0.229)	1.667 (0.262)	5.983 (1.544)
1981	1.501 (0.271)	1.892 (0.220)	5.987 (1.534)
1982	1.626 (0.300)	1.955 (0.271)	5.995 (1.528)
1983	1.642 (0.278)	2.202 (0.251)	6.001 (1.522)
1984	1.732 (0.232)	2.197 (0.273)	6.009 (1.518)
1985	1.862 (0.217)	1.943 (0.285)	6.016 (1.513)
1986	1.915 (0.207)	1.924 (0.269)	6.025 (1.509)
1987	1.949 (0.161)	1.887 (0.321)	6.033 (1.506)
1988	2.011 (0.161)	1.773 (0.341)	6.038 (1.502)
1989	2.070 (0.173)	1.641 (0.334)	6.042 (1.496)
1990	2.138 (0.170)	1.613 (0.243)	6.047 (1.490)
1991	2.171 (0.183)	1.667 (0.214)	6.052 (1.482)
All years	1.764 (0.359)	1.851 (0.328)	6.013 (1.506)

Note: Figure given is the variable's mean across states (each state is treated as one observation); figure in parentheses is the variable's standard deviation.