Lecture 11

Education

Today we start a three series lecture on education. The general outline is as follows:

The Human Capital Model

Heterogeneity in returns to schooling

    Instrumental Variables Approaches

Liquidity Constraints

Sub-optimal Individual choice

Evidence of Signalling (Acemoglu)

It's hard to know where to start with covering education. There has been so much written about it, and not just in economics. It's a hot topic because it's an easily identifiable public policy issue, with tons of data available for looking at it. The correlation between wages and schooling is one of the most striking relationships in labor economics. We spend 10 – 25 years in school. It's well worth understanding individual's reasons for schooling, its impact on individuals, the wage distribution, and externalities.

The standard human capital earnings function, or Mincerian model takes on the following form:

$$\log y_i = a + bS_i + cX_i + dX_i^2$$

where y_i is some earnings measure: sometimes hourly, weekly, annual. S_i is individual years of schooling. The U.S. census and BLS stopped asking number of years of school and switched to highest degree completed in 1990. Often this new definition is converted to years of schooling by guessing time to completion for each degree. X_i is experience, often measured as the number of predicted years in the labor force based on school attainment. This is 'potential experience', usually age – years of school – 6. This may not be an accurate measure as, more recently, individuals move

in and out of the labor market. The model assumes the same experience profile for all education levels. Recent research uses a quartic term for experience: X_i^3 and X_i^4

The Mincer equation assumes schooling only affects the experience profile in levels, and that the log earnings increases with schooling at a linear rate. This does not have to be the case, but empirically, these two functional form specifications seems to hold up reasonably well (e.g. see Card's Handbook Chapter).

Several issues arise in relation to this equation and the interpretation of b.

What is b?
1) an increase in productivity?
2) a signal of productivity?: employers use schooling to derive productivity expectations
3) omitted variables bias, measurement error?

Does b differ by:
4) SES background?
5) school quality?
6) level? (is the process really linear?)

7) How do people choose S?
8) Are there constraints in choosing S?
9) Are there externalities from choosing S?

The U.S. government spends about 4.5 percent of GDP on elementary and secondary education and 1 percent of GDP on post-secondary education. The U.S. spends about 6.5 percent of GDP on education annually. That's a lot of money. An important role in research of education is to answer the questions above to help determine how public funds might affect short and long run outcomes of interest.

Grilliches, 1977 Econometrica

Card's Model: (Econometrica, Handbook Chapter, and 'Earnings, Schooling, and Ability Revisited, 'Research in Labor Economics, Vol. 14, pp 23-48, 1995)

Grilliches's classic paper covers many of the central issues underlying the human capital earnings model: ability bias (both positive and negative), measurement error, and interpretation of the Mincerian returns to schooling. Card builds on this, and focuses on what measures of the effects of schooling are we interested in, what does OLS measures, IV, measure, and other empirical approaches.

Assume that individuals have an infinite planning horizon that starts at t = 0. Lifetime utility is:

$$V(S, c(t)) = \int_0^\infty u(c(t)) e^{-\rho t} dt$$

subject to the budget constraint:

$$\int_0^\infty c(t) e^{-Rt} dt = \int_S^\infty y(S) e^{-Rt} dt$$
$$= y(S) \frac{e^{-RS}}{R}$$

The model's been written up with earnings not a function of age. With schooling additively separable in age we can ignore earnings growth that does not depend on schooling. This is just a simplifying assumption.

As I've written it, s does not appear in the utility function. Therefore, the education decision is simply to choose how much to maximize the budget set: choose education to maximize income. The first order conditions are:

$$W'(S) = -R y(S) \frac{e^{-Rs}}{R} + y'(S) \frac{e^{-RS}}{R} = 0$$

Or: $\dfrac{y'(S)}{y(S)} = R$

Which states that individuals invest in schooling until the marginal return is equal to the interest rate. This expression is where the term 'returns to schooling' comes from. The left hand side of this expression is the marginal benefits from an investment in schooling (per dollar of foregone earnings) and the right hand side as the marginal cost (the opportunity cost of the dollar investment, in this case)

Suppose that $y(S) = e^{a_i + b_i S}$ describes the human capital production function. Then from F.O.C.,

$$\log y(S) = \alpha_i + b_i S = \alpha_i + RS$$

$$\dfrac{y'(S)}{y(S)} = b_i = R$$

With this functional form, either $b_i = R$, and an individual is indifferent to how much schooling they get, or $b_i > R$ and the individual obtains an infinite amount of schooling, or $b_i < R$ and the individual obtains no schooling. This peculiarity, with no solution to optimal schooling, arises because costs of schooling (R) are independent of the benefits. We need to introduce some curvature in either the marginal benefits or marginal costs of school to avoid corner solutions.

<draw figure>

Suppose instead that:

$$y(S) = e^{a_i + b_i S - \frac{1}{2} k_1 S^2},$$

So that the marginal increase in log earnings from schooling falls with additional schooling.

$$\frac{y'(S)}{y(S)} = b_i - k_1 S = R$$

Now we have a closed form solution for optimal schooling: $S^* = \dfrac{b_i - R}{k_1}$

<draw figure>

Heterogeneity in the return to schooling, b_i, leads to different optimal amounts.[1]

Heterogeneity in a_i does not affect schooling, but this is just because of the functional form. If

instead, $y(S) = e^{b_i S - \frac{1}{2} k_1 S^2} + A_i$, $\dfrac{y'(S)}{y(S)} = (b_i - k_1 S) \dfrac{e^{b_i S - \frac{1}{2} k_1 S^2}}{e^{b_i S - \frac{1}{2} k_1 S^2} + A_i} = R$

<draw figure>

Heterogeneity in the levels of initial earnings will tend to <u>lower</u> optimal schooling. Think Bill Gates, Mick Jagger, pro-athletes. This functional form is admittedly messy, but it does get across one of Grilliches' main points, which is that there are two forms of ability bias – differences in ability endowments (in the levels) which tends to lower optimal education attainment, and differences in the interaction between ability and schooling (in the slopes) which tends to raise optimal education.

There is no apriori reason for believing one dominates over the other, but most researchers are more concerned with ability bias from differences in b_i. As Card does, I will ignore the importance of ability differences in the levels.

---

[1] The relationship between log earnings and schooling does not appear, at first blush, to be linear, as in the Mincer equation. Sub in for S*, however, and notice that the ability bias in b leads to convexity in logy and S, while the quadratic expression leads to a concave relationship. The two may offset each other to generate a linear cross-section relationship with a heterogeneous population. See Card (95) for another paragraph and figure about this.

We can also introduce heterogeneity in costs. Let's modify the lifetime utility function to include distaste for school:

$$V(S,c(t)) = \int_0^S (u(c(t)) - \varphi_i(t)))e^{-\rho t}dt + \int_S^\infty u(c(t))e^{-\rho t}dt,$$

where $\varphi_i(t)$ is convex and reflects the disutility from school (we might run into trouble if individuals like school and $\varphi_i(t)$ was concave).

Now, the schooling decision affects utility directly, and we have to set up the Lagrangian when calculating the F.O.C. Let $u(c(t)) = \log c(t)$. The first order conditions are:

$$\frac{y'(S)}{y(S)} = R_i + \rho e^{-\rho S}\varphi_i(S) \approx r_i + k_2 S$$

I've also introduced heterogeneity in the interest rate. A high R_i is often used to proxy for liquidity constraints. Individuals that need to borrow to attend school but cannot easily do so can be modelled as individuals facing high values of R_i.

Individual differences in optimal schooling can arise from differences in the economic benefits from schooling, represented by:

$$\frac{y'(S)}{y(S)} = b_i - k_1 S$$

and differences in the marginal costs, represented by:

$$r_i + k_2 S$$

<draw figure>

$$S_i* = \frac{b_i - r_i}{k}, \ k = k_1 + k_2$$

Loosely speaking, variation in b_i corresponds to variation in 'ability', whereas variation in r_i corresponds to variation in 'access to funds' or 'tastes for schooling'. Note again that we are not looking at innate ability differences, independent of schooling (e.g. athletic ability), rather we are focussing on ability differences in the slope of the returns to schooling.

What are the main reasons for schooling differences? Note special case: when $r_i = \bar{r}$ and $k_2 = 0$, we have equality of opportunity: schooling differences only arise from ability differences, but costs associated with schooling the same (although, not clear what generates differences in b_i). Heckman suggests we have equality of opportunity, and b_i determined by young age.

What is the 'causal' effect of education?

The marginal effect of additional school on earnings, $b_i - k_1 S*_i$, differs across individuals. The average marginal return is $\bar{\beta} = E(b_i - k_1 S*_i) = \bar{b}\frac{k_2}{k} + \bar{r}\frac{k_1}{k}$. This is the expected increase in average log earnings if a random sample of the population acquired an additional unit of education. This is the 'Average Treatment Effect' (ATE) of schooling. The ATE may not be all that relevant for wanting to evaluate a particular schooling intervention that affects a specific sub-population. We might be interested in the average and local average impact from additional education, but also why education attainment differs in the first place: that is, how does b_i and r_i differ across the population and why?

Let's use the definition of $\bar{\beta}$ as a benchmark for comparing OLS and IV coefficients on the returns to education.

To see the implications of this when trying to estimate the average returns to schooling, consider this model implies a regression equation:

$$\log y_i = \alpha_i + b_i S_i - \frac{1}{2} k_1 S_i^2$$

$$= a_0 + \bar{b} S_i - \frac{1}{2} k_1 S_i^2 + a_i + (b_i - \bar{b}) S_i$$

where $a_i = \lambda_0 (S_i - \bar{S}) + u_i$, and $b_i - \bar{b} = \psi_0 (S_i - \bar{S}) + v_i$

λ_0 is the traditional 'ability' bias that is positive in this model when $\operatorname{cov}(r_i, a_i) < 0$ (and/or $\operatorname{cov}(b_i, a_i) > 0$). Note, λ_0 could be negative in the Grilliches case of Bill Gates and pro-athletes examples.

$$\lambda_0 = \frac{\operatorname{cov}(a_i, S_i)}{\operatorname{var}(S_i)} = \frac{\operatorname{cov}(a_i, (b_i - r_i)/k)}{\operatorname{var}(b_i - r_i)/k}$$

(check:)

$$= k \frac{\operatorname{cov}(a_i, b_i) - \operatorname{cov}(a_i, r_i)}{\operatorname{var}(b_i) + \operatorname{var}(r_i) + \operatorname{cov}(b_i, r_i)}$$

ψ_0 is the 'comparative advantage bias' that arises from differences in the slope of the earnings-schooling function.

Ordinary Least Squares: Our coefficient estimate for the relationship is:

$$\frac{\operatorname{cov}(\log_i, S_i)}{\operatorname{var}(S_i)} = \bar{b} - k_1 \bar{S} + \lambda_0 + \psi_0 \bar{S} ,$$

assuming that b_i and r_i are joinly symmetric, which means: $E[(b_i - \bar{b})^3] = E[(r_i - \bar{r})^3] = E[(b_i - \bar{b})(r_i - \bar{r})^2] = 0$. Use laws of expectations to verify this for yourselves (it's a bit of messy algebra).

$\bar{b} - k_1 \bar{S}$ is the ATE. Our estimate is biased (relative to ATE) by $\lambda_0 + \psi_0 \bar{S}$, A cross-sectional regression of earnings on schooling therefore also yields an upward biased estimate of the average

marginal return to schooling, even ignoring ability bias. Even without variation in the intercept of the earnings function, we over estimate the ATE with OLS (and Card was the first to point this out).

Instrumental Variables

Consider an exogenous change in college expenses, Z, as an instrument that affects r_i, but more for people with high values of r_i.

Let $k_1 = 0$. The log earnings, schooling equation is: $\log y_i = \alpha_i + b_i S_i + e_i$

Note, from a change in schooling: $\Delta \log y_i = b_i \Delta S_i$

With a valid instrument, Z,

$$p \lim b_{iv} = \frac{E[\log y_i \mid Z = 1] - E[\log y_i \mid Z_i = 0]}{E[S_i \mid Z_i = 1] - E[S_i \mid Z_i = 0]}$$

$$= \frac{E[\log y_i \mid Z = 0] - E[\log y_i \mid Z_i = 0] + E[\Delta \log y_i]}{E[S_i \mid Z_i = 0] - E[S_i \mid Z_i = 0] + E[\Delta S_i]}$$

$$= \frac{E[b_i \Delta S_i]}{E[\Delta S_i]}$$

Suppose the drop in college expenses leads to a fall in r and an increase in S, but mostly among persons with high B. If B is higher for people who tend to have high values of r, such an instrument will identify higher than average B (LATE). Lang (93), and Card refer to this as discount rate bias.

Card suggests this as a reason why IV estimates from supply-side variation tend to be higher than OLS estimates. (Measurement error might also be a possibility, but generally IV>OLS more than me would suggest).

Example: Card: Using Geographic Variation in College Proximity to Estimate the Return to Schooling

Card takes data from the U.S. National Longitudinal Survey of Young Men (NLSYM), which began with 5525 men aged 14-24 in 1966 and occasionally interviewed them later . 83% of these individuals were interviewed in 1976. The NLSYM contains some geographic information, including an indicator variable for whether an accredited 4-year college (university) is located close by. 70% of individuals did live close by to such a college.

Card's first stage equation is:

$$S_i = \gamma X_i + v_i$$

where X_i is a dummy variable for the presence of a nearby college. For this instrument to be valid, this variable must be correlated with education attainment, but uncorrelated with v, other factors that could affect both schooling and Y. Card's idea for this instrument is that students who grow up in an area without a college face a higher cost of college education, since the option of living at home is precluded. Once would expect this higher cost to reduce investments in higher education, at least among children from relatively low-income families.

We use the first stage to predict individual schooling, $\hat{S}_i = \hat{\gamma} X_i$, and then use these predictions to estimate the second stage:

$$Y_i = \beta_0 + \beta_1 \hat{S}_i + e_i$$

I want to emphasize here that we are estimating the average effect of schooling on earnings only off of the variation in schooling from living close to or far away from a 4-year college. Therefore, we are only estimating the average effect off of those affected by the instrument, and not the average effect for the entire sample. Who are the people affected by this instrument? Likely those from low-income parents who can't afford, or don't want to, send their children to a school out of town. Thus, for this group, the marginal cost to attend school may be higher than the average marginal cost if liquidity constraints play a role. The marginal benefits from education may also differ compared to the average in the sample, or for individuals under other circumstances. Instrumental variable approaches estimate causal effects only off of a particular group affected by the instrument. We therefore can only draw conclusions from our results for this group, and not for the rest of the sample.

What does Card find? Tables 3 and 4.

The potential problem with an instrumental variables strategy you should always first consider is the validity of the instrument. Is it really true that college proximity is unrelated to unobservable factors that correlate with Y? College proximity is obviously related to geography – in particular urban versus rural location. If innate ability or demographic characteristics that impact education attainment choices (but not through residence costs) are related to being in an urban or rural area, than this instrument is invalid – we still face omitted variable problems, and cannot draw definite conclusions from the results.

Card realizes this, and argues that if this instrument explains schooling differences only through the effect of lowering costs to some in attending school, then we should expect to see a first stage effect among individuals with low-income family background versus high-income family background.

See Figure 1: This is what he shows: The presence of a nearby college has its strongest effect on men with lowest propensities to continue their education (men from single headed families with low parental education in rural southern areas).

Is this convincing proof that the instrument is valid? If average family background characteristics for those living in rural or urban areas are the same regardless of income-class, then we would expect the same omitted variables bias when examining the effect of schooling from individuals of low or high income groups with this instrument. The finding that we do not suggests the omitted variables bias could be small, and we are, indeed, estimating the effect of schooling off of living away from a 4-year college. However, if differences in urban/rural background characteristics vary themselves by income group, then we may still face omitted variable bias in these estimates. If low income families living in urban areas are more likely to place a strong emphasis on education than those living in rural areas, but this is not true for high income families in urban and rural areas, Card's robustness check does not work. He tries to add additional controls, such as AFQT ability measures, and broader region controls. You decide.

Card finds that the IV estimates for the return to schooling are larger than the OLS results.

Possibilities:

1) heterogeneous returns to education, and marginal returns higher to those that obtain more school from being close to college
2) Measurement Error in OLS
3) OLS ability bias downwards (Grilliches argument)

More 'sinister' reasons:

4) 'specification bias': researchers working with IV and higher standard errors end up 'preferring' specifications that generate significant t-statistics, which tends to preferring higher estimates (see Ashenfelter)
5) Instruments are invalid