Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond

James Heckman
University of Chicago,
University College Dublin and
The American Bar Foundation

Econ 345
This draft, February 9, 2007

Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond

James J. Heckman
University of Chicago,
University College Dublin and
The American Bar Foundation

Lance Lochner
University of Western Ontario

Petra Todd
University of Pennsylvania

Part I

Introduction

Introduction

- Earnings functions are the most widely used empirical equations in labor economics and the economics of education.
- Almost daily, new estimates of "rates of return" to schooling are reported, based on numerous instrumental variable and ordinary least squares estimates.
- For many reasons, few of these estimates are true rates of return.

Introduction

- The internal rate of return to schooling was introduced as a central concept of human capital theory by Becker (1964).
- It is widely sought after and rarely obtained.
- Under certain conditions which we discuss in this chapter, high internal rates of return to education relative to those of other investment alternatives signal the relative profitability of investment in education.

Introduction

- Given the centrality of this parameter to economic policy making and the recent interest in wage inequality and the structure of wages, there have been surprisingly few estimates of the internal rate of return to education reported in the literature and surprisingly few justifications of the numbers that are reported as rates of return.
- The reported rates of return largely focus on the college-high school wage differential and ignore the full ingredients required to obtain a rate of return.
- The recent instrumental variable literature estimates various treatment effects which are only loosely related to rates of return.

Introduction

- In common usage, the coefficient on schooling in a regression of log earnings on years of schooling is often called a rate of return.
- In fact, it is a price of schooling from a hedonic market wage equation.
- It is a growth rate of market earnings with years of schooling and not an internal rate of return measure, except under stringent conditions which we specify, test and reject in this chapter.

Introduction

- The justification for interpreting the coefficient on schooling as a rate of return derives from a model by Becker and Chiswick (1966).
- It was popularized and estimated by Mincer (1974) and is now called the Mincer model.

Introduction

- This model is widely used as a vehicle for estimating "returns" to schooling quality, for measuring the impact of work experience on male-female wage gaps, and as a basis for economic studies of returns to education in developing countries.
- It has been estimated using data from a variety of countries and time periods.
- Recent studies in growth economics use the Mincer model to analyze the relationship between growth and average schooling levels across countries.

Introduction

- Using the same type of data and the same empirical conventions employed by Mincer and many other scholars, we test the assumptions that justify interpreting the coefficient on years of schooling as a rate of return.
- We exposit the Mincer model, showing conditions under which the coefficient in a pricing equation (the "Mincer" coefficient) is also a rate of return.
- These conditions are not supported in the data from the recent U.S. labor market.
- We then go on to summarize other methods that use repeated cross section and panel data to recover *ex ante* and *ex post* returns to schooling.

Introduction

- This lecture makes the following points:

Point 1

- We test important predictions underlying the Mincer model using six waves of U.S. Census data, 1940-1990.
- We find, as does other recent literature, that Mincer's original model fails to capture central features of empirical earnings functions in recent decades.
- The empirical analysis in this chapter is more comprehensive than previous analyses and tests more features of the model, including its predictions about the linearity of log earnings equations in schooling, parallelism in log earnings-experience profiles, and U-shaped patterns for the variance of log earnings over the life cycle.

Point 2

- In response to the evidence against the Mincer specification of the earnings function, we estimate more general earnings models, where the coefficient on schooling in a log earnings equation is not interpretable as a rate of return.
- From the estimated earnings functions, we compute marginal internal rates of return to education for black and white men across different schooling levels and for different decades.

Point 2

- Our estimates account for nonlinearities and nonseparabilities in earnings functions, taxes and tuition.
- A comparison of these estimated returns with estimated Mincer coefficients shows that both levels and trends in rates of return generated from the Mincer model are misleading.
- Caution must be used in applying the Mincer equation to modern economies to estimate rates of return.

Point 2

- The estimated marginal rates of return are often implausible, calling into question the empirical conventions followed by Mincer and the recent U.S. Census-based/Current Population Survey-based literature reviewed by Katz and Autor (1999) that ignore endogeneity of schooling, censoring and missing wages, uncertainty, sequential revelation of information and psychic costs of schooling.

Point 3

- We explore the importance of Mincer's implicit stationarity assumptions, which allowed him to use cross-section experience-earnings profiles as guides to the life cycle earnings of persons.
- In recent time periods, life cycle earnings-education-experience profiles differ across cohorts.
- Thus cross-sections are no longer useful guides to the life cycle earnings or schooling returns of any particular individual.

Point 3

- Accounting for the nonstationarity of earnings over time has empirically important effects on estimated rates of return to schooling.
- Since many economies have nonstationary earnings functions, these lessons apply generally.

Point 4

- Mincer implicitly assumes a world of perfect certainty about future earnings streams.
- We first consider a model of uncertainty in a static economic environment without updating of information, which can be fit on cross sections or repeated cross sections.
- Accounting for uncertainty substantially reduces high estimated internal rates of return to more plausible levels.
- These adjustments introduce *ex ante* and *ex post* distinctions into the analysis of the earnings functions, something missing in the Mincer model, but essential to modern dynamic economics.

Point 5

- We next consider a dynamic model of schooling decisions with the sequential resolution of uncertainty.
- Following developments in the recent literature, we allow for the possibility that, with each additional year of schooling, information about the value of different schooling choices and opportunities becomes available.
- This generates an option value of schooling.

Point 5

- Completing high school generates the option to attend college and attending college generates the option to complete college.
- Our findings suggest that part of the economic return to finishing high school or attending college includes the potential for completing college and securing the high rewards associated with a college degree.
- Both sequential resolution of uncertainty and non-linearity in returns to schooling can contribute to sizeable option values.

Point 5

- Accounting for option values challenges the validity of the internal rate of return as a guide to the optimality of schooling choices.
- The internal rate of return has been a widely sought-after parameter in the economics of education since the analysis of Becker (1964).
- When schooling decisions are made at the beginning of life, there is no uncertainty and age-earnings streams across schooling levels cross only once.

Point 5

- In this case, the internal rate of return (IRR) can be compared with the interest rate to produce a valid rule for making education decisions (Hirshleifer, 197).
- If the IRR exceeds the interest rate, further investment in education is warranted.
- However, when schooling decisions are made sequentially as information is revealed, a number of problems arise that invalidate this rule.
- We examine the consequences of option values in determining rates of return to schooling.

Point 5

- Our analysis points to a need for more empirical studies that incorporate the sequential nature of individual schooling decisions and uncertainty about education costs and future earnings to help determine their importance.
- We report evidence on estimated option values from the recent empirical literature using rich panel data sources that enable analysts to answer questions that could not be answered with the cross section data available to Mincer in the 1960s.

Point 6

- We then consider models that control for unobserved heterogeneity and endogeneity of schooling in computing "the rate of return to schooling" starting with the Card (1995, 1999) model and moving into the more recent analyses of Carneiro, Heckman, and Vytlacil (2005).
- These models focus on identifying the growth of earnings with respect to schooling (the causal effect of schooling) and not internal rates of return.

Point 6

- In many papers, an instrument, rather than some well-posed question, defines the parameter of interest.
- The models ignore the sequential resolution of uncertainty but account for heterogeneity in responses to schooling where "returns" are potentially correlated with schooling levels.

Point 6

- This correlation is ignored in the Census/CPS-based literature on “returns” to schooling.
- We review some new analytical results from the instrumental variables literature that aid in interpreting reported “Mincer coefficients” (growth rates of earnings in terms of years of schooling) within a willingness to pay framework.
- We link the rate of return literature to the recent literature on treatment effects.

Point 7

- The literature on the returns to schooling focuses on certain mean parameters.
- Yet the original Mincer (1974) model entertained the possibility that returns varied in the population.
- Chiswick (1974) and Chiswick and Mincer (1972) estimate variation in rates of return as a contributing factor to overall income inequality.

Point 7

- We survey recent developments in the literature that use rich panel data to estimate distributions of the response of earnings to schooling using the modern theory of econometric counterfactuals.
- They reveal substantial variability in *ex post* returns to schooling.

Point 8

- Finally, we review research from a very recent literature that decomposes variability in returns to schooling into components that are not forecastable by agents at the time they make their schooling decisions (uncertainty) and components that are predictable (heterogeneity).
- Both predictable and unpredictable components of *ex post* returns are found to be sizeable in most recent studies.

Point 8

- This analysis highlights the distinction between *ex ante* and *ex post* returns to schooling and the importance of accounting for uncertainty in the analysis of schooling decisions.
- This literature also identifies psychic costs of schooling, which are estimated to be substantial.
- Conventional rate of return calculations assume that they are negligible.
- These components help to explain why many people who might benefit financially from additional schooling do not take it up.

In this lecture

- We use the Mincer model as a point of departure because it is so influential.
- Mincer's model was developed to explain cross sections of earnings.
- While the model is no longer a valid guide for accurately estimating rates of return to schooling, the Mincer vision of using economics to explain earnings data remains valid.

In this lecture

- This chapter proceeds in the following way.
- We review two distinct theoretical arguments for using the Mincer regression model to estimate rates of return.
- They are algebraically similar but their economic content is very different.

In this lecture

- We present empirical evidence on the validity of the widely used Mincer specification.
- Using nonparametric estimation techniques, we formally test and reject key predictions of Mincer's model, while others survive.
- The predictions that are rejected call into question the practice of interpreting the Mincer coefficient as a rate of return.

In this lecture

- We extract internal rates of return from nonparametric estimates of earnings functions fit on cross sections.
- We show the effects on estimated rates of return of accounting for income taxes, college tuition and psychic costs, and length of working life that depends on the amount of schooling.
- We also consider how accounting for uncertainty affects estimated marginal internal rates of return.

In this lecture

- We introduce a dynamic framework for educational choices with sequential resolution of uncertainty, which produces an option value for schooling.
- We discuss why in such an economic environment the internal rate of return is no longer a valid guide for evaluating schooling investments.
- A more general measure of the rate of return used in modern capital theory is more appropriate.

In this lecture

- We consider the interpretation of Mincer regression estimates based on cross-section data in a changing economy.
- We contrast cross-sectional estimates with those based on repeated cross-sections drawn from the CPS that follow cohorts over time.
- Mincer's assumption that cross sections of earnings are accurate guides to the life cycles of different cohorts is not valid in recent years when U.S. labor markets have been changing.

In this lecture

- We discuss the recent literature on the consequences of endogeneity of schooling for estimating growth rates of earnings with schooling.
- We describe Card's (1999) version of Becker's Woytinsky Lecture (1967) and some simple instrumental variables (*IV*) estimators of the mean growth rate of earnings with schooling.

In this lecture

- We discuss the modern theory of instrumental variable estimation and interprets what *IV* estimates in the general case where growth rates of schooling are heterogeneous and potentially correlated with schooling levels.
- We consider what economic questions *IV* answers.
- The modern *IV* literature defines the parameter of interest by an instrument, rather than an economic question, and produces estimates of "rates of return" that have little to do with true rates of return.

In this lecture

- We survey a recent literature that estimates distributions of *ex post* returns.
- We decompose the distributions of returns and growth rates of earnings with schooling into *ex ante* and *ex post* components and presents option values for schooling as well as estimates of the psychic costs of schooling.
- Our analysis links the classical literature on rates of return to the modern literature on counterfactual analysis.

Part II

The Theoretical Foundations of Mincer's Earnings Regression

The Theoretical Foundations of Mincer's Earnings Regression

- The most widely used specification of empirical earnings equations and the point of departure for our analysis is the Mincer equation:

$$\ln[Y(s,x)] = \alpha + \rho_s s + \beta_0 x + \beta_1 x^2 + \varepsilon \qquad (1)$$

  where $Y(s,x)$ is the wage or earnings at schooling level s and work experience x, ρ_s is the "rate of return to schooling" (assumed to be the same for all schooling levels) and ε is a mean zero residual with $E(\varepsilon|s,x) = 0$.

- This regression model is motivated by two conceptually different frameworks used by Mincer (1958, 1974).
- While algebraically similar, their economic content is very different.

The Theoretical Foundations of Mincer's Earnings Regression

- We formally test and reject predictions of these models on the type of Census data originally used by Mincer.
- We implement a more general nonparametric approach to estimating internal rates of return that does not require an explicit model specification.

The Compensating Differences Model

- The original Mincer model (1958) uses the principle of compensating differences to explain why persons with different levels of schooling receive different earnings over their lifetimes.

- Individuals have identical abilities and opportunities, credit markets are perfect, the environment is perfectly certain, but occupations differ in the amount of schooling required.
- Individuals forego earnings while in school, but incur no direct costs.

The Compensating Differences Model

- Because individuals are *ex ante* identical, they require a compensating wage differential to work in occupations that require a longer schooling period.
- The compensating differential is determined by equating the present value of earnings streams net of costs associated with different levels of investment.
- This framework implicitly ignores uncertainty about future earnings as well as nonpecuniary costs and benefits of school and work, which are important determinants of the return to schooling and its distribution.

The Compensating Differences Model

- Let $Y(s)$ represent the annual earnings of an individual with s years of education, assumed to be constant over his lifetime.
- Let r be an externally determined interest rate and T the length of working life, assumed not to depend on s.
- The present value of earnings associated with schooling level s is

$$V(s) = Y(s) \int_s^T e^{-rt} dt = \frac{Y(s)}{r}(e^{-rs} - e^{-rT}).$$

The Compensating Differences Model

- Equilibrium across heterogeneous schooling levels requires that individuals be indifferent between schooling choices, with allocations being driven by demand conditions.
- Equating earnings streams across schooling levels and taking logs yields

$$\ln Y(s) = \ln Y(0) + rs + \ln((1 - e^{-rT})/(1 - e^{-r(T-s)})).$$

- The final term on the right-hand-side is an adjustment for finite life, which vanishes as T gets large.

The Compensating Differences Model

- This model implies that people with more education receive higher earnings.
- When T is large, the percentage increase in lifetime earnings associated with an additional year of school, ρ_s, must equal the interest rate, r.

The Compensating Differences Model

- Because the internal rate of return to schooling represents the discount rate that equates lifetime earnings streams for different education choices, it will also equal the interest rate in this model.
- Therefore, ρ_s in equation (1) yields an estimate of the internal rate of return, and when $\rho_s = r$, the education market is in equilibrium.
- If $\rho_s > r$, there is underinvestment in education.

The Accounting-Identity Model

- The model used by Mincer (1974), and now widely applied, is motivated differently from the compensating differences model, but yields an algebraically similar empirical specification of the earnings equation.
- It is much less clearly tied to an underlying optimizing model, although some of the assumptions are motivated by the dynamic human capital investment model of Ben-Porath (1967).

The Accounting-Identity Model

- Mincer's accounting identity model emphasizes life cycle dynamics of earnings and the relationship between observed earnings, potential earnings, and human capital investment, for both formal schooling and on-the-job investment.
- Persons are *ex ante* heterogeneous, so the compensating differences motivation of the first model is absent.
- ρ_s varies in the population to reflect heterogeneity in returns.

The Accounting-Identity Model

- Let P_t be potential earnings at age t, and express costs of investments in training C_t as a fraction k_t of potential earnings, $C_t = k_t P_t$.
- Let ρ_t be the *average* return to training investments made at age t.
- Potential earnings at t are

$$P_t \equiv P_{t-1}(1 + k_{t-1}\rho_{t-1}) \equiv \prod_{j=0}^{t-1}(1 + \rho_j k_j)P_0.$$

The Accounting-Identity Model

- Formal schooling is defined as years spent in full-time investment ($k_t = 1$), which is assumed to take place at the beginning of life and to yield a rate of return ρ_s that is constant across all years of schooling.

The Accounting-Identity Model

- Assuming that the rate of return to post-school investment is constant over ages and equals ρ_0, we can write

$$\begin{aligned} \ln P_t &\equiv \ln P_0 + s\ln(1+\rho_s) + \sum_{j=s}^{t-1} \ln(1+\rho_0 k_j) \\ &\approx \ln P_0 + s\rho_s + \rho_0 \sum_{j=s}^{t-1} k_j, \end{aligned}$$

where the last approximation is obtained for "small" ρ_s and ρ_0.

The Accounting-Identity Model

- Mincer approximates the Ben-Porath (1967) model by assuming a linearly declining rate of post-school investment: $k_{s+x} = \kappa\left(1 - \frac{x}{T}\right)$ where $x = t - s \geq 0$ is the amount of work experience as of age t.
- The length of working life, T, is assumed to be independent of years of schooling.
- Under these assumptions, the relationship between potential earnings, schooling and experience is given by

$$\ln P_{x+s} \approx \ln P_0 + s\rho_s + \left(\rho_0\kappa + \frac{\rho_0\kappa}{2T}\right)x - \frac{\rho_0\kappa}{2T}x^2.$$

The Accounting-Identity Model

- Observed earnings are potential earnings less investment costs, producing the relationship for observed earnings known as the Mincer equation,

$$\begin{aligned} & \ln Y(s,x) \\ & \quad \approx \ln P_{x+s} - \kappa\left(1 - \frac{x}{T}\right) \\ & \quad = [\ln P_0 - \kappa] + \rho_s s + \left(\rho_0\kappa + \frac{\rho_0\kappa}{2T} + \frac{\kappa}{T}\right)x - \frac{\rho_0\kappa}{2T}x^2. \end{aligned}$$

- This expression is equation (1) without an error term.

The Accounting-Identity Model

- Log earnings are linear in years of schooling, and linear and quadratic in years of labor market experience.
- Parameter ρ_s is an *average rate of return across* all schooling investments and not, in general, an internal rate of return or a marginal return that is appropriate for evaluating the optimality of educational investments.

The Accounting-Identity Model

- In many studies (see, e.g. Psacharopoulos, 1981, and Psacharopoulos and Patrinos, 2004), estimates of ρ_s are simply referred to as "rates of return" without any justification for doing so.
- In this formulation, ρ_s is the *ex post average* growth rate of earnings with schooling.
- It communicates how much average earnings increase with schooling, but it is not informative on the optimality of educational investments which requires knowledge of the *ex ante* marginal rate of return.

The Accounting-Identity Model

- In most applications of the Mincer model, it is assumed that the intercept and slope coefficients in equation (1) are identical across persons.
- This implicitly assumes that P_0, κ, ρ_0 and ρ_s are the same across persons and do not depend on the schooling level.

The Accounting-Identity Model

- However, Mincer formulates a more general model that allows for the possibility that κ and ρ_s differ across persons, which produces a random coefficient model,

$$\ln Y(s_i, x_i) = \alpha_i + \rho_{si} s_i + \beta_{0i} x_i + \beta_{1i} x_i^2$$

Letting $\bar{\alpha} = E(\alpha_i)$, $\bar{\rho}_s = E(\rho_{si})$, $\bar{\beta}_0 = E(\beta_{0i})$, $\bar{\beta}_1 = E(\beta_{1i})$, we may write this expression, dropping individual subscript "i", as

$$\begin{aligned} &\ln Y(s,x) \\ &\quad = \bar{\alpha} + \bar{\rho}_s s + \bar{\beta}_0 x + \bar{\beta}_1 x^2 \\ &\qquad\qquad + [(\alpha - \bar{\alpha}) + (\rho_s - \bar{\rho}_s)s + (\beta_0 - \bar{\beta}_0)x + (\beta_1 - \bar{\beta}_1)x^2], \end{aligned}$$

where the terms in brackets are part of the error.

The Accounting-Identity Model

- Mincer originally assumed that $(\alpha - \bar{\alpha}), (\rho_s - \bar{\rho}_s), (\beta_0 - \bar{\beta}_0), (\beta_1 - \bar{\beta}_1)$ are independent of (s, x); although he relaxes this assumption in later work (Mincer, 1997).
- Allowing for correlation between ρ_s and s motivates an entire instrumental variables literature which we survey below.

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- Both Mincer models predict that log earnings are linear in years of schooling although the two models have very different economic content.
- We test and reject this prediction on widely used Census and CPS data.

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- Assuming that post-school investment patterns are identical across persons and do not depend on the schooling level, the accounting identity model also predicts that

(i)

log-earnings experience profiles are parallel across schooling levels

$$\left(\frac{\partial \ln Y(s,x)}{\partial s \partial x} = 0 \right)$$

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- and

(ii)

log-earnings age profiles diverge with age across schooling levels

$$\left(\frac{\partial \ln Y(s,x)}{\partial s \partial t} = \frac{\rho_0 \kappa}{T} > 0 \right).$$

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- We extend Mincer's original empirical analysis of white males from the 1960 Census to white and black males from the 1940-1990 Censuses.
- The data from the 1940-1950 Censuses provide some empirical support for predictions (*i*) and (*ii*).
- The 1960 and 1970 data are roughly consistent with the model; prediction (*i*) does not pass conventional statistical tests for whites, although they pass an "eyeball" test.
- Data from the more recent Census years (1980-1990) are much less supportive of these predictions of the model, due in large part to the nonstationarity of recent labor markets.

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- Another implication of Mincer's model is that for each schooling class, there is an age in the life cycle at which the interpersonal variance in earnings is minimized.
- Consider the accounting identity for observed earnings in levels at experience x and schooling s, which we can write as

$$Y(s,x) \equiv P_s + \rho_s \sum_{j=s}^{s+x-1} C_j - C_{s+x}.$$

- This says that earnings at schooling level s equals initial endowment from schooling plus the return on past investments less the cost of current investment at age $s+x$ or experience class x.

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- In logs,

$$\ln Y(s,x) \approx \ln P_s + \rho_s \sum_{j=0}^{x-1} k_{s+j} - k_{s+x}.$$

- Interpersonal differences in observed log earnings of individuals with the same P_0 and ρ_s arise because of differences in $\ln P_s$ and in post-school investment patterns as determined by k_j.
- When $\ln P_s$ and κ or the k_{s+j} are uncorrelated, the variance of log earnings reaches a minimum when experience is approximately equal to $1/\rho_0$.

Derivation of the Overtaking Age

- Based on the text,

$$\begin{aligned} \ln Y(s,x) &= \ln P_{s+x} + \ln(1-k_{s+x}) \\ &\approx \ln P_s + \rho_0 \sum_{j=0}^{x-1} k_{s+j} - k_{s+x} \end{aligned}$$

Further using the assumption of linearly declining investment yields

$$\ln Y(s,x) \approx \ln P_s + \kappa \left(\rho_0 \sum_{j=0}^{x-1} (1 - j/T) - (1 - x/T) \right).$$

Derivation of the Overtaking Age

- Assuming only initial earnings potential (P_s) and investment levels (κ) vary in the population, the variance of log earnings is given by

$$
\begin{aligned}
&\operatorname{Var}(\ln Y(s,x)) \\
&\quad = \operatorname{Var}(\ln P_s) \\
&\qquad + \left(\rho_0 \sum_{j=0}^{x-1}(1-j/T) - (1-x/T)\right)^2 \operatorname{Var}(\kappa) \\
&\qquad + 2\left(\rho_0 \sum_{j=0}^{x-1}(1-j/T) - (1-x/T)\right) \operatorname{Cov}(\ln P_s, \kappa).
\end{aligned}
$$

Derivation of the Overtaking Age

- If κ and $\ln P_s$ are uncorrelated, then earnings are minimized (and equal to $\mathrm{Var}(\ln P_s)$) when

$$\rho_0 \sum_{j=0}^{x-1} (1 - j/T) = 1 - x/T, \qquad \text{or}$$

$$\rho_0 \left(x - \frac{x(x-1)}{2T} \right) = \rho_0 \left(1 - x/T\right).$$

- Clearly, $\lim_{T \to \infty} x^* = \frac{1}{\rho_0}$, so the variance minimizing age is $\frac{1}{\rho_0}$ when the work-life is long.

Derivation of the Overtaking Age

- More generally, re-arranging terms and solving for the root of this equation yields the variance minimizing experience level of

$$
\begin{aligned}
x^* &= T + \frac{1}{2} + \frac{1}{\rho_0} - \sqrt{\left(T + \frac{1}{2} + \frac{1}{\rho_0}\right)^2 - \frac{2T}{\rho_0}} \\
&\approx \left(\rho_0 + \frac{\rho_0}{2T} + \frac{1}{T}\right)^{-1},
\end{aligned}
$$

  where the final approximation comes from a first order Taylor approximation of the square root term around the squared term inside.

- The approximation suggests that the variance minimizing age will generally be less than or equal to $\frac{1}{\rho_0}$, with the difference disappearing as T grows large.

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- At this experience level, variance in earnings is solely a consequence of differences in schooling levels or ability and is unrelated to differences in post-school investment behavior.
- Prior to and after this time period (often referred to as the 'overtaking age'), there is an additional source of variance due to differences in post-school investment.

Implications for log earnings-age and log earnings-experience profiles and for the interpersonal distribution of life-cycle earnings

- Thus, the model predicts

(iii)

the variance of earnings over the life cycle has a U-shaped pattern.

- We show that this prediction of the model is supported in Census data from both early and recent decades.

Part III

Empirical Evidence on the Mincer Model

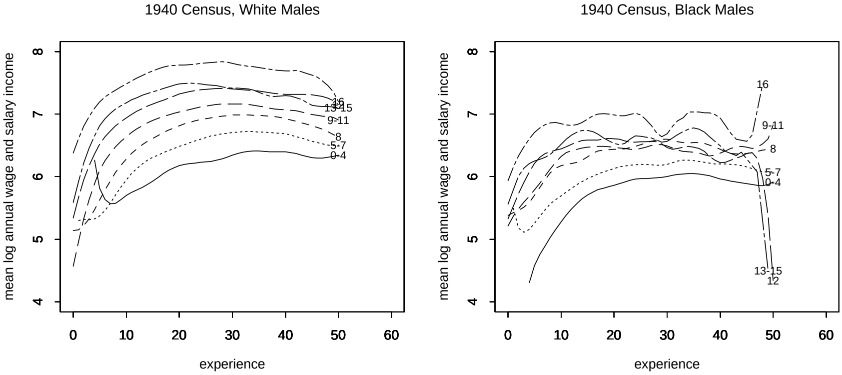
Empirical Evidence on the Mincer Model

- We now examine the empirical support for three key implications of Mincer's accounting identity model given above by (*i*), (*ii*), and (*iii*) using data on white and black males from the 1940-1990 decennial Censuses.
- Mincer conducted his original studies on Census and CPS data.
- Earnings correspond to annual earnings, which includes both wage and salary income and business income.

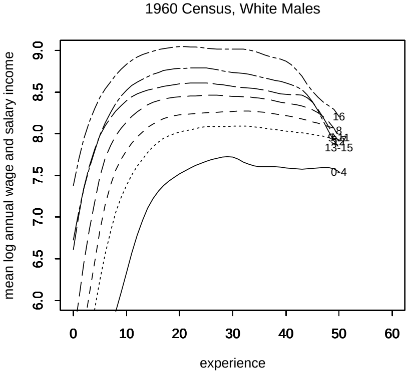
Empirical Evidence on the Mincer Model

- Figures 1a and 1b present nonparametric estimates of the experience – log earnings profiles for each of the Census years for white and black males.
- Nonparametric estimates of the age – log earnings profiles are shown for 1940, 1960 and 1980 in Figure 2.
- These estimates are based on a synthetic cohort assumption: that the cross-section is a guide to the life cycle of individuals.
- We question the validity of this assumption as a characterization of the recent U.S. labor market.

Figure 1a: Experience-Earnings Profiles, 1940-1960



1950 Census, White Males
mean log annual wage and salary income
experience
16
13-15
9-11
5-7
12
0-4
1950 Census, Black Males
mean log annual wage and salary income
experience
9-11
8
5-7
0-4
12
13-15


1960 Census, White Males
mean log annual wage and salary income
experience
0
10
20
30
40
50
60
6.0
6.5
7.0
7.5
8.0
8.5
9.0
16
8
9-11
12
13-15
0-4

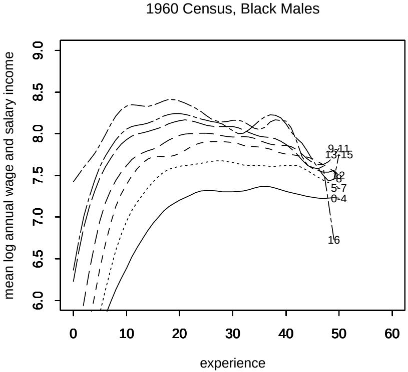
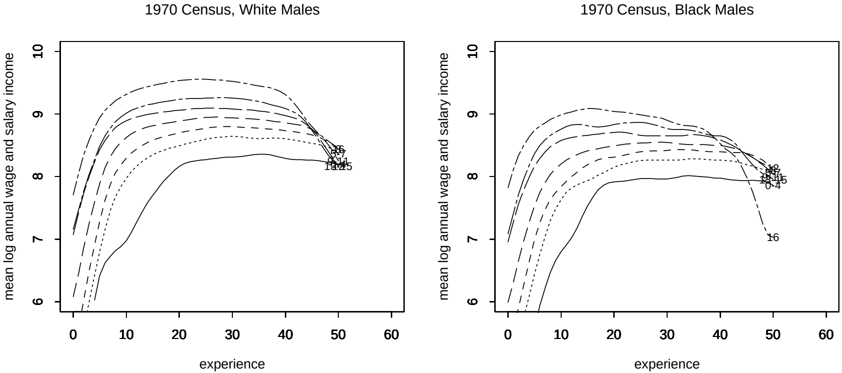
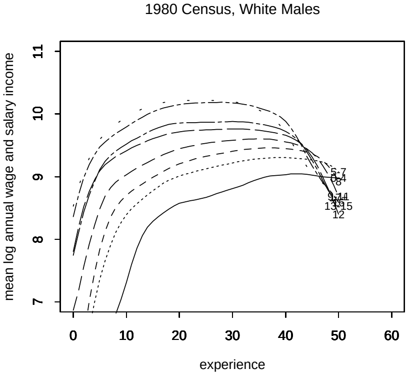
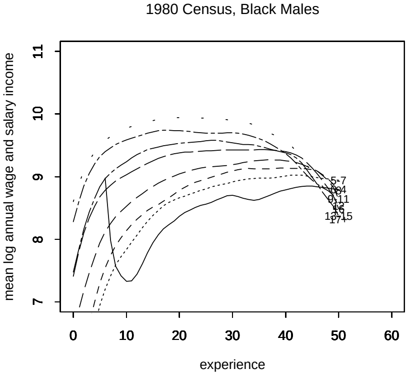
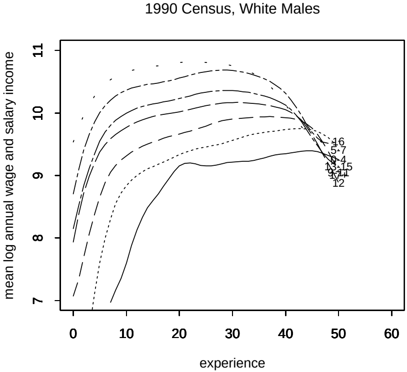

1960 Census, Black Males
mean log annual wage and salary income
experience
0
10
20
30
40
50
60
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9-11
13-15
12
8
5-7
0-4
16

Figure 1b: Experience-Earnings Profiles, 1970-1990




1980 Census, White Males
mean log annual wage and salary income
experience
7
8
9
10
11
0
10
20
30
40
50
60
5-7
0-4
8
9-11
13-15
12


1980 Census, Black Males
mean log annual wage and salary income
experience
7
8
9
10
11
0
10
20
30
40
50
60
5-7
0-4
8
9-11
12
16
13-15
17+


1990 Census, White Males
mean log annual wage and salary income
experience
7
8
9
10
11
0
10
20
30
40
50
60
16
5-7
0-4
13-15
9-11
17+
12

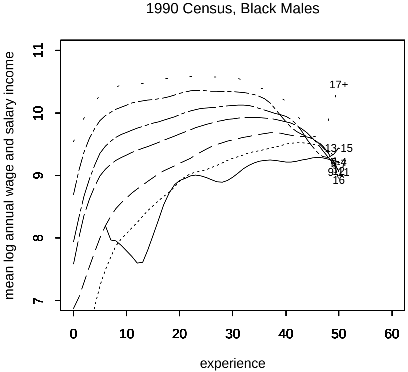
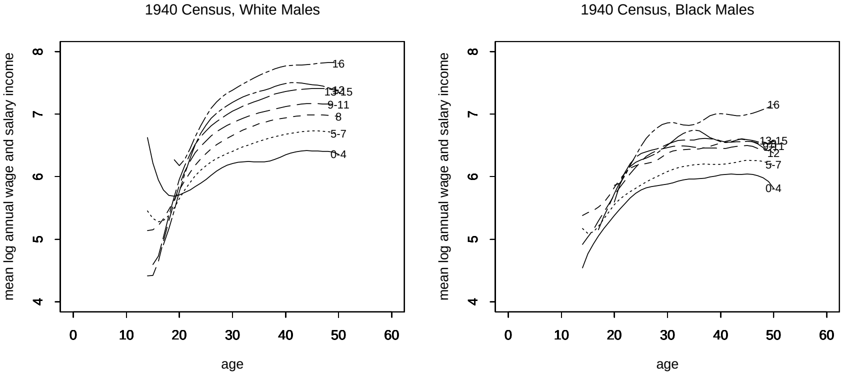
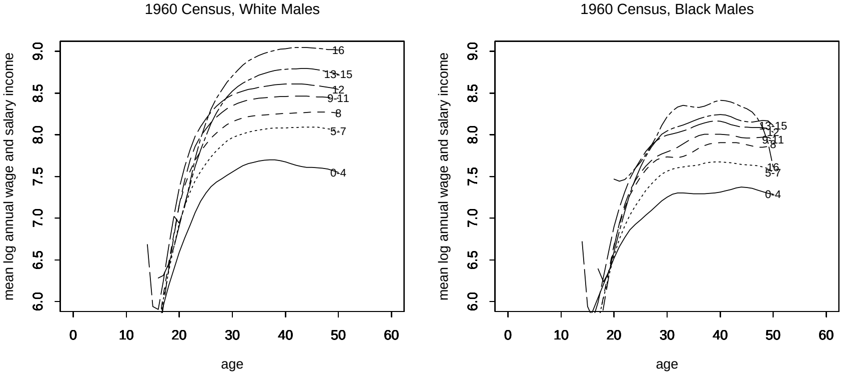
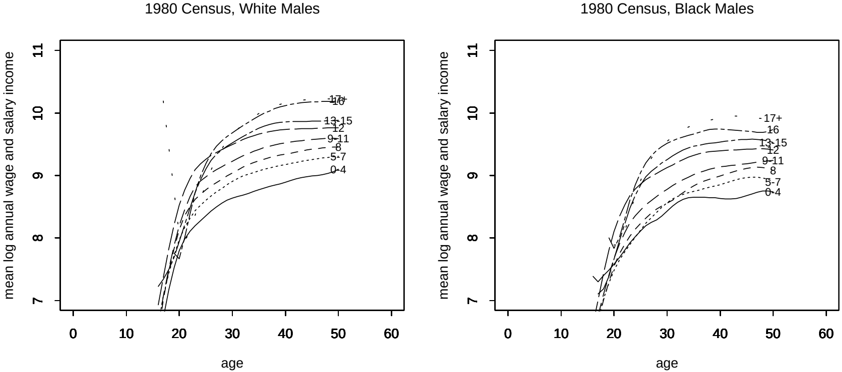

1990 Census, Black Males
mean log annual wage and salary income
experience
7
8
9
10
11
0
10
20
30
40
50
60
17+
13-15
9-11
16

Figure 2: Age-Earnings Profiles, 1940,1960,1980




1960 Census, White Males
mean log annual wage and salary income
6.0
6.5
7.0
7.5
8.0
8.5
9.0
0
10
20
30
40
50
60
age
16
13-15
12
9-11
8
5-7
0-4
1960 Census, Black Males
mean log annual wage and salary income
6.0
6.5
7.0
7.5
8.0
8.5
9.0
0
10
20
30
40
50
60
age
13-15
12
9-11
8
16
5-7
0-4


1980 Census, White Males
mean log annual wage and salary income
age
17+
16
13-15
12
9-11
8
5-7
0-4
1980 Census, Black Males
mean log annual wage and salary income
age
17+
16
13-15
12
9-11
8
5-7
0-4

Empirical Evidence on the Mincer Model

- Nonparametric local linear regression is used to generate the estimates.
- The estimated profiles for white males from the 1940-1970 Censuses generally support the hypothesis of the fanning-out by age and the parallelism by experience patterns (implications (*i*) and (*ii*) above) predicted by the accounting identity model.

Empirical Evidence on the Mincer Model

- For black males, the patterns are less clear, partly due to the much smaller sample sizes which result in less precise estimates.

- For 1960 and 1970, when the sample sizes of black males are much larger relative to earlier years, experience - log earnings profiles for black males show convergence across education levels over the life cycle.

Empirical Evidence on the Mincer Model

- Log earnings-experience profiles for the 1980-1990 Censuses show convergence for both white and black males.
- Thus, while data from the 1940-1950 Censuses provide support for implications (*i*) and (*ii*) of Mincer's model, the evidence for implication (*i*) is weaker for 1960 and 1970.
- The data from 1980 and 1990 do not support the model.

Empirical Evidence on the Mincer Model

- Formal statistical tests, reported in Table 1, reject the hypothesis of parallel experience - log earnings profiles for whites during all years except 1940 and 1950.
- Thus, even in the 1960 data used by Mincer, we reject parallelism, although it appears roughly consistent with the data.
- For black males, parallelism is only rejected in 1980 and 1990, although the samples are much smaller.

Table 1: Tests of Parallelism in Log Earnings Experience Profiles for Men

Sample	Experience Level	Estimated Difference Between College and High School Log Earnings at Different Experience Levels 1940	1950	1960	1970	1980	1990
Whites	10	0.54	0.30	0.46	0.41	0.37	0.59
	20	0.40	0.40	0.43	0.49	0.45	0.54
	30	0.54	0.27	0.46	0.48	0.43	0.52
	40	0.58	0.21	0.50	0.45	0.27	0.30
	p-value	0.32	0.70	<0.001	<0.001	<0.001	<0.001
Blacks	10	0.20	0.58	0.48	0.38	0.70	0.77
	20	0.38	0.05	0.25	0.22	0.48	0.69
	30	-0.11	0.24	0.08	0.33	0.36	0.53
	40	-0.20	0.00	0.73	0.26	0.22	-0.04
	p-value	0.46	0.55	0.58	0.91	<0.001	<0.001

Notes: Data taken from 1940-90 Decennial Censuses without adjustment for inflation. Because there are very few blacks in the 1940 and 1950 samples with college degrees, especially at higher experience levels, the test results for blacks in those years refer to a test of the difference between earnings for high school graduates and persons with 8 years of education. See Appendix B for data description. See Appendix C for the formulae used for the test statistics.

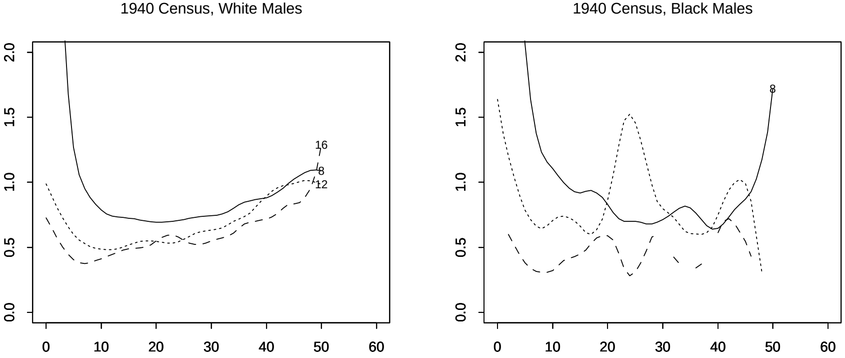
Empirical Evidence on the Mincer Model

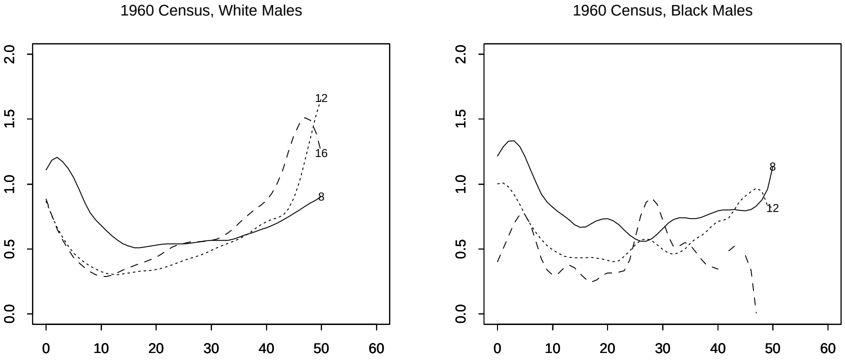
- We also formally test the hypothesis that log earnings are linear in education and quadratic in experience against an alternative that allows the coefficient on education to differ across schooling levels.
- The hypothesis of linearity is rejected for all Census years and for both blacks and whites (p-values $< .001$).

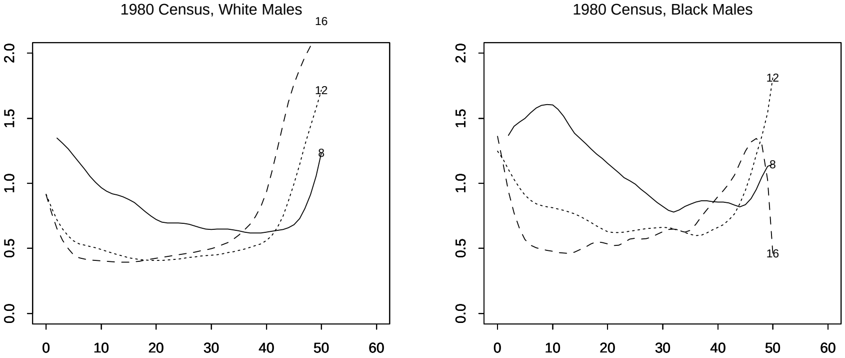
Empirical Evidence on the Mincer Model

- Figure 3 examines the support for implication (*iii*) – a U-shaped variance in earnings – for three different schooling completion levels: eighth grade, twelfth grade, and college (16 years of school).
- For the 1940 Census year, the variance of log-earnings over the life cycle is relatively flat for whites.
- It is similarly flat in 1950, with the exception of increasing variance at the tails.

Figure 3: Experience-Variance Log Earnings




1960 Census, White Males
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
12
16
8
1960 Census, Black Males
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
8
12


1980 Census, White Males
16
12
8
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
1980 Census, Black Males
12
8
16
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60

Empirical Evidence on the Mincer Model

- However, data for black and white men from the 1960-1990 Censuses clearly exhibit the U-shaped pattern predicted by Mincer's accounting-identity model.
- The evidence in support of predictions (*ii*) and (*iii*) gives analysts greater confidence in using the Mincer model to study earnings functions and rates of return to schooling, while failure of prediction (*i*) in recent decades raises a note of caution.

Empirical Evidence on the Mincer Model

- A major limitation of cross sectional analyses of variances is that they are silent about which components are predictable by the agent and which components represent true uncertainty, which is important in assessing the determinants of schooling decisions.

Empirical Evidence on the Mincer Model

- Table 2 reports standard cross-section regression estimates of the Mincer return to schooling for all Census years derived from earnings specification (1).
- The estimates indicate an *ex post* average rate of return to schooling of around 10-13% for white men and 9-15% for black men over the 1940-1990 period.
- While estimated coefficients on schooling tend to be lower for blacks than whites in the early decades, they are higher in 1980 and 1990.

Table 2: Estimated Coefficients from Mincer Log Earnings Regression for Men

		Whites		Blacks	
		Coefficient	Std. Error	Coefficient	Std. Error
1940	Intercept	4.4771	0.0096	4.6711	0.0298
	Education	0.1250	0.0007	0.0871	0.0022
	Experience	0.0904	0.0005	0.0646	0.0018
	Experience-Squared	-0.0013	0.0000	-0.0009	0.0000
1950	Intercept	5.3120	0.0132	5.0716	0.0409
	Education	0.1058	0.0009	0.0998	0.0030
	Experience	0.1074	0.0006	0.0933	0.0023
	Experience-Squared	-0.0017	0.0000	-0.0014	0.0000
1960	Intercept	5.6478	0.0066	5.4107	0.0220
	Education	0.1152	0.0005	0.1034	0.0016
	Experience	0.1156	0.0003	0.1035	0.0011
	Experience-Squared	-0.0018	0.0000	-0.0016	0.0000
1970	Intercept	5.9113	0.0045	5.8938	0.0155
	Education	0.1179	0.0003	0.1100	0.0012
	Experience	0.1323	0.0002	0.1074	0.0007
	Experience-Squared	-0.0022	0.0000	-0.0016	0.0000
1980	Intercept	6.8913	0.0030	6.4448	0.0120
	Education	0.1023	0.0002	0.1176	0.0009
	Experience	0.1255	0.0001	0.1075	0.0005
	Experience-Squared	-0.0022	0.0000	-0.0016	0.0000
1990	Intercept	6.8912	0.0034	6.3474	0.0144
	Education	0.1292	0.0002	0.1524	0.0011
	Experience	0.1301	0.0001	0.1109	0.0006
	Experience-Squared	-0.0023	0.0000	-0.0017	0.0000

Notes: Data taken from 1940-90 Decennial Censuses. See Appendix B for data description.

Empirical Evidence on the Mincer Model

- The estimates suggest that the rate of return to schooling for blacks increased substantially over the 50 year period, while it first declined and then rose for whites.
- The coefficient on experience rose for both whites and blacks over the five decades.

Empirical Evidence on the Mincer Model

- The economic content of these numbers is far from clear.
- What does a high "rate of return"—really a high growth rate of earnings with schooling—mean? The clearest interpretation is as a marginal price of schooling in the labor market and not as an internal rate of return.
- We next show how to use empirical earnings functions to estimate marginal internal rates of return.

Part IV

Estimating Internal Rates of Return

Estimating Internal Rates of Return

- Given the evidence against the validity of the Mincer earnings specification presented in below and in recent studies of the changing wage structure (e.g. Murphy and Welch, 1990; Katz and Murphy, 1992; Katz and Autor, 1999), it is fruitful to develop an alternative approach to estimating marginal internal rates of return without imposing the Mincer specification on the data.

Estimating Internal Rates of Return

- Using a simple income maximizing framework under perfect certainty of the sort developed in Rosen (1977) and Willis (1986), this section first presents estimates of the internal rate of return based on progressively more general formulations of the earnings function.
- We then relax the assumption of perfect certainty below.

Estimating Internal Rates of Return

- We initially assume that individuals choose education levels to maximize the present value of their lifetime earnings.
- They take as given a post-school earnings profile, which may be determined through on-the-job investment as in the previous accounting-identity model.
- The model estimated in this section relaxes many of the conditions of the models studied below, such as the restriction that log earnings increase linearly with schooling and the restriction that log earnings-experience profiles are parallel across schooling classes.

Estimating Internal Rates of Return

- To estimate marginal internal rates of return, which we refer to as internal rates of return in this section, analysts must account for direct costs, including both monetary and psychic costs as well as indirect costs.
- They must also account for income taxes and length of working life that may depend on the schooling level.

Estimating Internal Rates of Return

- With these additional considerations, the coefficient on schooling in a log earnings equation need no longer equal the real interest rate (the rate of return on capital), and it loses its interpretation as the internal rate of return to schooling.
- However, the internal rate of return can still be estimated using an alternative direct solution method, as we discuss below.

Estimating Internal Rates of Return

- Let $Y(s,x)$ be wage income at experience level x for schooling level s; $T(s)$, the last age of earnings, which may depend on the schooling level; v, private tuition and non-pecuniary costs of schooling; τ, a proportional income tax rate; and r, the before-tax interest rate.

- Individuals are assumed to choose s to maximize the present discounted value of lifetime earnings,

$$V(s) = \int_0^{T(s)-s} (1-\tau)e^{-(1-\tau)r(x+s)}Y(s,x)dx \qquad (2)$$
$$- \int_0^{s} ve^{-(1-\tau)rz}dz.$$

Estimating Internal Rates of Return

- The first order condition for a maximum yields

$$[T'(s) - 1]e^{-(1-\tau)r(T(s)-s)}Y(s, T(s) - s) \quad (3)$$
$$- (1-\tau)r\int_0^{T(s)-s} e^{-(1-\tau)rx}Y(s,x)dx$$
$$+ \int_0^{T(s)-s} e^{-(1-\tau)rx}\frac{\partial Y(s,x)}{\partial s}dx - v/(1-\tau) = 0.$$

Estimating Internal Rates of Return

- Defining $\tilde{r} = (1-\tau)r$ (the after-tax interest rate) and re-arranging terms yields

$$\begin{aligned}\tilde{r} \;=\;& \underbrace{\frac{[T'(s)-1]e^{-\tilde{r}(T(s)-s)}Y(s,T(s)-s)}{\int_0^{T(s)-s} e^{-\tilde{r}x}Y(s,x)dx}}_{\text{(Term 1)}} \\ &+ \underbrace{\frac{\int_0^{T(s)-s} e^{-\tilde{r}x}\left[\frac{\partial \log\ Y(s,x)}{\partial s}\right] Y(s,x)dx}{\int_0^{T(s)-s} e^{-\tilde{r}x}Y(s,x)dx}}_{\text{(Term 2)}} \\ &- \underbrace{\frac{v/(1-\tau)}{\int_0^{T(s)-s} e^{-\tilde{r}x}Y(s,x)dx}}_{\text{(Term 3)}}. \qquad (4)\end{aligned}$$

Estimating Internal Rates of Return

- Term 1 represents a life-earnings effect—the change in the present value of earnings due to a change in working-life associated with additional schooling (expressed as a fraction of the present value of earnings measured at age s).
- Term 2 is the weighted average effect of schooling on log earnings by experience, and Term 3 is the cost of tuition and psychic costs expressed as a fraction of lifetime income measured at age s.

Estimating Internal Rates of Return

- The special case assumed by Mincer and many other economists writes $v = 0$ (i.e., no tuition or psychic costs).
- The traditional assumption is that tuition costs are a small (and negligible) component of total earnings or that earnings in college offset tuition.

Estimating Internal Rates of Return

- In light of the substantial estimates of psychic costs presented in Carneiro, Hansen, and Heckman (2003); Cunha, Heckman, and Navarro (2005, 2006); and Cunha and Heckman (2006a,b), the assumption that $v = 0$ is very strong even if tuition costs are a small component of the present value of income.
- We discuss this evidence below. Accounting for psychic costs lowers the internal rate of return.

Estimating Internal Rates of Return

- Consider the additional commonly invoked assumption that $T'(s) = 1$ (i.e., no loss of work life from schooling).
- These assumptions simplify the first order condition to

$$\tilde{r} \int_0^{T(s)-s} e^{-\tilde{r}x} Y(s,x) dx = \int_0^{T(s)-s} e^{-\tilde{r}x} \frac{\partial Y(s,x)}{\partial s} dx.$$

- Mincer's model implies multiplicative separability between the schooling and experience components of earnings, so $Y(s,x) = \mu(s)\varphi(x)$ (i.e., log earnings profiles are parallel in experience across schooling levels).

Estimating Internal Rates of Return

- In this special case, $\tilde{r} = \mu'(s)/\mu(s)$.
- If this holds for all s, then wage growth must be log linear in schooling and $\mu(s) = \mu(0)e^{\rho_s s}$, where $\rho_s = \tilde{r}$.
- If all of these assumptions hold, then the coefficient on schooling in a Mincer equation (ρ_s) estimates the internal rate of return to schooling, which should equal the after-tax interest rate.

Estimating Internal Rates of Return

- From equation (4) we observe, more generally, that the difference between after-tax interest rates (and the marginal internal rate of return) and the Mincer coefficient can be decomposed into three parts: a life-earnings part (Term 1), a second part which depends on the structure of the schooling return over the life cycle, and a tuition and psychic cost part (Term 3).

Estimating Internal Rates of Return

- Term 2 is averaged over all experience levels.
- Under multiplicative separability, it is the Mincer rate of return estimated from equation (1).
- In general nonseparable models, it is not the Mincer coefficient.

Estimating Internal Rates of Return

- The evidence for 1980 and 1990 presented here and in the recent literature argues strongly against the assumption of multiplicative separability of log earnings in schooling and experience.
- In recent decades, cross section log earnings-experience profiles are not parallel across schooling groups.
- In addition, college tuition costs are nontrivial and are not offset by work in school for most college students.
- These factors account for some of the observed disparities between the after-tax interest rate and the steady-state Mincer coefficient.

Estimating Internal Rates of Return

- One can view $\tilde{r}$ as a marginal internal rate of return to schooling after incorporating tuition costs, earnings increases, and changes in the retirement age.
- That is, $\tilde{r}$ is the discount rate that equates the net lifetime earnings for marginally different schooling levels at an optimum.
- As in the model of Mincer (1958), this internal rate of return should equal the interest rate in a world with perfect credit markets, once all costs and benefits from schooling are considered.

Estimating Internal Rates of Return

- After allowing for taxes, tuition, variable length of working life, and a flexible relationship between earnings, schooling and experience, the coefficient on years of schooling in a log earnings regression need no longer equal the internal rate of return.
- However, it is still possible to calculate the internal rate of return using the observation that it is the discount rate that equates lifetime earnings streams for two different schooling levels.
- Typically, internal rates of return are based on non-marginal differences in schooling.

Estimating Internal Rates of Return

- Incorporating tuition (and psychic costs) and taxes, the internal rate of return for schooling level s_1 versus s_2, $r_I(s_1, s_2)$, solves (suppressing the argument of $r_I(s_1, s_2)$)

$$\int_0^{T(s_1)-s_1} (1-\tau)e^{-r_I} Y(s_1, x)dx - \int_0^{s_1} ve^{-r_I z}dz \tag{5}$$
$$= \int_0^{T(s_2)-s_2} (1-\tau)e^{-r_I} Y(s_2, x)dx - \int_0^{s_2} ve^{-r_I z}dz.$$

Estimating Internal Rates of Return

- As with $\tilde{r}$ above, r_I will equal the Mincer coefficient on schooling under the assumptions of parallelism in experience across schooling categories (i.e., $Y(s,x) = \mu(s)\varphi(x)$), linearity of log earnings in schooling ($\mu(s) = \mu(0)e^{\rho_s s}$), no tuition and psychic costs ($v = 0$), no taxes ($\tau = 0$), and equal work-lives irrespective of years of schooling ($T'(s) = 1$).
- In the next section, we compare rate of return estimates based on specification (1) to those obtained by directly solving for r_I in equation (5).

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Using data for white and black men from 1940-1990 decennial Censuses, we examine how estimates of the internal rate of return change when different assumptions about the model are relaxed.
- Tables 3a and 3b report internal rates of return to schooling for each Census year and for a variety of pairwise schooling level comparisons for white and black men, respectively.

Table 3a: Internal Rates of Return for White Men: Earnings Function Assumptions
(Specifications Assume Work Lives of 47 Years)

	Schooling Comparisons					
	6-8	8-10	10-12	12-14	12-16	14-16
1940						
Mincer Specification	13	13	13	13	13	13
Relax Linearity in S	16	14	15	10	15	21
Relax Linearity in S & Quad. in Exp.	16	14	17	10	15	20
Relax Lin. in S & Parallelism	12	14	24	11	18	26
1950						
Mincer Specification	11	11	11	11	11	11
Relax Linearity in S	13	13	18	0	8	16
Relax Linearity in S & Quad. in Exp.	14	12	16	3	8	14
Relax Linearity in S & Parallelism	26	28	28	3	8	19
1960						
Mincer Specification	12	12	12	12	12	12
Relax Linearity in S	9	7	22	6	13	21
Relax Linearity in S & Quad. in Exp.	10	9	17	8	12	17
Relax Linearity in S & Parallelism	23	29	33	7	13	25
1970						
Mincer Specification	13	13	13	13	13	13
Relax Linearity in S	2	3	30	6	13	20
Relax Linearity in S & Quad. in Exp.	5	7	20	10	13	17
Relax Linearity in S & Parallelism	17	29	33	7	13	24
1980						
Mincer Specification	11	11	11	11	11	11
Relax Linearity in S	3	-11	36	5	11	18
Relax Linearity in S & Quad. in Exp.	4	-4	28	6	11	16
Relax Linearity in S & Parallelism	16	66	45	5	11	21
1990						
Mincer Specification	14	14	14	14	14	14
Relax Linearity in S	-7	-7	39	7	15	24
Relax Linearity in S & Quad. in Exp.	-3	-3	30	10	15	20
Relax Linearity in S & Parallelism	20	20	50	10	16	26

Notes: Data taken from 1940-90 Decennial Censuses. In 1990, comparisons of 6 vs. 8 and 8 vs. 10 cannot be made given data restrictions. Therefore, those columns report calculations based on a comparison of 6 and 10 years of schooling. See Appendix B for data description.

Table 3b: Internal Rates of Return for Black Men: Earnings Function Assumptions
(Specifications Assume Work Lives of 47 Years)

	Schooling Comparisons					
	6-8	8-10	10-12	12-14	12-16	14-16
1940						
Mincer Specification	9	9	9	9	9	9
Relax Linearity in S	18	7	5	3	11	18
Relax Linearity in S & Quad. in Exp.	18	8	6	2	10	19
Relax Linearity in S & Parallelism	11	0	10	5	12	20
1950						
Mincer Specification	10	10	10	10	10	10
Relax Linearity in S	16	14	18	-2	4	9
Relax Linearity in S & Quad. in Exp.	16	14	18	0	3	6
Relax Linearity in S & Parallelism	35	15	48	-3	6	34
1960						
Mincer Specification	11	11	11	11	11	11
Relax Linearity in S	13	12	18	5	8	11
Relax Linearity in S & Quad. in Exp.	13	11	18	5	7	10
Relax Linearity in S & Parallelism	22	15	38	5	11	25
1970						
Mincer Specification	12	12	12	12	12	12
Relax Linearity in S	5	11	30	7	10	14
Relax Linearity in S & Quad. in Exp.	6	11	24	10	11	12
Relax Linearity in S & Parallelism	15	27	44	9	14	23
1980						
Mincer Specification	12	12	12	12	12	12
Relax Linearity in S	-4	1	35	10	15	19
Relax Linearity in S & Quad. in Exp.	-4	6	29	11	14	17
Relax Linearity in S & Parallelism	10	44	48	8	16	31
1990						
Mincer Specification	16	16	16	16	16	16
Relax Linearity in S	-5	-5	41	15	20	25
Relax Linearity in S & Quad. in Exp.	-3	-3	35	17	19	22
Relax Linearity in S & Parallelism	16	16	58	18	25	35

Notes: Data taken from 1940-90 Decennial Censuses. In 1990, comparisons of 6 vs. 8 and 8 vs. 10 cannot be made given data restrictions. Therefore, those columns report calculations based on a comparison of 6 and 10 years of schooling. See Appendix B for data description.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- These estimates assume that workers spend 47 years working irrespective of their educational choice (i.e., a high school graduate works until age 65 and a college graduate until 69).
- To calculate each of the IRR estimates, we first estimate a log wage equation under the assumptions indicated in the tables.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Then, we predict earnings under this specification for the first 47 years of experience, and the IRR is taken to be the root of equation (5).
- As a benchmark, the first row for each year reports the IRR estimate obtained from the Mincer specification for log wages (equation (1)).
- The IRR could equivalently be obtained from a Mincer regression coefficient.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Relative to the Mincer specification, row 2 relaxes the assumption of linearity in schooling by including indicator variables for each year of schooling.
- This modification alone leads to substantial differences in the estimated rate of return to schooling, especially for schooling levels associated with degree completion years (12 and 16) which have much larger returns than other schooling years.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- For example, the IRR to finishing high school is 30% for white men in 1970, while the rate of return to finishing 10 rather than 8 years of school is only 3%.
- In general, imposing linearity in schooling leads to upward biased estimates of the rate of return to grades that do not produce a degree, while it leads to downward biased estimates of the degree completion years (high school or college).

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Sheepskin effects are an important feature of the data.
- There is a considerable body of evidence against linearity of log earnings in schooling.
- See, e.g. Heckman, Layne-Farrar, and Todd, 1996, Jaeger and Page, 1996, Hungerford and Solon, 1987.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Row 3 relaxes both linearity in schooling and the quadratic specification for experience, which produces similar estimates.
- The assumption that earnings are quadratic in experience is empirically innocuous for estimating returns to schooling once linearity in years of schooling is relaxed.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Finally, row 4 relaxes all three Mincer functional form assumptions.
- Earnings functions are nonparametrically estimated as a function of experience, separately within each schooling class as shown in Figure 1.
- This procedure does not impose any assumption other than continuity on the earnings-experience relationship.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Comparing these results with those of row three provides a measure of the bias induced by assuming separability of earnings in schooling and experience.
- In many cases, especially in recent decades, there are large differences.
- This finding is consistent with the results we reported, which show that earnings profiles in recent decades are no longer parallel in experience across schooling categories.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- The general estimates in Tables 3a and b show a large increase in the return to completing high school for whites (Table 3a), which goes from 24% in 1940 to 50% in 1990, and even more dramatic increases for blacks (Table 3b).
- The estimates for 1990 seem implausible but are the rates of return that are implicit in recent Census- and CPS-based estimates.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- It is possible that these increases in rates of return over time partially reflect a selection effect, stemming from a decrease in the average quality of workers over time who drop out of high school.
- Given the limitations of Census and CPS data, we do not correct for censoring or selection bias in our analysis of these data.
- Below, we consider estimation when schooling choices are endogenous.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Since 1950, there has been a sizeable increase over time in the marginal internal rate of return to attending and completing college, consistent with changes in demand favoring highly skilled workers.
- For most grade comparisons and years, the Mincer coefficient implies a lower return to schooling than do the nonparametric estimates, with an especially large disparity for the return to high school completion.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- For whites, the return to a 4-year college degree is similar under the Mincer and nonparametric models, but for blacks the Mincer coefficient substantially understates the return in recent decades.
- While the recent literature has focused on rising returns to college relative to high school, the increase in returns to completing high school appears to have been substantially greater.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- A comparison of the IRR estimates based on the most flexible model for black males and white males shows that for all years except 1940, the return to high school completion is higher for black males, reaching a peak of 58% in 1990 (compared with 50% for whites in 1990).
- The internal rate of return to completing 16 years is also higher for blacks in most years (by about 10% in 1990).

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Estimated internal rates of return differ depending on the set of assumptions imposed by the earnings model.
- Murphy and Welch (1990) note that allowing for quartic terms in experience is empirically important for fitting the earnings equation (the hedonic pricing equation), but do not report any effects of relaxing the quadratic-in-experience assumption on estimated marginal rates of return to schooling.
- We find that imposing the quadratic-in-experience assumption is fairly innocuous for computing rates of return.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- The assumptions of linearity in schooling and separability in schooling and experience are not.
- Comparing the unrestricted estimates in row 4 with the Mincer-based estimates in row 1 reveals substantial differences for nearly all grade progressions and all years.
- If imposing linearity and separability is innocuous, relaxing these conditions should not have such a dramatic effect on estimates of rates of return.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Table 4 examines how the IRR estimates for post-secondary education change when we account for income taxes (both flat and progressive) and college tuition.
- Below, we discuss the relevance of psychic costs.
- For ease of comparison, the first row for each year reports estimates of the IRR for the most flexible earnings specification, not accounting for tuition and taxes.

Table 4: Internal Rates of Return for White & Black Men: Accounting for Taxes and Tuition
(General Non-Parametric Specification Assuming Work Lives of 47 Years)

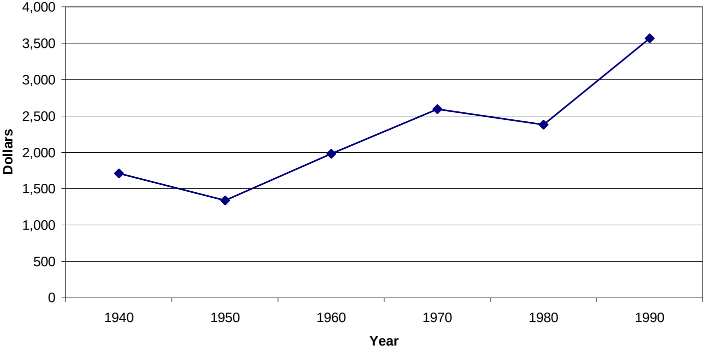
		Schooling Comparisons					
		Whites			Blacks		
		12-14	12-16	14-16	12-14	12-16	14-16
1940	No Taxes or Tuition	11	18	26	5	12	20
	Including Tuition Costs	9	15	21	4	10	16
	Including Tuition & Flat Taxes	8	15	21	4	9	16
	Including Tuition & Prog. Taxes	8	15	21	4	10	16
1950	No Taxes or Tuition	3	8	19	-3	6	34
	Including Tuition Costs	3	8	16	-3	5	25
	Including Tuition & Flat Taxes	3	8	16	-3	5	24
	Including Tuition & Prog. Taxes	3	7	15	-3	5	21
1960	No Taxes or Tuition	7	13	25	5	11	25
	Including Tuition Costs	6	11	21	5	9	18
	Including Tuition & Flat Taxes	6	11	20	4	8	17
	Including Tuition & Prog. Taxes	6	10	19	4	8	15
1970	No Taxes or Tuition	7	13	24	9	14	23
	Including Tuition Costs	6	12	20	7	12	18
	Including Tuition & Flat Taxes	6	11	20	7	11	17
	Including Tuition & Prog. Taxes	5	10	18	7	10	16
1980	No Taxes or Tuition	5	11	21	8	16	31
	Including Tuition Costs	4	10	18	7	13	24
	Including Tuition & Flat Taxes	4	9	17	6	12	21
	Including Tuition & Prog. Taxes	4	8	15	6	11	20
1990	No Taxes or Tuition	10	16	26	18	25	35
	Including Tuition Costs	9	14	20	14	18	25
	Including Tuition & Flat Taxes	8	13	19	13	17	22
	Including Tuition & Prog. Taxes	8	12	18	13	17	22

Notes: Data taken from 1940-90 Decennial Censuses. See discussion in text and Appendix B for a description of tuition and tax amounts.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- These estimates are identical to the fourth row in Tables 3a and 3b.
- All other rows account for private tuition costs for college (v) assumed equal to the average college tuition paid in the U.S. that year.
- The average college tuition paid by students increased steadily since 1950 as shown in Figure 4a.
- In 1990, it stood at roughly $3,500 (in 2000 dollars).

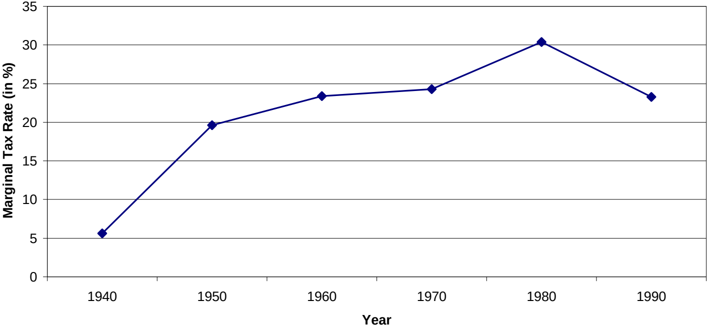
Figure 4a: Average College Tuition Paid (in 2000 dollars)



How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Row three of Table 4 accounts for flat wage taxes using estimates of average marginal tax rates (τ) from Barro and Sahasakul (1983) and Mulligan and Marion (2000), which are plotted for each of the years in Figure 4b.
- Average marginal tax rates increased from a low of 5.6% in 1940 to a high of 30.4% in 1980 before falling to 23.3% in 1990.

Figure 4b: Marginal Tax Rates
(from Barro & Sahasakul, 1983, Mulligan & Marion, 2000)



How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- The final row of Table 4 accounts for the progressive nature of our tax system using federal income tax schedules (Form 1040) for single adults with no dependents and no unearned income.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- When costs of schooling alone are taken into account (comparing row 2 with row 1), the return to college generally falls by a few percentage points.
- Because the earnings of blacks are typically lower than for whites but tuition payments are assumed here to be the same, accounting for tuition costs has a bigger effect on the estimates for the black samples.
- For example, internal rates of return to the final two years of college decline by about one-fourth for whites and one-third for blacks.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Further accounting for taxes on earnings (rows 3 and 4) has little additional impact on the estimates.
- Interestingly, the progressive nature of the tax system typically reduces rates of return by less than a percentage point.

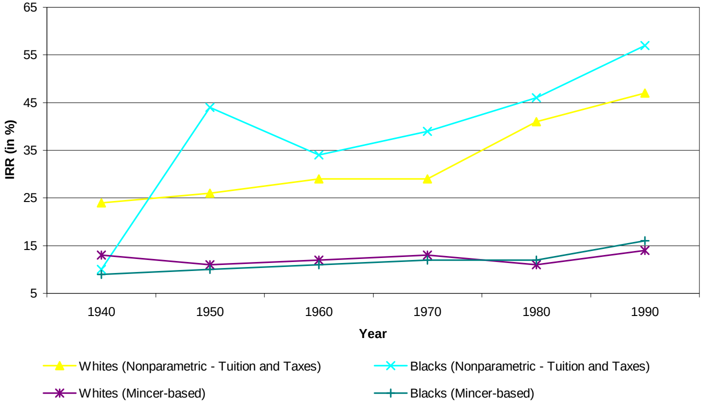
How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Overall, failure to account for tuition and taxes leads to an overstatement of the return to college, but the time trends in the return are fairly similar whether or not one adjusts for taxes and tuition.
- However, accounting for psychic costs has a substantial effect on estimated rates of return.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Figure 5 graphs the time trend in the IRR to high school completion for white and black males, comparing estimates based on (i) the Mincer model and (ii) the flexible nonparametric earnings model accounting for progressive taxes and tuition.
- Estimates based on the Mincer specification tend to understate returns to high school completion and also fail to capture the substantial rise in returns to schooling that has taken place since 1970.
- Furthermore, the sizeable disparity in returns by race is not captured by the estimates based on the Mincer equation.

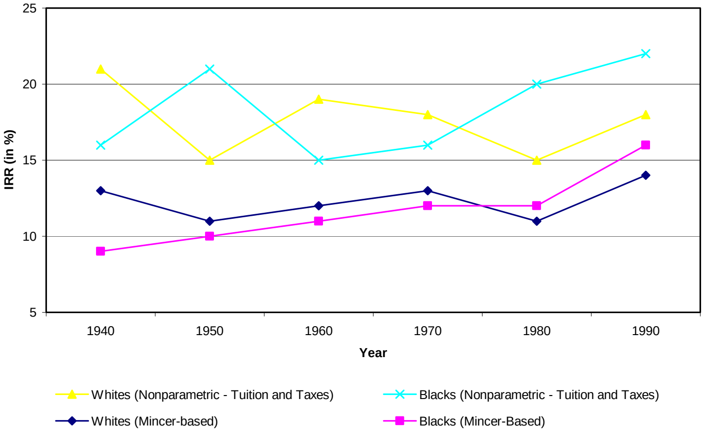
Figure 5: IRR for High School Completion (White and Black Men)



How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Figure 6 presents similar estimates for college completion (14 vs. 16 years of school).
- Again, the Mincer model yields much lower estimates of the IRR in comparison with the more flexible model that also takes into account taxes and tuition.

Figure 6: IRR for College Completion (White and Black Men)



How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Nonparametric estimates of the return to college completion are generally 5-10% higher than the corresponding Mincer-based estimates even after accounting for taxes and tuition.
- Additionally, the more general specification reveals a substantial decline in the IRR to college between 1950 and 1960 for blacks that is not reflected in the Mincer-based estimates.

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- Using our flexible earnings specification, we also examine how estimates depend on assumptions about the length of working life, comparing two extreme cases.
- The estimates just reported assume that individuals work for 47 years regardless of their schooling (i.e., $T'(s) = 1$).
- An alternative assumption posits that workers retire at age 65 regardless of their education (i.e., $T'(s) = 0$).

How alternative specifications of the Mincer equation and accounting for taxes and tuition affect estimates of the internal rate of return (IRR)

- We find virtually identical results for all years and schooling comparisons for both assumptions about the schooling – worklife relationship.
- Because earnings at the end of the life cycle are heavily discounted, they have little impact on the total value of lifetime earnings and, therefore, have little effect on internal rate of return estimates.

Accounting for Uncertainty in a Static Version of the Model

- To this point, we have computed internal rates of return using fitted values from estimated earnings equations.
- Mincer's approach and more general nonparametric approaches pursued in the literature make implicit assumptions about how individuals forecast their future earnings.
- The original formulations ignore uncertainty, making no distinction between *ex post* and *ex ante* returns.
- It is essential to know *ex ante* returns in order to understand schooling choices, because they are the returns on which individuals act.

Accounting for Uncertainty in a Static Version of the Model

- In this subsection, we explore alternative approaches for estimating the IRR used by agents in making their schooling choices that are based on alternative assumptions about expectation formation mechanisms.
- These analyses are based on cross section data.
- We present a more general dynamic analysis in the next section.

Accounting for Uncertainty in a Static Version of the Model

- As previously discussed, it is common in the literature to use log specifications for earnings.
- Thus, using a general notation, it is common to assume $\ln Y = Z\gamma + \varepsilon$, so $Y = e^{Z\gamma}e^{\varepsilon}$ and that expected earnings given Z are

$$E(Y|Z) = e^{Z\gamma}E(e^{\varepsilon}).$$

Accounting for Uncertainty in a Static Version of the Model

- Assume for the sake of argument (but contrary to the evidence) that equation (1) describes the true earnings process and that $E(\varepsilon|x, s) = 0$.
- To this point, when we have fit Mincer equations, we have estimated internal rates of return using fitted values for Y in place of the true values.
- That is, we use the following estimate for earnings: $\hat{Y}(s, x) = \exp(\hat{\alpha}_0 + \hat{\rho}_s s + \hat{\beta}_0 x + \hat{\beta}_1 x^2)$, where $\hat{\alpha}_0$, $\hat{\rho}_s$, $\hat{\beta}_0$, and $\hat{\beta}_1$ are the regression estimates.

Accounting for Uncertainty in a Static Version of the Model

- This procedure implicitly assumes that individuals place themselves at the mean of the log earnings distribution when forecasting their earnings and making their schooling choices.
- Individuals take fitted log earnings profiles as predictions for their own future earnings, ignoring any potential person-specific deviations from that profile.

Accounting for Uncertainty in a Static Version of the Model

- Ignoring taxes, for this case, the IRR estimator $\hat{r}_I$ solves

$$\sum_{x=0}^{T} \frac{\hat{Y}(s+j,x)}{(1+\hat{r}_I)^{s+j+x}} - \sum_{x=0}^{T} \frac{\hat{Y}(s,x)}{(1+\hat{r}_I)^{s+x}} - v \sum_{x=1}^{j} \frac{1}{(1+\hat{r}_I)^{s+x}} = 0,$$

  which is the discrete time analogue to the model of equation (2) for two schooling levels s and $s+j$.

- If tuition and psychic costs are negligible ($v = 0$),

$$\text{plim } \hat{r}_I = e^{\rho_s} - 1 \approx \rho_s.$$

Accounting for Uncertainty in a Static Version of the Model

- Given our assumptions on expectations, this is an *ex ante* rate of return.
- *Ex ante* returns are the theoretically appropriate ones for studying schooling behavior, because they are the returns on which schooling decisions are based.

Accounting for Uncertainty in a Static Version of the Model

- Suppose instead that agents base their expectations of future earnings at different schooling levels on the mean earnings profiles for each schooling level, or on $E(Y|s,x)$.
- In this case, the estimator of the *ex ante* rate of return is given by the root of

$$\sum_{x=0}^{T} \frac{E(Y(s+j,x)|s,x)}{(1+\hat{r}_I)^{s+j+x}} - \sum_{x=0}^{T} \frac{E(Y(s,x)|s,x)}{(1+\hat{r}_I)^{s+x}} - \sum_{x=1}^{j} \frac{v}{(1+\hat{r}_I)^{s+x}} = 0. \qquad (6)$$

Accounting for Uncertainty in a Static Version of the Model

- If $v = 0$ and Mincer's assumptions hold, this formula specializes to

$$\frac{e^{\rho_s j}}{(1+\hat{r}_I)^j} \sum_{x=0}^{T} \frac{e^{\beta_0 x + \beta_1 x^2} E(e^{\varepsilon(s+j,x)} | s, x)}{(1+\hat{r}_I)^x}$$
$$= \sum_{x=0}^{T} \frac{e^{\beta_0 x + \beta_1 x^2} E(e^{\varepsilon(s,x)} | s, x)}{(1+\hat{r}_I)^x}.$$

- If $E[e^{\varepsilon(s,x)} | s, x] = E[e^{\varepsilon(s+j,x)} | s, x]$ for all x, then the two sums are equal and $\text{plim } \hat{r}_I = e^{\rho_s} - 1$ as before.

Accounting for Uncertainty in a Static Version of the Model

- In this special case, using $\hat{Y}(s,x) = exp(\hat{\alpha}_0 + \hat{\rho}_s s + \hat{\beta}_0 x + \hat{\beta}_1 x^2)$ or $E(Y(s,x)|s,x)$ will yield estimates of the internal rate of return that are asymptotically equivalent.
- However, if $E(e^{\varepsilon(s+j,x)}|s,x)$ is a more general function of s and x, then the estimators of the *ex ante* return will differ.

Accounting for Uncertainty in a Static Version of the Model

- In the more general case, using estimates of $E(Y(s,x)|s,x)$ under a Mincer specification yields an estimated rate of return with a probability limit

$$\text{plim } \hat{r}_I = e^{\rho_s}\left[M(s,j)\right]^{1/j} - 1 \approx \rho_s + \frac{1}{j}(\ln M(s,j)),$$

where

$$M(s,j) = \frac{\sum_{x=0}^{T} e^{\beta_0 x + \beta_1 x^2} E(e^{\varepsilon(s+j,x)}|s,x)(1+r_I)^{-x}}{\sum_{x=0}^{T} e^{\beta_0 x + \beta_1 x^2} E(e^{\varepsilon(s,x)}|s,x)(1+r_I)^{-x}}. \tag{7}$$

Accounting for Uncertainty in a Static Version of the Model

- This estimator of the *ex ante* internal rate of return will be larger than ρ_s if the variability in earnings is greater for more educated workers (i.e., $M(s,j) > 1$) and smaller if the variability is greater for less educated workers (i.e., $M(s,j) < 1$).
- If individuals use mean earnings at given schooling levels in forming expectations, then this estimator is more appropriate.
- However, this approach equates all variability across people with uncertainty, even though some aspects of variability across persons are predictable.

Accounting for Uncertainty in a Static Version of the Model

- We discuss how to decompose variability into predictable and unpredictable components below.
- Inspection of Figure 3 reveals that, at young ages, the variability in earnings for low education groups is the highest among all groups.
- If discounting dominates wage growth with experience, we would expect that $M(s, j) < 1$.

Accounting for Uncertainty in a Static Version of the Model

- These calculations assume that agents are forecasting the unknown $\varepsilon(s, x)$ using (s, x).
- If they also use another set of variables q, then the rate of return should be defined conditional on q $(\hat{r}_I = \hat{r}_I(q))$ and we would have to average over q to obtain the average *ex ante* rate of return.
- If agents know $\varepsilon(s, x)$ at the time they make their schooling decisions, then the *ex ante* return and the *ex post* return are the same, and $\hat{r}_I$ now depends on the full vector of "shocks" confronting agents.

Accounting for Uncertainty in a Static Version of the Model

- Returns would then be averaged over the distribution of all "shocks" to calculate an expected return.
- Due to the nonlinearity of the equation used to calculate the internal rate of return, the rate of return based on an average earnings profile is not the same as the mean rate of return.
- Thus, mean *ex ante* and mean *ex post* internal rates of return are not the same.

Accounting for Uncertainty in a Static Version of the Model

- When ρ_s varies in the population, these results must be further modified.
- Assume that ρ_s varies across individuals, that $E(\rho_s) = \bar{\rho}_s$, and that ρ_s is independent of x and $\varepsilon(s+j, x)$ for all x, j.
- Also, assume $v = 0$ for expositional purposes (no tuition or psychic costs).

Accounting for Uncertainty in a Static Version of the Model

- Using fitted earnings, $\hat{w}(s, x)$, to calculate internal rates of return yields an estimator, $\hat{r}_I$, that satisfies

$$\text{plim } \hat{r}_I = e^{\bar{\rho}_s} - 1 \approx \bar{\rho}_s.$$

- This estimator calculates the *ex ante* internal rate of return for someone with the mean increase in annual log earnings $\rho_s = \bar{\rho}_s$ and with the mean deviation from the overall average $\varepsilon(s, x) = \varepsilon(s + j, x) = 0$ for all x.

Accounting for Uncertainty in a Static Version of the Model

- On the other hand, assuming agents cannot forecast ρ_s, using estimates of mean earnings $E(Y(s,x)|s,x)$ will yield an estimator for r with

$$\text{plim } \hat{r}_I = e^{\bar{\rho}_s}\left[kM(s,j)\right]^{1/j} - 1 \approx \bar{\rho}_s + \frac{1}{j}[\ln k + \ln M(s,j)],$$

where $k = \frac{E(e^{(s+j)(\rho_s - \bar{\rho}_s)}|s,x)}{E(e^{s(\rho_s - \bar{\rho}_s)}|s,x)}$ and $M(s,j)$ is defined in equation (7).

Accounting for Uncertainty in a Static Version of the Model

- For $\bar{\rho}_s > 0$, it is straightforward to show that $k > 1$, which implies that everything else the same, the estimator, $\hat{r}_I$, based on mean earnings will be larger when there is variation in the return to schooling than when there is not.
- Furthermore, the internal rate of return is larger for someone with the mean earnings profile than it is for an individual with the mean value of ρ_s.

Accounting for Uncertainty in a Static Version of the Model

- Again, if agents know ρ_s, we should compute $\hat{r}_I$ conditioning on ρ_s and construct the mean rate of return from the average of those $\hat{r}_I$.
- Again, the mean *ex post* and *ex ante* rates of return are certain to differ unless agents have perfect foresight.

Accounting for Uncertainty in a Static Version of the Model

- Table 5 reports estimates of the *ex ante* IRR based on our general nonparametric specification.
- We compute the IRR under two alternative assumptions: (i) that agents forecast future earnings using the earnings function that sets $\varepsilon = 0$ ("unadjusted earnings") and (ii) that agents forecast using mean earnings within each education and experience category rather than using predicted earnings placing themselves at $\varepsilon = 0$ ("adjusted earnings").

Table 5: Internal Rates of Return for White & Black Men: Residual Adjustment
(General Non-Parametric Specification Accounting for Tuition and Progressive Taxes)

			Schooling Comparisons					
			6-8	8-10	10-12	12-14	12-16	14-16
a. Whites								
	1940	Unadjusted	12	14	24	8	15	21
		Adjusted	2	2	8	9	13	16
	1950	Unadjusted	25	26	26	3	7	15
		Adjusted	17	19	14	5	8	14
	1960	Unadjusted	21	27	29	6	10	19
		Adjusted	13	19	16	7	11	16
	1970	Unadjusted	16	27	29	5	10	18
		Adjusted	11	18	16	6	10	16
	1980	Unadjusted	14	64	41	4	8	15
		Adjusted	9	28	24	5	8	13
	1990	Unadjusted	19	19	47	8	12	18
		Adjusted	11	11	31	8	12	17
b. Blacks								
	1940	Unadjusted	11	0	10	4	10	16
		Adjusted	3	0	-8	4	6	7
	1950	Unadjusted	33	14	44	-3	5	21
		Adjusted	53	8	21	1	9	15
	1960	Unadjusted	20	14	34	4	8	15
		Adjusted	14	12	16	6	6	8
	1970	Unadjusted	14	25	39	7	10	16
		Adjusted	12	16	22	7	10	12
	1980	Unadjusted	9	43	46	6	11	20
		Adjusted	7	21	29	6	9	15
	1990	Unadjusted	16	16	57	13	17	22
		Adjusted	8	8	42	11	15	20

Notes: Data taken from 1940-90 Decennial Censuses. In 1990, comparisons of 6 vs. 8 and 8 vs. 10 cannot be made given data restrictions. Therefore, those columns report calculations based on a comparison of 6 and 10 years of schooling. See discussion in text and Appendix B for a description of tuition and tax amounts. Unadjusted sets the residual from the earnings equation to be the same for everyone (= 0). Adjusted uses mean earnings within each age-schooling cell.

Accounting for Uncertainty in a Static Version of the Model

- Procedure (ii) is described in equation (6).
- Procedure (i) sets $E\left(e^{\varepsilon(s,x)} \mid s, x\right) = 1$ for all s, x.
- Both the adjusted and unadjusted estimates account for tuition and progressive taxes.
- The adjusted estimates generate much lower (and more reasonable) IRR estimates than the unadjusted ones.

Accounting for Uncertainty in a Static Version of the Model

- Using mean earnings rather than earnings for someone with the mean residual generally leads to lower estimated *ex ante* internal rates of return for most schooling comparisons.
- Even if the Mincer specification for log earnings is correct, the internal rate of return guiding individual decisions is lower than the Mincer estimated rate of return when individuals base their schooling decisions on average earnings levels within schooling and experience categories.

Accounting for Uncertainty in a Static Version of the Model

- In other words, predicted earnings obtained using the coefficients from a log earnings regression evaluated where $\varepsilon = 0$ is an inaccurate measure of the average earnings within each schooling and experience category.

Accounting for Uncertainty in a Static Version of the Model

- The adjustment for uncertainty reported in this section based on mean earnings makes the strong assumption that all variation is unforecastable at the time schooling decisions are made.
- A better approach is to extract components of variation that are forecastable at the time schooling decisions are being made (heterogeneity) from components that are unforecastable (true uncertainty).

Accounting for Uncertainty in a Static Version of the Model

- Only the latter components should be used to compute $M(s, j)$.
- Methods for separating forecastable heterogeneity from uncertainty are available (Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2005; Heckman and Navarro, 2006) but require panel data and cannot be applied to Census cross-sections. We review the evidence from the panel literature below.

Accounting for Uncertainty in a Static Version of the Model

- Another major issue about the entire enterprise of calculating rates of return is whether the marginal rate of return is an economically interesting concept when agents are sequentially revising their information about returns to schooling.
- As shown in the next section, in general it is not.
- This casts doubt on the policy relevance of the entire rate of return literature, that was initially motivated by Becker (1964), and suggests that the literature should be refocused to account for intrinsic uncertainty.

Part V

The Internal Rate of Return and The Sequential Resolution of Uncertainty

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Human capital theory was developed in an era before the tools of dynamic decision making under uncertainty were fully developed.
- Concepts central to human capital theory like the internal rate of return are not generally appropriate to the evaluation of investment programs under sequential resolution of uncertainty.
- The recent literature has made progress towards empirical analysis of schooling decisions in dynamic settings.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Our analysis of this issue in this section is mainly theoretical and aimed at clarifying a number of important features of dynamic schooling decisions under uncertainty.
- We discuss other dynamic models with option values developed in the recent dynamic literature below.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- This section makes three main points.
- First, ignoring the sequential revelation of information, Mincer's assumption of the linearity of log earnings in years of schooling rules out option values that can arise even in an environment where the agent perfectly anticipates future earnings.
- We show how nonlinearity is a source of option values, and accounting for option values affects estimated returns to schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Second, sequential revelation of information is an additional source of option values.
- Accounting for information updating is a force toward generating a *downward* bias in least squares estimates of returns to schooling.
- Intuitively, people drop out of school when they have good draws, leaving only the unlucky to continue on in their schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- This result runs counter to the intuition often expressed in the conventional ability bias literature that the most able continue on to school.
- For a survey of the conventional literature see, e.g., Griliches, 1977.
- Third, we show that the internal rate of return is not a correct investment criterion when earnings are uncertain and there are option values.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- For two reasons, the dynamic nature of schooling suggests that the returns to education may include an option value.
- First, the return to one year of school may include the potential for larger returns associated with higher levels of education when the returns to school are not constant across all schooling levels.
- For example, finishing high school provides access to college, and attending college is a necessary first step for obtaining a college degree.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Given the large increase in earnings associated with college completion, the total return to high school or college attendance includes the potential for even greater returns associated with finishing college.
- The return in excess of the direct return (the lifetime income received at a given schooling level) is the option value.
- Mincer's assumption that earnings are log linear in schooling implicitly rules out this type of option value if the growth rate in earnings is the same as the interest rate.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The traditional approach to schooling computes the rate of return using the lifetime income arising from stopping at schooling level s with the lifetime income from stopping at $s+1$ using the direct return, i.e., the return of stopping at s versus the return from stopping at $s+1$, and does not consider the continuation value.
- Second, when there is uncertainty about college costs or future earnings and when each additional year of schooling reveals new information about those costs or earnings, the full returns to schooling will include the expected value of newly revealed information that can be acted on.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Finishing high school opens the possibility of attending college, which will be realized if tuition costs and opportunity costs turn out to be low.
- Therefore, the returns to high school completion include both the increase in earnings associated with completing high school and the *ex ante* expected value of continuing beyond high school, including the expected value of all future information, including information about wage shocks, costs of additional schooling, ability in various tasks and the like.
- The value of this information depends on the probability that the individual decides to continue on to college and the expected return if he does so.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Failing to finish high school precludes an individual from learning the information that arises from high school completion as well as the value of exercising the option to go to college.
- Dropping out eliminates the college option.
- Earnings each period may also be uncertain.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The decision to continue on in school will depend on both current and expected future labor market conditions.
- By ignoring uncertainty, the literature based on the Mincer earnings equation neglects this source of option values.
- Sequential arrival of information implies that education decisions are made sequentially and should not be treated as a static discrete choice problem made once in a lifetime by individuals—the traditional approach used in human capital theory (see, e.g. Mincer, 1958; Willis and Rosen, 1979; Willis, 1986; Card, 2001).

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The empirical evidence we present (see also Bound, Jaeger, and Baker, 1995; Heckman, Layne-Farrar, and Todd, 1996; Hungerford and Solon 1987) strongly rejects Mincer's (1958) implicit assumption that marginal internal rates of return to each year of schooling are identical and equal to a common interest rate, i.e., the assumption that log earnings are linear in years of schooling.
- This observation alone undermines the interpretation of the coefficient on schooling in a log earnings regression as a rate of return.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- But this non-linearity, combined with the sequential resolution of uncertainty, creates additional problems for estimating rates of returns using Mincer regressions.
- Because the returns to college completion are high, it may be worthwhile to finish high school to keep open the option of attending college.
- The total return to high school and earlier schooling choices includes a non-trivial option value.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- To analyze this option value, we present two simple dynamic models with uncertainty about the value of future schooling choices.
- Following most of the literature, we assume that individuals maximize the expected value of lifetime earnings given their current education level and the available information.
- We briefly discuss more general dynamic models with option values below.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- To gain some understanding about the separate roles of nonlinearity and uncertainty in generating option values, we first consider the option value framework of Comay, Melnik, and Pollatschek (1973), which assumes that there is no uncertainty about earnings conditional on final schooling attainment but that individuals face an exogenously specified probability ($\pi_{s+1,s}$) of being accepted into grade $s+1$ if they choose to apply after finishing grade s.
- Thus they face a lottery where the chance of being admitted to the next round of schooling does not depend on earnings.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- For someone attending exactly s years of school, define the discounted present value of lifetime earnings as of the schooling completion date as:

$$Y_s = \sum_{x=0}^{T} (1+r)^{-x} Y(s,x),$$

  where the interest rate, r, is assumed to be exogenously specified and common across persons.

- This expression is assumed to be known with certainty.
- If an individual who chooses to apply for grade $s+1$ is rejected, he or she begins working immediately, earning Y_s.
- This is the direct value of schooling as conventionally measured.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- In this environment, the total expected value of attaining $s \in \{1, 2, \dots, \bar{S}\}$ years of school, given the information available at the end of stage $s - 1$, is

$$E_{s-1}(V_s) = (1 - \pi_{s+1,s}) Y_s + \pi_{s+1,s} E_{s-1} \max \left\{ Y_s, \frac{E_s(V_{s+1})}{1+r} \right\}$$

  for $s < \bar{S}$ and $E_{\bar{S}-1}(V_{\bar{S}}) = Y_{\bar{S}}$.

- This expression assumes that each grade of school takes one period and that direct costs of schooling are negligible.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The *ex ante* option value of grade s as perceived at the end of $s-1$ is defined as the difference between the total expected value of that opportunity, $E_{s-1}(V_s)$, and the direct value or the present discounted value of earnings if the person does not continue in school, Y_s:

$$\begin{aligned} O_{s,s-1} &= E_{s-1}[V_s - Y_s] \\ &= E_{s-1} \max\left\{0, \pi_{s+1,s}\left(\frac{E_s(V_{s+1})}{1+r} - Y_s\right)\right\} \\ &= \max\left\{0, \pi_{s+1,s}\left(\frac{E_{s-1}(V_{s+1})}{1+r} - Y_s\right)\right\}, \end{aligned}$$

where the final equality follows from the assumption that there is no uncertainty about earnings conditional on the final schooling outcome.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Notice that if the growth rate of earnings is the same as the interest rate, as is assumed by Mincer (1958), or if the growth in earnings with schooling is at the same rate as the individual-specific interest rate in the accounting identity model, then $Y_s = \frac{Y_{s+1}}{1+r}$ for each individual and all s.
- Under this assumption, Mincer's assumption of linearity of log earnings in schooling implicitly rules out any option value of schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Intuitively, if the earnings profiles associated with all schooling choices provide the same present value when discounted back to the same date, then there is no value attached to the possibility of continuation of schooling.
- Thus, linearity of log wages in years of schooling with a growth rate equal to the interest rate implies no option value of education in the Comay, Melnik, and Pollatschek (1973) framework.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- This model generates option values when future wage growth is greater than $1 + r$ for an additional year of schooling.
- For example, if college graduation offers large returns, finishing high school will carry an option value since there is some probability that an individual will be accepted into college.
- In this case, the total value of a high school degree includes the value of a lottery ticket that pays the rewards of a college degree to 'winners'.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The option value of high school represents the value of this lottery ticket scaled by the probability that the option will arise.
- Notice that even if the probability of being accepted to college is one ($\pi_{s+1,s} = 1$), if s corresponds to the state of high school graduation, there is an option value.
- Thus even in a certain environment, because of the staged nature of the schooling process, option values may arise.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The Comay, Melnik, and Pollatschek (1973) model assumes that the probability of transiting to higher grades is exogenously determined by a lottery.
- Because there is no uncertainty about future earnings paths conditional on schooling or about the future costs, their model isolates the role played by a non-linear log earnings – schooling relationship in determining option values.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- We now consider an economically more interesting model of the schooling choice problem that incorporates uncertainty in future earnings (or school costs) and sheds light on the impact of that uncertainty on the option value of education.
- This model motivates recent work in the economics of education by Keane and Wolpin (1997), Belzil and Hansen (2002) and Heckman and Navarro (2006).
- Suppose that there is uncertainty about net earnings conditional on s, so that actual lifetime earnings for someone with s years of school are

$$Y_s = \left[\sum_{x=0}^{T} (1+r)^{-x} Y(s,x) \right] \epsilon_s .$$

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- This form of uncertainty is a one time, schooling-specific shock.

- The literature discussed below considers more general models with age or period-specific shocks, but we start with this simple set up to motivate ideas.

- We assume that $E_{s-1}(\epsilon_s) = 1$ and define expected earnings associated with schooling s conditional on current schooling $s-1$,

$$\bar{Y}_s = E_{s-1}(Y_s).$$

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The disturbance, ϵ_s, may represent a shock to additional schooling costs or to current earnings that is revealed after the decision to attend grade s is made at the end of $s-1$ but prior to any future schooling decisions.
- Individuals with s years of schooling must decide whether to quit school and receive lifetime earnings of Y_s, or continue on in school for an additional year and receive an expected lifetime earnings of $E_s(V_{s+1})$.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The decision problem for a person with s years of schooling given the sequential revelation of information is to complete another year of schooling if

$$Y_s \leq \frac{E_s(V_{s+1})}{1+r},$$

so the value of schooling level s, V_s, is

$$V_s = \max\left\{ Y_s, \frac{E_s(V_{s+1})}{1+r} \right\}$$

for $s < \bar{S}$.

- At the maximum schooling level, $\bar{S}$, after all information is revealed, we obtain $V_{\bar{S}} = Y_{\bar{S}} = \bar{Y}_{\bar{S}} \epsilon_{\bar{S}}$.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The endogenously determined probability of going on from school level s to $s+1$ is

$$p_{s+1,s} = Pr\left(\epsilon_s \leq \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right),$$

where $E_s(V_{s+1})$ may depend on ϵ_s because it enters the agent's information set.

- The average earnings of a person who stops at schooling level s are

$$\bar{Y}_s E_{s-1}\left(\epsilon_s \middle| \epsilon_s > \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right). \tag{8}$$

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Thus, the expected value of schooling level s as perceived at current schooling $s-1$ is:

$$E_{s-1}(V_s) = (1 - p_{s+1,s})\bar{Y}_s E_{s-1}\left(\epsilon_s \middle| \epsilon_s > \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right) + p_{s+1,s}\left(\frac{E_{s-1}(V_{s+1})}{1+r}\right).$$

- The first component is the direct return. The second component arises from the option to go on to higher levels of schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Assuming that schooling choices are irreversible, the *option value* of schooling s, as perceived after completing $s-1$ levels of schooling given that the agent has the information about all of the shocks ϵ_{s-j}, $j \geq 1$, is the difference between the expected value of the earnings associated with termination at schooling level s and the corresponding value function:

$$O_{s,s-1} = E_{s-1}\left[V_s - Y_s\right].$$

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- These option values can be defined for all s. Option values are non-negative for all schooling levels, since $V_s \geq Y_s$ for all s.
- The option value for the highest schooling level is zero, since there is no tomorrow and $V_{\bar{S}} = Y_{\bar{S}}$ although in reality even final schooling opens up other choices beyond schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The *ex ante* rate of return to schooling s as perceived at the end of stage $s-1$, before the information is revealed, is

$$R_{s,s-1} = \frac{E_{s-1}(V_s) - Y_{s-1}}{Y_{s-1}}. \tag{9}$$

- This expression assumes no direct costs of schooling.
- If there are up-front direct costs of schooling, C_{s-1}, to advance beyond level $s-1$, the *ex ante* return is

$$\widetilde{R}_{s,s-1} = \frac{E_{s-1}(V_s) - (Y_{s-1} + C_{s-1})}{Y_{s-1} + C_{s-1}}.$$

- This expression assumes that tuition or direct costs are incurred up front and that returns are revealed one period later.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- $\widetilde{R}_{s,s-1}$ is an appropriate *ex ante* rate of return concept because if

$$Y_{s-1} + C_{s-1} \leq \frac{E_{s-1}(V_s)}{1+r}, \tag{10}$$

i.e.,

$$r \leq \frac{E_{s-1}(V_s) - (Y_{s-1} + C_{s-1})}{Y_{s-1} + C_{s-1}} = \widetilde{R}_{s,s-1},$$

then it would be optimal to advance one more year of schooling (from $s-1$ to s) given the assumed certain return on physical capital r.

- The *ex post* return as of period s is

$$\frac{V_s - (Y_{s-1} + C_{s-1})}{Y_{s-1} + C_{s-1}}.$$

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The distinction between *ex ante* and *ex post* returns to schooling is an important one that is not made in the conventional literature on "returns to schooling" surveyed in Willis (1986) or Katz and Autor (1999).
- Below, we survey a literature that demonstrates that uncertainty is an empirically important feature of lifetime earnings.
- Hence, option values play an important role in computing the theoretically motivated *ex ante* return.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- This analysis highlights the sequential nature of the schooling choice problem under uncertainty.
- The schooling allocations that arise out of this framework differ from those implied by the standard Mincer approach, which uses a static decision rule based on expected earnings profiles as of some initial period.
- The sequential approach recognizes that individuals face uncertainty at the time they make their schooling decisions and that some of that uncertainty is resolved after each schooling decision is made.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- After completing a schooling level, individuals observe the shock associated with that level and can base their decision to continue in school on its realization.
- This, along with any nonlinearity in the reward function, can create an option value of attending school.
- If the shock at stage s is bad, one can always continue to the next higher schooling level, $s + 1$.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- It is interesting to note that even when $\bar{Y}_s = \frac{\bar{Y}_{s+1}}{1+r}$ as assumed by Mincer's models, there is still an option value in this framework.
- This is so because after completing s, new information about the actual returns associated with that choice offers the option of continuing on to level $s+1$ with fresh draws of the ϵ.
- This is in contrast to the role of uncertainty in the simple Comay, Melnik, and Pollatschek (1973) model.
- More generally, when future earnings choices (Y_{s+1} vs. Y_s in this example) offer very large expected returns, the option value might be quite substantial – both sources of option values are at work.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Conventional rate of return calculations for comparing the "returns" to schooling levels s and $s+1$ base the calculation only on the direct or terminal earnings streams associated with s and $s+1$.
- Taking into account the option value also requires consideration of the earnings stream associated with higher schooling levels.
- That is, the value of graduating from high school instead of dropping out is affected by the expected earnings associated with graduating from college.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Keane and Wolpin (1997) develop sequential models of schooling.
- Although not the focus of their analysis, option values can be derived from the estimated value functions associated with different schooling levels.
- Heckman and Navarro (2006) present a more general approach to information revelation by allowing for serially correlated unobservables.
- They also establish semiparametric identification of their model.
- We briefly discuss their work below.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- To illustrate the role of uncertainty and non-linearity of log earnings in terms of schooling, we simulate a five schooling-level version of our model with uncertainty.
- Results are reported in Tables 6a and 6b.
- In both tables, we assume an interest rate of $r = 0.1$ and further assume that ϵ_s is independent and identically distributed log-normal: $\log(\epsilon_s) \sim N(0, \sigma)$ for all s.
- We assume that $\sigma = 0.1$ in the results presented in the tables.

Table 6: Simulated Returns under Uncertainty with Option Values

(Log Wages Linear in Schooling: $\bar{Y}_{s+1} = (1+r)\bar{Y}_s$)

Ed. Level (s)	Transition Probability ($p_{s,s-1}$)	Proportional Increase in $\bar{Y}$	Proportional Increase in Observed Earnings	Option/ Total Value $\frac{O_{s,s-1}}{E_{s-1}(V_s)}$	Avg. Return $E_{s-1}[R_{s,s-1}]$	Treatment on Treated	Treatment on Untreated
2	0.796	0.100	0.086	0.075	0.201	0.242	0.041
3	0.746	0.100	0.082	0.060	0.182	0.231	0.037
4	0.669	0.100	0.072	0.038	0.155	0.216	0.032
5	0.520	0.100	0.016	0.000	0.111	0.196	0.019

OLS (Mincer) estimate of the rate of return is 0.063.

(Sheepskin Effects: $\bar{Y}_{s+1} = (1+\rho_{s+1})\bar{Y}_s$ with $\rho_2 = 0.1$, $\rho_3 = 0.3$, $\rho_4 = 0.1$, $\rho_5 = 0.2$)

Ed. Level (s)	Transition Probability ($p_{s,s-1}$)	Proportional Increase in $\bar{Y}$	Proportional Increase in Observed Earnings	Option/ Total Value $\frac{O_{s,s-1}}{E_{s-1}(V_s)}$	Avg. Return $E_{s-1}[R_{s,s-1}]$	Treatment on Treated	Treatment on Untreated
2	0.997	0.100	0.101	0.239	0.459	0.460	0.068
3	0.997	0.300	0.116	0.100	0.459	0.460	0.068
4	0.846	0.100	0.092	0.093	0.224	0.257	0.045
5	0.822	0.200	0.041	0.000	0.212	0.249	0.043

OLS (Mincer) estimate of the rate of return is 0.060.

Table 6: Notes

The simulated model assumes lifetime earnings for someone with s years of school equal $\bar{Y}_s \epsilon_s$ where ϵ_s are independent and identically distributed $\log(\epsilon_s) \sim N(0, 0.1)$. An interest rate of $r = 0.10$ is assumed. The transition probability from $s-1$ to s is given by $p_{s,s-1} = \Pr_{s-1}\left(\epsilon_{s-1} \leq \frac{E_{s-1}(V_s)}{(1+r)\bar{Y}_{s-1}}\right)$, where the subscript means that the agent conditions his/her information on that available at $s-1$. Observed earnings for someone with s years of school are $\bar{Y}_s E_{s-1}\left(\epsilon_s | \epsilon_s > \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right)$, and option values are $E_{s-1}(V_s - Y_s)$. The return to school year s for someone with earnings Y_{s-1} is $R_{s,s-1} = \frac{E_{s-1}(V_s) - Y_{s-1}}{Y_{s-1}}$.

Table 6: Notes (cont.)

Average returns reflect the expected return over the full distribution of Y_{s-1}, or $E_{s-1}[R_{s,s-1}]$. "Treatment on Treated" reflects returns for those who continue to grade s, or $E_{s-1}\left[R_{s,s-1}|\epsilon_{s-1} \leq \frac{E_{s-1}(V_s)}{(1+r)\bar{Y}_{s-1}}\right]$. "Treatment on Untreated" reflects returns for those who do not continue to grade s, or $E_{s-1}\left[R_{s,s-1}|\epsilon_{s-1} > \frac{E_{s-1}(V_s)}{(1+r)\bar{Y}_{s-1}}\right]$. The marginal treatment effect equals $r = 0.10$. OLS (Mincer) estimate is the coefficient on schooling in a log earnings regression (the Mincer return).

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Table 6a reports various outcomes related to the returns to schooling when we assume log earnings are linear in years of schooling (i.e., $\bar{Y}_{s-1} = \bar{Y}_s/(1+r)$).
- Schooling continuation probabilities ($p_{s,s-1}$) and the proportional increase in $\bar{Y}$ associated with an increase in schooling from $s-1$ to s are shown.
- By assumption, the latter is equal to $r = 0.1$ for all education levels.
- Column 4 displays the proportional increases in observed earnings (where observed earnings are measured by equation 8) from period $s-1$ to s, which are always less than r.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- In the presence of uncertainty, self-selection leads to a substantial *downward* bias in the observed returns to schooling, especially for the schooling transitions associated with higher grades.
- The traditional ability bias model discussed below predicts an *upward* bias in *OLS* estimates of the return to schooling.
- In a sequential model with serially independent shocks, people with a good draw at lower schooling levels drop out, thus producing a downward bias.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Option values as a fraction of the total expected value of a schooling level $(O_{s,s-1}/E_{s-1}(V_s))$ are reported in column 5.
- They show a pattern of decline with schooling levels attained.
- The final three columns report average measures of the return to schooling for different sets of individuals.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Column 6 reports the average return for the entire population ($E_{s-1}[R_{s,s-1}]$), while column 7 reports estimates of the return for those who choose to continue on to grade s ("treatment on the treated") and column 8 reports the expected return that would be received by those who choose not to continue in school ("treatment on the untreated").
- Here the treatment is schooling at the stated level.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Comparing average returns with the proportional increase in $\bar{Y}$ with schooling or in observed earnings with schooling, observe that total rates of return to schooling are substantially higher for all but the final schooling transition due to the additional effect of the option value of school and the self-selection that takes place.
- When log earnings are linear in schooling, true returns are actually declining in accumulated schooling since option values are decreasing in s.
- Returns for those who choose to continue in school are noticeably larger than average returns, while returns for those who choose not to continue are all less than r.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The least squares estimate of the rate of return to school (i.e., the coefficient on schooling in a log earnings regression or the "Mincer coefficient") is only 0.063, far below the estimates of the true average growth rate (ATE) or treatment on the treated (TT), the growth rate among the treated.
- It also under-estimates the rate of increase in expected earnings, $\bar{Y}_s$, and does not accurately reflect the pricing relationship for wages and schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Even under linearity of mean log earnings in schooling, Mincer-based estimates of the return are substantially downward biased in the presence of sequential resolution of uncertainty.
- Not surprisingly, this bias (along with option values) disappears as the variance of ϵ_s goes to zero.
- However, we find a bias as large as -0.01, roughly 10% of the true return, when σ is as low as 0.01.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Table 6b adds nonlinearity in the wage equation in terms of schooling to the base model to demonstrate its added effect on rates of return and option values.
- The simulation reported in this table assumes that increases in population mean log earnings from the first to the second and third to fourth levels of school are both 0.1, but the increase associated with going from level two to three is 0.3 and from four to five is 0.2.
- This roughly mimics the patterns observed in the later Census years with schooling levels three and five representing high school and college graduation, respectively.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- These simulations show substantially larger returns to the lower school transitions as a result of the sizeable sheepskin effects in later years.
- Option values are particularly large in early schooling years.
- In general, the greater the nonlinearity, the greater the option value.
- Estimates from a Mincer regression suggest a rate of return of only 0.060, substantially less than the true average growth rate or the treatment on the treated growth rate estimates, which range from 0.21 to 0.46.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- While true returns increase relative to those reported in Table 6a, the Mincer estimate actually declines slightly.
- Because most individuals are choosing to continue to higher schooling levels in this simulation, there is little difference between "average returns" and estimated treatment on the treated parameters.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The simulations presented in Tables 6a and 6b point to the potentially important role of both sources of option values in determining total returns to schooling.
- Turning to real data, we use the nonparametrically estimated earnings profiles for white males in the 1990 Census to compute the option value of high school completion and college attendance for a range of reasonable schooling transition probabilities, p, and interest rates, r.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- These estimates are unbiased measures of the option value within the framework of Comay, Melnik, and Pollatschek (1973) where $p_{s+1,s} = \pi_{s+1,s}$ are the empirical transition probabilities for the schooling levels we examine because selection is random with respect to individual earnings levels.

- For a model of sequential resolution of uncertainty, where $p_{s+1,s}$ is $Pr\left(\epsilon_s \leq \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right)$ and ϵ_s is in the information set used to define E_s, they under-estimate the option value and return to schooling, since observed earnings are $\bar{Y}_s E_{s-1}\left(\epsilon_s \middle| \epsilon_s > \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right)$ rather than $\bar{Y}_s$ (i.e., observed earnings are based on a sample selecting not to continue).

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Table 7 reports the average discounted lifetime earnings for individuals making different schooling choices, denoted by $\hat{Y}_s$.
- It also reports the total expected value of a schooling choice, $E_{s-1}(V_s)$, the implied option value, $\hat{O}_{s,s-1}$, and return to schooling, $R_{s,s-1}$.
- The table reports estimates based on interest rates of 7% and 10% and transition probabilities ranging from 0.1 to 0.5 (empirically, about half of all 1990 high school graduates attended college and about half of those went on to graduate).

Table 7: Present Value of Earnings, Option Values, and Return to Schooling
(White Men, 1990 Census)

Interest Rate	Transition Probability	PV Lifetime Earnings (in \$1000's)			Option Value (in \$1000's)		Total Value (in \$1000's)		Return to Schooling	
r	p	$\hat{Y}_{12}$	$\hat{Y}_{14}$	$\hat{Y}_{16}$	$\hat{O}_{12,10}$	$\hat{O}_{14,12}$	$E(V_{12})$	$E(V_{14})$	$\hat{R}_{12,10}$	$\hat{R}_{14,12}$
0.07	0.1	226.46	274.15	394.97	1.92	7.08	228.38	281.23	0.24	0.11
0.07	0.3	226.46	274.15	394.97	9.47	21.25	235.92	295.40	0.26	0.14
0.07	0.5	226.46	274.15	394.97	21.96	35.41	248.42	309.56	0.30	0.17
0.1	0.1	149.26	181.17	266.12	0.37	3.88	149.63	185.05	0.27	0.11
0.1	0.3	149.26	181.17	266.12	3.02	11.63	152.29	192.80	0.28	0.14
0.1	0.5	149.26	181.17	266.12	8.24	19.38	157.51	200.56	0.31	0.16

Notes: Transition probability, p, represents the probability of continuing in school conditional on current education. "PV of lifetime earnings" is $\hat{Y}_s = \sum_{x=0}^{65} (1+r)^{-x}\hat{Y}(s,x)$ where $\hat{Y}(s,x)$ is the nonparametrically estimated earnings for a white man with s years of school and x years of experience (based on the 1990 Census). "Total value", $E(V_s) = (1-p)\hat{Y}_s + p(1+r)^{-1}E(V_{s+1})$, is recursively solved backward from $E(V_{16}) = \hat{Y}_{16}$. "Option value" is $\hat{O}_{s,s-1} = E(V_s) - \hat{Y}_s$. "Return to school" $\hat{R}_{s,s-1} = \frac{E(V_s)-\hat{Y}_{s-1}}{\hat{Y}_{s-1}}$ is annualized. See Appendix B for data description.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- As expected, both the present value of earnings for each schooling choice and the option value of continuing are declining in the interest rate.
- Option values rise with increases in the transition probability.
- The option value for high school completion ranges from a low of only \$370 when the interest rate is 10% and $p = 0.1$ to a high of \$22,000 when interest rates are 7% and $p = 0.5$.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The major component of this option value comes from the return to completing college rather than the return to attending college, because the difference in earnings between high school graduates and those with some college is quite small.
- Accordingly, option values are noticeably higher for college attendance, reaching a high of \$35,000 when the interest rate is 7% and $p = 0.5$.
- Simply comparing the earnings streams for two schooling levels fails to recognize a potentially important component of the returns to education.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Rates of return, shown in the final two columns, increase by about 50% for college attendance when the transition probability is raised from 0.1 to 0.5.
- Returns to high school completion are less sensitive to assumptions about p and the option values.
- Failing to consider option values leads to biased estimates of the true return to schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- We conclude this section by considering whether the internal rate of return has any relevance in a model with sequential updating of information or in a model with a lottery structure, like the framework of Comay, Melnik, and Pollatschek (1973).
- Investment criterion (10) based on (9) is the appropriate criterion for *ex ante* calculations.
- *Ex post* returns, of the sort traditionally reported in the labor economics literature, are obtained by using realized values of earnings.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- The natural generalization of the IRR to an environment with sequential revelation of information would be as that rate that equates value functions across different schooling levels defined relative to some information set at the date schooling choices are being made.
- However, even for a particular information set, single crossings of realized age-earnings profiles, a near universal feature of schooling-earnings data, do not guarantee unique internal rates of return applied to the valuation function when option values are taken into account.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Hirshleifer (1970) shows that there is always a unique positive internal rate of return when comparing two deterministic earnings streams which cross at only one age.
- This is the typical case when comparing the earnings profiles for any two schooling levels.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Accounting for options to continue in school, it is possible for multiple roots to arise in the computation of more sophisticated internal rates of return that account for the option value of schooling even if earnings are monotonically increasing in schooling for workers conditional on age, and there are single crossings of any two earnings streams.
- Intuitively, the value function is a weighted average of future earnings streams so a single crossing property for earnings streams is not enough to guarantee unique internal rates of return for value functions.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- To explore this intuition formally, consider a model of exogenous schooling transition probabilities like that of Comay, Melnik, and Pollatschek (1973) for the case where earnings are zero until the end of school, age s, at which time they jump up to $\alpha_s + \beta s$ and linearly increase thereafter at rate $\beta > 0$.
- Assume that there are no direct or psychic costs of schooling.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- As long as $\alpha_s > \alpha_{s'}$ for all $s > s'$, any two earnings streams will only cross once at the age where the higher schooling level ends.
- Letting $Y(s,a)$ denote the earnings for someone with s years of school at age a, we have

$$Y(s,a) = \begin{cases} 0 & \text{if } a < s \\ \alpha_s + \beta a & \text{if } a \geq s. \end{cases}$$

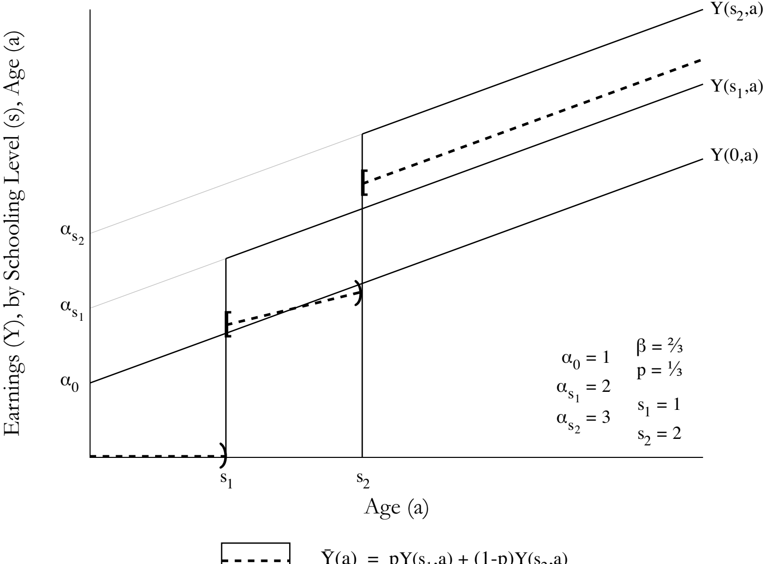
The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Consider three schooling choices, $s \in \{0, s_1, s_2\}$.
- Suppose p is the exogenously specified probability that someone with $s_1 < s_2$ years of school continues on to s_2 years.
- The expected earnings stream at age a of someone choosing to attend s_1 years of school with the option of continuing will be $\bar{Y}(a) = (1 - p)Y(s_1, a) + pY(s_2, a)$.
- An option value arises because the agent has a chance of getting into schooling level s_2 after completing schooling level s_1.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- For $\alpha_0 < \alpha_{s_1} < \alpha_{s_2}$, $\bar{Y}(a)$ will cross $Y(0, a)$ three times whenever
$$\frac{\alpha_0 + \beta s_1}{\alpha_{s_1} + \beta s_1} < 1 - p < \frac{\alpha_0 + \beta s}{\alpha_{s_1} + \beta s}$$
for any s, where $s_1 < s < s_2$.
- This possibility is illustrated in Figure 7.
- Because $\bar{Y}(a)$ crosses $Y(0, a)$ three times, the internal rate of return equations for the value functions produced from this model can generate multiple roots.
- Even if pairwise earnings streams cross only once, there may be multiple internal rates of return when we use the appropriate value function, invalidating their use as a guide to selecting human capital investment projects.

Figure 7: Three Crossings of Prospective Earnings Profiles in a Model with Option Values



The Internal Rate of Return and The Sequential Resolution of Uncertainty

- In the more general case of sequential resolution of uncertainty, the schooling transition probability is not exogenous.
- Multiple roots are even more likely in this case, since the transition probability depends on the discount rate.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Writing equations out explicitly in terms of interest rate r, we obtain

$$\begin{aligned} & E_{s-1}\left(V_s\left(r\right)\right) \\ & \quad = \Pr_{s-1}\left(\epsilon_s \geq \frac{E_s\left[V_{s+1}\left(r\right)\right]}{\left(1+r\right)\bar{Y}_s\left(r\right)}\right)\bar{Y}_s\left(r\right) \\ & \qquad \times E_{s-1}\left(\epsilon_s \,\middle|\, \epsilon_s \geq \frac{E_s\left[V_{s+1}\left(r\right)\right]}{\left(1+r\right)\bar{Y}_s\left(r\right)}\right) \\ & \qquad + \Pr_{s-1}\left(\epsilon_s < \frac{E_s\left[V_{s+1}\left(r\right)\right]}{\left(1+r\right)\bar{Y}_s\left(r\right)}\right)\frac{E_{s-1}\left[V_{s+1}\left(r\right)\right]}{\left(1+r\right)}. \end{aligned}$$

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- In this setting, the natural generalization of the IRR is the value (or values) of r_I that solves

$$Y_s(r_I) = \frac{E_{s-1}\left(V_{s+1}(r_I)\right)}{1+r_I}.$$

- Take a three period example.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- In this case, the IRR for the second level of schooling solves

$$
\begin{aligned}
&\bar{Y}_1(r_I) \\
&\quad = \Pr_1\left(\epsilon_2 \geq \frac{\bar{Y}_3(r_I)}{(1+r_I)\,\bar{Y}_2(r_I)}\right)\frac{\bar{Y}_2(r_I)}{1+r_I} \\
&\qquad\qquad \times E_1\left(\epsilon_2 \,\middle|\, \epsilon_2 \geq \frac{\bar{Y}_3(r_I)}{(1+r_I)\,\bar{Y}_2(r_I)}\right) \\
&\qquad + \Pr_1\left(\epsilon_2 < \frac{\bar{Y}_3(r_I)}{(1+r)\,\bar{Y}_2(r_I)}\right)\frac{\bar{Y}_3(r_I)}{(1+r_I)^2}
\end{aligned}
$$

The fact that the continuation probabilities also depend on r_I makes multiple roots more likely.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- To gain some intuition in this case, take a limiting case where the variance of ϵ_2 goes to zero.
- This implies that the probability of continuing to level three will be either zero or one, depending on whether or not $\bar{Y}_2$ is greater or less than $\frac{\bar{Y}_3}{(1+r_I)}$.
- We may, therefore, get two valid solutions to the above IRR equation:

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Case 1 (individual always continues): r_I^1 satisfies

$$\bar{Y}_1\left(r_I^1\right) = \frac{\bar{Y}_3\left(r_I^1\right)}{\left(1 + r_I^1\right)^2} > \frac{\bar{Y}_2\left(r_I^1\right)}{1 + r_I^1}.$$

- The latter inequality guarantees that the person always wants to continue to schooling level three upon reaching level two.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Case 2 (individual never continues): r_I^2 satisfies

$$\bar{Y}_1\left(r_I^2\right) = \frac{\bar{Y}_2\left(r_I^2\right)}{\left(1 + r_I^2\right)} > \frac{\bar{Y}_3\left(r_I^2\right)}{\left(1 + r_I^2\right)^2}.$$

- The latter inequality guarantees that the person always stops his schooling at level two.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Both of these cases can arise for the same person depending on the realization of ε_2 as long as $\text{Var}(\varepsilon_2) > 0$ if log earnings are not parallel in experience.
- Consider the case where wage gaps are small initially and large later in the life cycle.
- In this case, r_I^1 would be less than r_I^2.
- In Case 1, the high wage differential later on is not discounted very much, so the individual always wants to attend schooling level three.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- A low IRR must, therefore, equate level one earnings with discounted level three earnings.
- On the other hand, the high late wage differential may be discounted so much with a high discount rate that the individual never chooses to go on to college at that rate.
- In this case, a high IRR, r_I^2, must equate level one earnings with discounted level two earnings.
- These examples are extreme, but multiple roots can arise more generally as long as the variance of ε_s is not too large.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- This type of multiplicity of roots could also come more directly out of the Comay, Melnik, and Pollatschek (1973) type of model, where the probability of continuing to level three would be either zero (if individuals do not want to continue) or p (if individuals wish to continue), depending on the discount rate.
- Given the lack of parallelism in cross section log earnings profiles, multiplicity of roots is likely to be empirically important.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- These issues call into serious question the usefulness of internal rates of return as a measure of the return to education in an environment where the schooling decision is dynamic and sequential.
- A central tool of policy evaluation from classical human capital theory loses its validity in the presence of option values.
- Criterion (9) does not suffer from this criticism and is the appropriate measure of the *ex ante* rate of return to use but it is rarely reported.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- For an exception, see Cunha, Heckman, and Navarro (2005) and Cunha and Heckman (2006b) who estimate this rate of return.
- In the absence of sequential resolution of uncertainty and option values, $R_{s,s-1}$ is the same as the classical internal rate of return applied to pairwise earnings streams, so it is the natural generalization of that concept.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Empirical work on the option value of schooling is in its infancy.
- If option values are empirically relatively unimportant in models with the sequential resolution of uncertainty, conventional investment evaluation methods based on the IRR may well be informative on the optimality of schooling investments.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- Even if option values are negligible, the analysis presented throughout this paper suggests that the Mincer model will not estimate theoretically appropriate rates of return to schooling.
- In the absence of option values, other key assumptions required to equate Mincer coefficients with internal rates of return are violated.
- Even in an environment without the sequential resolution of uncertainty, more general methods of the type presented below are required to obtain internal rates of return.

The Internal Rate of Return and The Sequential Resolution of Uncertainty

- We next turn to an analysis of cross section bias.
- In doing so, we ignore option values, following conventions in the labor economics literature, and focus on "rates of return" as conventionally measured to concentrate on the issue of whether cross section estimates of "rates of return" are valid for life cycle "rates of return."

Part VI

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Thus far we have considered estimation of rates of return to schooling using cross-section data which applies the standard synthetic cohort approach followed by most of the literature.
- For an *ex ante* analysis it assumes that younger workers base their earnings expectations on the current experiences of older workers.
- For an *ex post* analysis, it assumes that the experiences of older workers at a point in time will be those of younger workers when they reach those ages.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- If skill prices are changing over time and workers at least partially anticipate these changes, the estimates of the *ex ante* return to different schooling levels based on cross-sectional data may not represent the *ex ante* rates of return governing human capital investment decisions.
- Similarly, if the environment is nonstationary, the *ex post* returns of the younger cohort are not accurately estimated.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- While estimates based on cross-section data reflect current price differentials and opportunity costs, they do not capture future skill price differentials that forward-looking individuals would take into account.
- The U.S. labor market in recent years is highly nonstationary as are the labor markets of many economies around the world.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- If cohorts anticipate future changes in the skill premium, they will base their schooling decisions on their true cohort-specific rate of return and not the rate of return estimated from a cross-section of current workers.
- However, if individuals do not anticipate the future price changes, cross-section estimates may better represent their expectations about the returns to school.
- Expectations play a crucial role in determining whether cross-section or cohort-based estimates of the rate of return influence schooling decisions.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Another possible source of discrepancy between cross-section and cohort-based rate of return estimates is change in cohort quality, as might arise from changes in the quality of schools over time.
- If relative skills for some schooling classes increase permanently, then cohort rates of return jump up with the first ‘new’ cohort and remain higher for all succeeding cohorts.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Cross-section estimates only reflect the changes slowly as more high quality cohorts enter the sample each year.
- As a result, they under-estimate true rates of return for cohorts entering the labor market after the change in school quality, with the bias disappearing as time progresses.
- While future price changes are difficult to predict, changes in cohort or school quality are more identifiable.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Mincer (1974) addressed cross section bias in his pioneering work.
- He found that patterns for wage growth in a 1956 cross-section of male workers were quite similar to the 1956 to 1966 growth in wages for individual cohorts.
- The empirical discrepancy between cross-section and cohort-based estimates was relatively small.
- Recent analyses reveal that wage patterns have changed dramatically across cohorts and that cross-sections no longer approximate cohort or life cycle change (MaCurdy and Mroz, 1995; Card and Lemieux, 2001).

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- While these studies do not agree on whether or not these changes are due to changes in relative skill prices or cohort quality, there is little question in the U.S. data that life cycle earnings profiles based on a cross-section of workers no longer accurately reflect the true earnings patterns for any given cohort.
- As a result, the rates of return to schooling estimated from cross-sections of workers reported in the previous section are likely to differ from the rates of return faced by cohorts making their schooling decisions.

$$\ln W_i = \alpha_0 + \underset{\text{age}}{\alpha_1 \underset{\uparrow}{a_i}} + \underset{\text{year}}{\alpha_2 \underset{\uparrow}{y}}$$

$$\underset{\substack{\uparrow \\ \text{experience}}}{\alpha_3 e_i} + \underset{\substack{\uparrow \\ \text{schooling}}}{\alpha_4 s_i} + \underset{\substack{\uparrow \\ \text{vintage (birth cohort)}}}{\alpha_5 c_i} + u_i$$

Two Identities

$$e_i = a_i - s_i \qquad \text{“experience”} \tag{11}$$

$$y = a_i + c_i \qquad c_i = \text{birth year} \tag{12}$$

- Solve out for c_i and a_i to get estimable combinations.

- Take the simpler case first:

$$\begin{aligned} \ln W(a, y, c) &= \beta_0 + \underset{\text{(age)}}{\beta_1 a_i} + \underset{\text{(year)}}{\beta_2 y_i} + \underset{\text{(cohort)}}{\beta_3 c_i} + u_i \\ y_i &= a_i + c_i, \end{aligned}$$

where y_1 is the current year, and c_i is the year of birth.

- Obviously, we get an exact linear dependence:

$$(\beta_0, \beta_1, \beta_2, \beta_3)$$

- Substitute $c_i = y_i - a_i$.

- $$\begin{aligned} \ln W_i &= \alpha_0 + \beta_1 a_i + \beta_2 y_i + \beta_3 (y_i - a_i) + u_i \\ &= \alpha_0 + (\beta_1 - \beta_3) a_i + (\beta_2 + \beta_3) y_i + u_i \end{aligned}$$

  can identify only combinations of coefficients.

- In a cross section, y_i is the same for everyone. The intercept is

$$[\alpha_0 + (\beta_2 + \beta_3) y_i].$$

- We can estimate $(\beta_1 - \beta_3)$: age minus cohort effect.
- If $\beta_3 > 0$, we underestimate true β_1.
- Will longitudinal data rescue us? — Not necessarily.
- With panels, y_i moves with time. Recall that $y_i = a_i + c_i$.
- So we still have exact linear dependence. This is true if we have dummy variables in place of continuous variables (verify). Panel data will rescue us — if we have no year effects.

- We acquire similar problems in models with nonlinear terms:

$$y = a + c$$

$$\left.\begin{aligned} y^2 &= a^2 + 2ac + c^2 \\ ay &= a^2 + ac \\ cy &= ca + c^2 \end{aligned}\right\} \text{3 linear dependencies in these set-ups}$$

- Thus when we write

$$\begin{aligned} \ln W &= \beta_0 + \beta_1 a + \beta_2 y + \beta_3 c + \beta_4 a^2 + \beta_5 ac \\ &\quad + \beta_6 ay + \beta_7 cy + \beta_8 c^2 + \beta_9 y^2 + u, \end{aligned}$$

we cannot identify all of the parameters (only 3 second order parameters are estimable out of 6 total.

Theorem

. In a model with interactions of order k with j variables and one linear restriction among the j variables, then of the $\binom{j+k-1}{k}$ coefficients of order k, only $\binom{j+k-2}{k}$ are estimable. (Heckman and Robb, in S. Feinberg and W. Mason, *Age, Period and Cohort Effects: Beyond the Identification Problem*, Springer, 1986).

E.g. $k = 2$, $j = 3$; 6 coefficients and 3 are estimable, as in the preceding example.

Theorem. In a model with ℓ restrictions on the j variables, then $\binom{j+k-\ell-1}{k}$ kth order coefficients are estimable (Heckman and Robb, 1986).

- Return to the more general case. Substitute out for c_i and a_i, using (11) and (13):

$$\ln W_i = \alpha_0 + (\alpha_2 + \alpha_5)y + (\alpha_1 + \alpha_3 - \alpha_5)e_i + (\alpha_1 + \alpha_4 - \alpha_5)s_i + u_i.$$

- In a single cross section, y is the same for everyone. The intercept is then $\alpha_0 + (\alpha_2 + \alpha_5)y$, where y is year of cross section.

- Experience coefficient $= \alpha_1 + \alpha_3 - \alpha_5 = \alpha_3 + (\alpha_1 - \alpha_5)$ if later vintages get higher skills, $\alpha_5 > 0$ and downward bias (*e.g.* higher quality of schooling). If there is an aging effect (> 0, *e.g.* maturation) cannot separate. Produces upward bias for α_3.

Schooling Coefficient

- $\alpha_1 + \alpha_4 - \alpha_5 = \alpha_4 + (\alpha_1 - \alpha_5)$

- Vintage (cohort) effects lead to downward bias.

- Age effects, upward bias.

- Observe that from the
experience coefficient – schooling coefficient:

$$(\alpha_1 + \alpha_3 - \alpha_5) - (\alpha_1 + \alpha_4 - \alpha_5) = \alpha_3 - \alpha_4.$$

- Can estimate difference in "returns" to experience net of schooling.

- Observe that even if α_1=0 (no aging effect), still can't estimate these coefficients.

- Is the solution **longitudinal data** (observations n the same people over time)—or **repeated cross section data** (observations on the same population over time but sampling different persons)?

- If $\alpha_2 = 0$,(no year effects), we can estimated α_5.

- Alternatively, for each c_i we can estimate $\alpha_1 + \alpha_3$, and hence we can estimate α_5.

- We also know $\alpha_1 + \alpha_4$. If $\alpha_1 = 0$, then $\alpha_{3,}$ $\alpha_{4,}$ α_5 identified.

- Observe the weakness in the procedure.
- If year effects are present, we have that there is no gain to going to longitudinal or repeated cross section data.
- We gain a parameter when we move to the panel or repeated cross sectional data.

Solutions in Literature

(1) Redefine vintage (cohort) e.g. vintage fixed over period of years (e.g. a cohort of Depression babies.

- Then $\ln W = (\alpha_0 + \alpha_5 c) + \alpha_1 a + \alpha_2 y + \alpha_3 e + \alpha_4 s + u$.
- In single cross section, c and y are fixed.

- Substitute for e:

$$e = a_i - s_i$$

- Then

$$\ln W = [\alpha_0 + \alpha_5 c + \alpha_2 y] + (\alpha_1 + \alpha_3)a_i + (\alpha_4 - \alpha_3)s_i.$$

- We can estimate $\alpha_1 + \alpha_3$ and $\alpha_4 - \alpha_3$, and thus $\alpha_1 + \alpha_4$.
- Successive time periods for the same vintage gives us α_2 directly [since c doesn't move].
- If no age effect , we get $\alpha_3, \alpha_4, \alpha_2$, and from successive vintage estimations, we get α_5.

(2) If we measure experience, $a_i \neq e_i + s_i$ (non-market breaks), we get break in linear dependence.

- Cost: better proxies may be endogenous.
- E.g. experience = cumulated hours.
- Results carry over in an obvious way to nonlinear models.

Example of Interpretive Pitfall

(1) Johnson and Stafford (AER, 1974)
(2) Weiss and Lillard (JPE, 1979)

- **Fact:** Disparity in real wages between recent Ph.D. entrants and experienced workers rose in *physics* and *mathematics* in the late 60s and early 70s. Not observed in the *social sciences*.

- **Why?** — Johnson-Stafford story.

- Supplies of Ph.D.s enlarged by federal grants while demand for scientific personnel declined. Wage rigidity at the top end motivated by specific human capital. Spot market / entrant market bears the brunt of the burden.

- Weiss & Lillard: "experience – vintage" interaction (ec).
- Ignore age effect:

$$\begin{aligned}\ln W(e,c,s,y) &= \varphi_0 + \varphi_1 e + \varphi_2 c + \varphi_3 y + \varphi_4 s \\ &\quad + \varphi_5 e^2 + \varphi_6 c^2 + \varphi_7 ec \\ &\quad + \varphi_8 ey + \varphi_9 cy + \varphi_{10} y^2\end{aligned}$$

- Assume other powers and interactions are zero. Assume $\varphi_{10} = 0$.
- Johnson-Stafford: $\varphi_8 > 0$ or $\varphi_9 < 0$
- Weiss-Lillard: $\varphi_7 > 0$
- Recall that $y = e + s + c$.

- Weiss-Lillard ignore year effects.
- We get Weiss-Lillard by substituting for y:

$$\begin{aligned}\ln W(e,c,s) &= \varphi_0 + (\varphi_1+\varphi_3)e + (\varphi_3+\varphi_4)s \\ &+(\varphi_2+\varphi_3)c + (\varphi_5+\varphi_8)e^2 \\ &+\varphi_8 es + (\varphi_7+\varphi_8+\varphi_9)ec \\ &+(\varphi_6+\varphi_8)c^2\end{aligned}$$

- Note that if $\varphi_7 = 0$ but $\varphi_8 > 0$, we get *ec* interaction, but it is "really" a year effect. If entry level wages fall relative to wages of experienced workers, the wage / experience profile is steeper in more recent cross-sections.

- Looking at social scientists where no interaction appears favors Johnson-Stafford.
- Moral: auxiliary evidence and theory break the identification problem.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Next, we present a cohort analysis focusing on the actual returns earned by each cohort without taking a position on whether changes in those returns over time are due to changes in cohort quality or skill prices.
- Arias and McMahon (2001) present a similar analysis in estimating *ex post* dynamic rates of return.
- We study how the actual *ex post* returns earned by individual cohorts compare with returns estimated from a cross-section of individuals at the time those cohorts made their schooling decisions.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- We use repeated cross-section data from the 1964-2000 Current Population Survey (CPS) March Supplements, comparing cross-section estimates of the return to schooling with estimates that combine all years of the CPS to follow cohorts over their life cycles.
- Given the sensitivity noted in the previous sections of this chapter to specifications of the functional forms of earnings equations, we adopt a flexible earnings specification and compute internal rates of return to high school completion (12 vs. 10 years of schooling) and college completion (16 vs. 12 years of schooling) that relax the assumptions that log earnings are parallel in experience and linear in schooling.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Our estimates also take into account average marginal tax rates and tuition costs using the time series generated from CPS data.
- Because earnings are not observed at every experience level for any cohort in the sample (an obvious practical problem in estimating cohort rates of return), a fully non-parametric approach is infeasible.
- To extrapolate the earnings function to work experience levels not observed in the data, we assume that log earnings profiles are quadratic in experience in a specification that allows the intercept and coefficients on experience and experience-squared to vary by schooling class and year or cohort of data.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- We estimate log earnings for each year or for each cohort using regressions of the form

$$\log(Y(s,x)) = \alpha_s + \beta_{0s}x + \beta_{1s}x^2 + \varepsilon_s,$$

where the regression coefficients are allowed to vary by schooling group.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Two sets of estimates are generated:
  1. regressions are estimated separately for each year of CPS data (to produce a set of cross-section estimates), and
  2. all CPS cross-sections are combined and separate regressions are estimated for each cohort by following them over their life cycles (to produce a set of cohort-based estimates).
- Both sets of estimates are used to generate predicted life cycle earnings profiles for each cohort or cross-section of individuals, which are then used to compute internal rates of return to high school and college by the method previously described, setting the residual of the wage equation to its mean.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Figures 8a and 8b show cohort and cross-section high school and college completion IRR estimates for white men, which are based on the CPS estimates reported in Table 8a.
- Cross-section estimates are shown for each year of the sample from 1964-1995, and cohort-based estimates are shown for cohorts turning age 18 in 1950 through 1983.
- The cohort-based estimates reported in Figure 8a reveal relative stability in the return to high school for cohorts making their high school completion decisions prior to 1960, followed by a large increase in the IRR for cohorts making their decisions over the first half of the 1960s, followed by another period of relative stability.

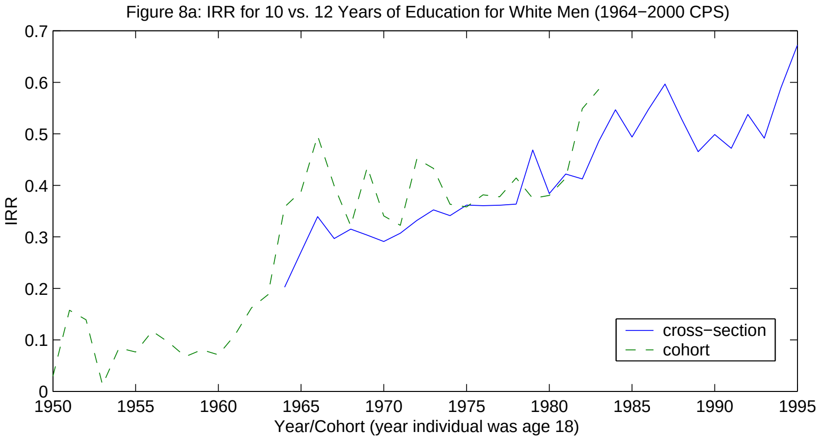


Figure 8a: IRR for 10 vs. 12 Years of Education for White Men (1964−2000 CPS)
IRR
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Year/Cohort (year individual was age 18)
cross−section
cohort

Figure 8b: IRR for 12 vs. 16 Years of Education for White Men (1964−2000 CPS)

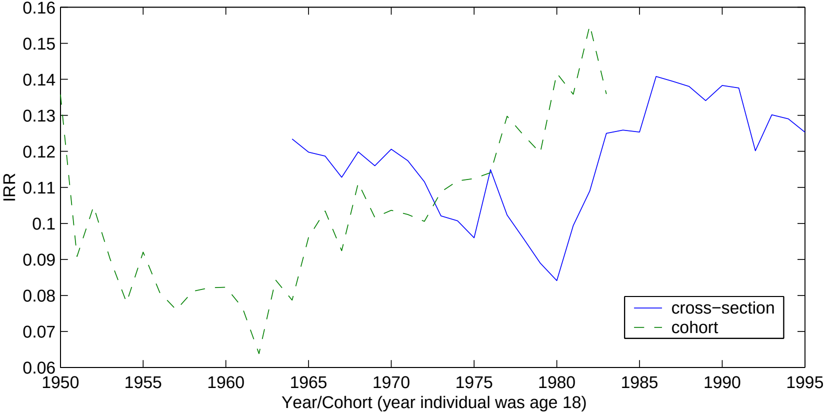


Table 8a: Internal Rates of Return for White Men: Best Census and CPS Estimates

Schooling Comparison	Year	Mincer	Census Data: General Spec. (No Residual Adjustment)	Census Data: General Spec. (Residual Adjustment)	CPS Data: Cross Section	CPS Data: Cohort
10 vs. 12	1940	13	24	8	-	-
	1950	11	26	14	-	3
	1960	12	29	16	-	7
	1970	13	29	16	29	34
	1980	11	41	24	38	38
	1990	14	47	31	50	-
12 vs. 16	1940	13	15	13	-	-
	1950	11	7	8	-	14
	1960	12	10	11	-	8
	1970	13	10	10	12	10
	1980	11	8	8	8	14
	1990	14	12	12	14	-

Notes: Mincer estimates make no adjustment for taxes or tuition. Census General Specification estimates account for tuition and progressive taxes with fully non-parametric wage specification. CPS Cross Section Estimates use cross sectional data and a general wage specification accounting for tuition and flat taxes. CPS Cohort estimates follow a cohort turning age 18 in the reported year, using a general wage specification accounting for tuition and flat taxes. See Appendix B for data description.

Table 8b: Internal Rates of Return for Black Men: Best Census and CPS Estimates

			Census Data:		CPS Data:	
			General Spec.	General Spec.		
Schooling			(No Residual	(Residual		
Comparison	Year	Mincer	Adjustment)	Adjustment)	Cross Section	Cohort
10 vs. 12	1940	9	10	-8	-	-
	1950	10	44	21	-	4
	1960	11	34	16	-	18
	1970	12	39	22	32	49
	1980	12	46	29	55	70
	1990	16	57	42	64	-
12 vs. 16	1940	9	10	6	-	-
	1950	10	5	9	-	15
	1960	11	8	6	-	6
	1970	12	10	10	12	14
	1980	12	11	9	14	17
	1990	16	17	15	16	-

Notes: Mincer estimates make no adjustment for taxes or tuition. Census General Specification estimates account for tuition and progressive taxes with fully non-parametric wage specification. CPS Cross Section Estimates use cross sectional data and a general wage specification accounting for tuition and flat taxes. CPS Cohort estimates follow a cohort turning age 18 in the reported year, using a general wage specification accounting for tuition and flat taxes. Each CPS estimate is based on three adjoining years/cohorts worth of data. See Appendix B for data description.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Returns increased from around 10% among 1950-60 cohorts to around 40% for post-1965 cohorts.
- Cross-section based estimates increase slowly but consistently over most of the 1964-1995 period.
- In general, cross-section estimated rates of return under-estimate the true rates of return earned by cohorts of white men making their schooling decisions in the late 1960s and 1970s.
- However, basic time patterns are consistent across the two sets of estimates.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- More dramatic differences are observed for the college-going decision of white men as shown in Figure 8b.
- While cross-section estimates show declining “returns” to college over the 1970s (from 12% down to 8%), cohort-based estimates show continually increasing returns from the early 1960s to the early 1980s.
- The rate of return estimated from cross-sections does not begin to increase until 1980.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- Cross-section estimates overestimate the rate of return faced by cohorts making their college attendance decisions around 1965 by as much as 4 percentage points, while estimates in the early 1980s under-estimate the return by nearly the same amount.
- Table 8b reports comparable numbers for black men.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- If the observed discrepancies between cross-section and cohort-based estimated "rates of return" are due to price changes over time that could be at least partly anticipated or are due to changing cohort quality, then cross-section estimates would not reflect the *ex ante* "rates of return" that governed schooling decisions.
- On the other hand, if changes in skill prices were entirely unanticipated, then cross-section estimates may provide a better indication of the *ex ante* returns governing schooling decisions than would the actual *ex post* returns experienced by each cohort.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- A better understanding of the underlying causes for such dramatic changes in wages and of individual expectations is needed.
- Buchinsky and Leslie (2000), Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) present empirical explorations of alternative expectation-formation models.
- We review methods for estimating agent information sets below.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- In summary, cross-section estimates of the “rate of return” to schooling should be cautiously interpreted, particularly when skill prices are changing over time or when cohort quality is changing.
- If one is interested in empirically estimating historical rates of return, a cohort analysis is clearly preferable.
- Data from the 1964-2000 March CPS suggest that “returns” estimated from a cross-section of workers are not only biased in levels, but they also suggest patterns that sometimes differ from those obtained using a cohort-based estimation strategy.

Cohort vs. Cross-Section Internal Rate of Return

- Take a cohort rate of return.

(1) $Y^h_{a,c}$ is the earnings of a high school graduate of cohort c at age a.

(2) $Y^d_{a,c}$ is the earnings of a droupout of cohort c at age a.

(3) $\rho_c = IRR_c$ (cohort internal rate of return).

(4)
$$\sum_{a=0}^{A} \frac{Y^h_{a,c} - Y^d_{a,c}}{(1+\rho_c)^a} = 0.$$

- The cross-section consists of a set of member of different cohorts.
- Start with $c = 1$ as the youngest age group and proceed.
- At a point in time, we have $a = 0 \Longrightarrow c = 1;\ c + a = t.$.
- The cross-section internal rate of return is

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,1-a} - Y^d_{a,1-a}\right)}{\left(1+\rho_t\right)^a} = 0,$$

where $A + 1$ is the maximum age in the population.

- When can $\rho_c = \rho_t$?
- This can occur if the environment is stationary.
- With steady growth in differentials, it cannot help explain $\rho_c = \rho_t$.
- The case

$$\begin{aligned} \Delta^{h,d}_{a,c} &= Y^h_{a,c} - Y^d_{a,c} \\ \Delta^{h,d}_{a,c+j} &= \left(\Delta^{h,d}_{a,c}\right)(1+g)^j \end{aligned} \tag{13}$$

will not work.

- With constant growth, g cannot explain $\rho_t = \rho_c$:

$$c = 0, 1 \qquad t = a + c.$$

- Consider a model with 2 cohorts, focus on cohort $c = 0$. ρ_c is the root of

$$0 = Y^h_{0,0} - Y^d_{0,0} + \frac{Y^h_{1,0} - Y^d_{1,0}}{1 + \rho_c}.$$

- Cross-section at $t = 1$, when cohort c enters, is

$$0 = Y^h_{0,0} - Y^d_{0,0} + \frac{Y^h_{1,-1} - Y^d_{1,-1}}{1 + \rho_t}.$$

- In general, $\rho_c \neq \rho_t$. More generally, for cohort $\bar{c}$, the benchmark cohort, $\rho_{\bar{c}}$ is the IRR that solves

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c}} - Y^d_{a,\bar{c}}\right)}{(1 + \rho_{\bar{c}})^a} = 0.$$

- Cross section in year $t = \bar{c}$ produces the equation

$$\sum_{a=0}^{A} \frac{\left(Y^{h}_{a,\bar{c}-a} - Y_{a,\bar{c}-}a^{d}\right)}{(1+\rho_t)^{a}} = 0,$$

where ρ_t is the root.

- If growth rates across cohorts are benchmarked against $\bar{c}$, we obtain

$$\begin{aligned} \sum_{a=0}^{A} \frac{\left(Y^{h}_{a,\bar{c}} - Y^{d}_{a,\bar{c}}\right)(1+g)^{-a}}{(1+\rho_t)^{a}} &= 0 \\ \sum_{a=0}^{A} \frac{\left(Y^{h}_{a,\bar{c}} - Y^{d}_{a,\bar{c}}\right)}{\left[(1+\rho_t)(1+g)\right]^{a}} &= 0, \end{aligned}$$

so clearly $\rho_t < \rho_c$.

- Suppose that there are no cohort effects but that there are smooth time effects, say, $1+\varphi$.

- Then the cohort rate of return is calculated as the root of the following equation in which the choice of a cohort $\bar{c}$ as a benchmark is innocuous:

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^{h} - Y_{a,\bar{c}}^{d}\right)(1+\varphi)^{a}}{(1+\rho_{\bar{c}})^{a}} = 0$$

- The cross-section rate at time $t=\bar{c}$ is

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^{h} - Y_{a,\bar{c}}^{d}\right)}{(1+\rho_{t})^{a}} = 0, \qquad t=\bar{c},$$

where clearly if $\varphi > 0$, then $\rho_{\bar{c}} > \rho_{t}$.

- Better notation — distinguish outcomes at age a, cohort c, period t:

$$
\begin{aligned}
& Y^h_{a,c,t};\ Y^d_{a,c,t} \\
\Delta^{h,d}_{a,c,t} &= Y^h_{a,c,t} - Y^d_{a,c,t}.
\end{aligned}
$$

- No cohort effects means $Y^j_{a,c,t} = Y^j_{a,-,t}\ \forall c$. "–" sets the argument to a constant.

Pure Time Effects

- Take cohort $c = 0$ at time t:

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,0,t+a} - Y^d_{a,0,t+a}\right)}{(1+\rho_c)^a} = 0$$

- Cross section at $t = 0$ for $c = 0$:

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,-a,t} - Y^d_{a,-a,t}\right)}{(1+\rho_t)^a} = 0, \qquad t = 0$$

- No time effects means $Y^j_{a,c,t} = Y^j_{a,c,-}\ \forall t$.

- A model with pure cohort effects and no time effects writes, for cohort $\bar{c}$,

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c},-} - Y^d_{a,\bar{c},-}\right)}{\left(1+\rho_{\bar{c}}\right)^a} = 0.$$

- This defines a cohort rate of return.

- The cross-section at time $t = \bar{c}$ writes

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c},\bar{c}+a} - Y^d_{a,\bar{c},\bar{c}+a}\right)\left(1+g\right)^{\bar{c}}}{\left(1+\rho_{\bar{c}}\right)^a} = 0.$$

- So if $g > 0$, then $\rho_{\bar{c}} > \rho_t$ $(t = \bar{c})$.

- A model with pure time effects $(1+\varphi)$ writes, for time $t=\bar{c}$, the cohort return for entry cohort $\bar{c}$ as

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c},\bar{c}+a} - Y^d_{a,\bar{c},\bar{c}+a}\right)(1+g)^{\bar{c}}}{(1+\rho_{\bar{c}})^a} = 0 text.$$

- Benchmarking on the $c=0$ cohort,

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c},\bar{c}} - Y^d_{a,\bar{c},\bar{c}}\right)(1+\varphi)^a (1+g)^{\bar{c}}}{(1+\rho_{\bar{c}})^a} = 0.$$

- The cross-section return at time $\bar{c}$ is

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c}-a,\bar{c}} - Y^d_{a,\bar{c}-a,\bar{c}}\right)}{\left(1+\rho_t\right)^a} = 0,$$

where $Y^h_{a,\bar{c}-a,\bar{c}} = Y^h_{a,c^*,\bar{c}}$ for all c^*, $t = \bar{c}$, if there are only pure time effects.

- Suppose we have both time and cohort effects. Then we have that the cross-section is

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c}-a,\bar{c}} - Y^d_{a,\bar{c}-a,\bar{c}}\right)}{\left(1+\rho_t\right)^a} = 0.$$

- These can be written at time $t = \bar{c}$ as

$$\sum_{a=0}^{A} \frac{\left(Y^h_{a,\bar{c},\bar{c}} - Y^d_{a,\bar{c},\bar{c}}\right)\left(1+g\right)^{\bar{c}-a}}{\left(1+\rho_t\right)^a} = 0.$$

- Thus, if the cohort rate $\left(1+g\right)^{\bar{c}-a} = \left(1+\varphi\right)^a \left(1+g\right)^{\bar{c}}$ for all $\bar{c}$, we can get the result.

- This requires that

$$1 + g = \frac{1}{1+\varphi} \Rightarrow g = \frac{-\varphi}{1+\varphi}.$$

- This seems to characterize the IRR for high school vs. dropouts. Cohort growth rate factor is the inverse of the time rate.

How do Cross-sectional IRR Estimates Compare with Cohort-based Estimates?

- If one is interested in estimating the conventional rates of return governing school investment decisions, then whether to use cross-section or cohort-based estimates depends on the extent to which individuals are able to forecast future changes in wages and skill prices.
- We next turn to a review of the recent instrument-based "rate of return" literature.

Part VII

Accounting for the Endogeneity of Schooling

Accounting for the endogeneity of schooling

- Much of the CPS-Census literature on the returns to schooling ignores the choice of schooling and its consequences for estimating "the rate of return".
- It ignores uncertainty.
- It is static and ignores the dynamics of schooling choices and the sequential revelation of uncertainty.
- It also ignores ability bias.

Accounting for the endogeneity of schooling

- Economists since C. Reinhold Noyes (1945) in his comment on Friedman and Kuznets (1945) have raised the specter of ability bias, noting that the estimated return to schooling may largely be a return to ability that would arise independently of schooling.
- Griliches (1977) and Willis (1986) summarize estimates from the conventional literature on ability bias.
- For the past 30 years, labor economists have been in pursuit of good instruments to estimate "the rate of return" to schooling, usually interpreted as a Mincer coefficient.

Accounting for the endogeneity of schooling

- However, the previous sections show that, for many reasons, the Mincer coefficient is not informative on the true rate of return to schooling, and therefore is not the appropriate theoretical construct to gauge educational policy.
- Card (1999) is a useful reference for empirical estimates from instrumental variable models.

Accounting for the endogeneity of schooling

- Even abstracting from the issues raised by the sequential updating of information, and the distinction between *ex ante* and *ex post* returns to schooling, which we discuss further below, there is the additional issue that returns, however defined, vary among persons.
- A random coefficients model of the economic return to schooling has been an integral part of the human capital literature since the papers by Becker and Chiswick (1966), Chiswick (1974), Chiswick and Mincer (1972) and Mincer (1974).

Accounting for the endogeneity of schooling

- In its most stripped-down form and ignoring work experience terms, the Mincer model writes log earnings for person i with schooling level S_i as

$$\ln y_i = \alpha_i + \rho_i S_i, \tag{14}$$

  where the "rate of return" ρ_i varies among persons as does the intercept, α_i.

- For the purposes of this discussion think of y_i as an annualized flow of lifetime earnings.
- Unless the only costs of schooling are earnings foregone, and markets are perfect, ρ_i is a percentage growth rate in earnings with schooling and not a rate of return to schooling.

Accounting for the endogeneity of schooling

- Let $\alpha_i = \bar{\alpha} + \varepsilon_{\alpha_i}$ and $\rho_i = \bar{\rho} + \varepsilon_{\rho_i}$ where $\bar{\alpha}$ and $\bar{\rho}$ are the means of α_i and ρ_i.
- Thus the means of ε_{α_i} and ε_{ρ_i} are zero.
- Earnings equation (14) can be written as

$$\ln y_i = \bar{\alpha} + \bar{\rho} S_i + \{\varepsilon_{\alpha_i} + \varepsilon_{\rho_i} S_i\}. \qquad (15)$$

- Equations (14) and (15) are the basis for a human capital analysis of wage inequality in which the variance of log earnings is decomposed into components due to the variance in S_i and components due to the variation in the growth rate of earnings with schooling (the variance in $\bar{\rho}$), the mean growth rate across regions or time ($\bar{\rho}$), and mean schooling levels ($\bar{S}$).
- See, e.g. Mincer, 1974, and Willis, 1986.

Accounting for the endogeneity of schooling

- Given that the growth rate ρ_i is a random variable, it has a distribution that can be studied using the methods surveyed below.
- Following the representative agent tradition in economics, it has become conventional to summarize the distribution of growth rates by the mean, although many other summary measures of the distribution are possible.
- For the prototypical distribution of ρ_i, the conventional measure is the "average growth rate" $E(\rho_i)$ or $E(\rho_i|X)$, where the latter conditions on X, the observed characteristics of individuals.
- Other means are possible such as the mean growth rates for persons who attain a given level of schooling.

Accounting for the endogeneity of schooling

- The original Mincer model assumed that the growth rate of earnings with schooling, ρ_i, is uncorrelated with or is independent of S_i.
- This assumption is convenient but is not implied by economic theory.
- It is plausible that the growth rate of earnings with schooling declines with the level of schooling.

Accounting for the endogeneity of schooling

- It is also plausible that there are unmeasured ability or motivational factors that affect the growth rate of earnings with schooling and are also correlated with the level of schooling.
- Rosen (1977) discusses this problem in some detail within the context of hedonic models of schooling and earnings.
- A similar problem arises in analyses of the impact of unionism on relative wages and is discussed in Lewis (1963).

Accounting for the endogeneity of schooling

- Allowing for correlated random coefficients (so S_i is correlated with ε_{ρ_i}) raises substantial problems that are just beginning to be addressed in a systematic fashion in the recent literature.
- Here, we discuss recent developments starting with Card's (1999) random coefficient model of the growth rate of earnings with schooling, a model that is derived from economic theory and is based on the analysis of Becker's model by Rosen (1977).
- We consider conditions under which it is possible to estimate the mean effect of schooling and the distribution of returns in his model.
- The next section considers the more general and recent analysis of Carneiro, Heckman, and Vytlacil (2005).

Accounting for the endogeneity of schooling

- In Card's (1999, 2001) model, the preferences of a person over income (y) and schooling (S) are

$$U(y, S) = \ln y\,(S) - \varphi(S) \qquad \varphi'(S) > 0$$

and

$$\varphi''(S) > 0.$$

- The schooling-earnings relationship is $y = g(S)$.
- This is a hedonic model of schooling, where $g(S)$ reveals how schooling is priced out in the labor market.

Accounting for the endogeneity of schooling

- This specification is written in terms of annualized earnings and abstracts from work experience.
- It assumes perfect certainty and abstracts from the sequential resolution of uncertainty that is central to the modern literature.
- In this formulation, discounting of future earnings is kept implicit.
- The first order condition for optimal determination of schooling is

$$\frac{g'(S)}{g(S)} = \varphi'(S). \tag{16}$$

Accounting for the endogeneity of schooling

- The term $\frac{g'(s)}{g(s)}$ is the percentage change of earnings with schooling or the "growth rate" at level s.
- Card's model reproduces Rosen's (1977) model if r is the common interest rate at which agents can freely lend or borrow and if the only costs are S years of foregone earnings.
- In Rosen's setup, an agent with an infinite lifetime maximizes $\frac{1}{r}e^{-rS}g(S)$ so $\varphi(S) = rS + \ln r$, and $\frac{g'(S)}{g(S)} = r$.

Accounting for the endogeneity of schooling

- Linearizing the model, we obtain

$$\begin{aligned} \frac{g'(S_i)}{g(S_i)} &= \beta_i(S_i) = \rho_i - k_1 S_i, \qquad k_1 \geq 0, \\ \varphi'(S_i) &= \delta_i(S_i) = r_i + k_2 S_i, \qquad k_2 \geq 0. \end{aligned}$$

- Substituting these expressions into the first order condition (16), we obtain that the optimal level of schooling is $S_i = \dfrac{(\rho_i - r_i)}{k}$, where $k = k_1 + k_2$.
- Observe that if both the growth rate and the returns are independent of S_i, $(k_1 = 0, k_2 = 0)$, then $k = 0$ and if $\rho_i = r_i$, there is no determinate level of schooling at the individual level.
- This is the original Mincer (1958) model.

Accounting for the endogeneity of schooling

- One source of heterogeneity among persons in the model is ρ_i, the way S_i is transformed into earnings.
- School quality may operate through the ρ_i for example, as in Behrman and Birdsall (1983), and ρ_i may also differ due to inherent ability differences.
- A second source of heterogeneity is r_i, the "opportunity cost" (cost of schooling) or "cost of funds."
- Higher ability leads to higher levels of schooling.
- Higher costs of schooling results in lower levels of schooling.

Accounting for the endogeneity of schooling

- We integrate the first order condition (16) to obtain the following hedonic model of earnings,

$$\ln y_i = \alpha_i + \rho_i S_i - \frac{1}{2} k_1 S_i^2. \tag{17}$$

- To achieve the familiar looking Mincer equation, assume $k_1 = 0$.
- This assumption rules out diminishing "returns" to schooling in terms of years of schooling.

Accounting for the endogeneity of schooling

- Even under this assumption, ρ_i is the percentage growth rate in earnings with schooling, but is not in general an internal rate of return to schooling.
- It would be a rate of return if there were no direct costs of schooling and everyone faces a constant borrowing rate.
- This is a version of the Mincer (1958) model, where $k_2 = 0$, and r_i is constant for everyone but not necessarily the same constant.

Accounting for the endogeneity of schooling

- If $\rho_i > r_i$, person i takes the maximum amount of schooling.
- If $\rho_i < r_i$, person i takes no schooling and if $\rho_i = r_i$, schooling is indeterminate.
- In the Card model, ρ_i is the person-specific growth rate of earnings and overstates the true rate of return if there are direct and psychic costs of schooling.

Accounting for the endogeneity of schooling

- This simple model is useful in showing the sources of endogeneity in the schooling earnings model.
- Since schooling depends on ρ_i and r_i, any covariance between $\rho_i - r_i$ (in the schooling equation) and ρ_i (in the earnings function) produces a random coefficient model.
- Least squares will not estimate the mean growth rate of earnings with schooling unless, $\text{Cov}(\rho_i, \rho_i - r_i) = 0$.

Accounting for the endogeneity of schooling

- Dropping the i subscripts, the conditional expectation of log earnings given s is

$$E(\ln y \mid S = s) = E(\alpha \mid S = s) + E(\rho \mid S = s)\, s.$$

- The first term produces the conventional ability bias if there is any dependence between s and raw ability α.
- Raw ability is the contribution to earnings independent of the schooling level attained.
- The second term arises from sorting on returns to schooling that occurs when people make schooling decisions on the basis of growth rates of earnings with schooling.
- It is an effect that depends on the level of schooling attained.

Accounting for the endogeneity of schooling

- In his Woytinsky Lecture (1967), Becker points out the possibility that many able people may not attend school if ability (ρ_i) is positively correlated with the cost of funds (r_i).
- A meritocratic society would eliminate this positive correlation and might aim to make it negative.
- Schooling is positively correlated with the growth rate (ρ_i) if $\text{Cov}(\rho_i, \rho_i - r_i) > 0$.
- If the costs of schooling are sufficiently positively correlated with the growth rate, then schooling is negatively correlated with the growth rate.

Accounting for the endogeneity of schooling

- Observe that S_i does not directly depend on the random intercept α_i.
- Of course, α_i may be statistically dependent on (ρ_i, r_i).
- In the context of Card's model, we consider conditions under which one can identify $\bar{\rho}$, the mean growth rate of earnings in the population as well as the full distribution of ρ.
- First we consider the case where the marginal cost of funds, r_i, is observed and consider other cases in the following subsections.

Estimating the mean growth rate of earnings when r_i is observed

- A huge industry surveyed in Card (1999) seeks to estimate the mean growth rate in earnings, $E(\rho_i)$, calling it the "causal effect" of schooling.
- For reasons discussed earlier in this chapter, in general, it is not an internal rate of return.
- However, it is one of the ingredients used in calculating the rate of return as we develop further below.
- The "causal effect" may also be of interest in its own right if the goal is to estimate pricing equations for labor market characteristics.
- We discuss some simple approaches for identifying causal effects before turning to a more systematic analysis below.

Estimating the mean growth rate of earnings when r_i is observed

- Suppose that the cost of schooling, r_i, is measured by the economist.
- Use the notation "$\perp\!\!\!\perp$" to denote statistical independence.
- Assume
$$r_i \perp\!\!\!\perp (\rho_i, \alpha_i).$$
- This assumption rules out any relationship between the cost of funds (r_i) and raw ability (α_i) with the growth rate of earnings with schooling.

Estimating the mean growth rate of earnings when r_i is observed

- For example, it rules out fellowships based on ability.
- We make this assumption to illustrate some ideas and not because of its realism.
- Observing r_i implies that we observe ρ_i up to an additive constant.
- Recall that $S_i = \dfrac{(\rho_i - r_i)}{k}$, so that $\rho_i = r_i + kS_i$ and $\bar{\rho} = E(\rho_i) = \bar{r} + kE(S_i)$.

Estimating the mean growth rate of earnings when r_i is observed

- r_i is a valid instrument for S_i under the assumption that $k_1 = 0$.
- It is independent of α_i, ρ_i (and hence $\varepsilon_{\alpha i}$, $\varepsilon_{\rho i}$) and is correlated with S_i because S_i depends on r_i.

Estimating the mean growth rate of earnings when r_i is observed

- Form

$$\frac{\text{Cov}(\ln y_i, r_i)}{\text{Cov}(S_i, r_i)}$$
$$= \frac{E\left\{(r_i - \bar{r})\left[(\alpha_i - \bar{\alpha}) + (\rho_i - \bar{\rho})(S_i - \bar{S}) + \bar{\rho}S_i + \rho_i\bar{S} - \bar{\rho}\bar{S})\right]\right\}}{E\left\{\left[\frac{\rho_i - r_i}{k}\right][r_i - \bar{r}]\right\}}$$
$$= \frac{\frac{1}{k}E[(\Delta r)(\Delta\rho)(\Delta\rho - \Delta r)] - \frac{\bar{\rho}}{k}\sigma_r^2}{-\frac{\sigma_r^2}{k}},$$

where $\Delta X = X - E(X)$.

Estimating the mean growth rate of earnings when r_i is observed

- As a consequence of the assumed independence between r_i and (ρ_i, α_i), $E[(\Delta r)(\Delta\rho)^2] = 0$ and $E[(\Delta r)^2 \Delta\rho] = 0$, so

$$\left[\frac{\mathrm{Cov}(\ln y_i, r_i)}{\mathrm{Cov}(S_i, r_i)}\right] = \bar{\rho}.$$

Estimating the mean growth rate of earnings when r_i is observed

- Observe that $\bar{\rho}$ is not identified by this argument if $\rho_i \not\perp\!\!\!\perp r_i$ (so the mean growth rate of earnings depends on the cost of schooling).
- In that case, $E[(\Delta r)(\Delta\rho)^2] \neq 0$ and $E[(\Delta r)^2(\Delta\rho)] \neq 0$.
- If r_i is known and $r_i = L_i\gamma + M_i$, where the L_i are observed variables that explain r_i and $E(M_i \mid L_i) = 0$, then γ is identified, provided a rank condition for instrumental variables is satisfied.
- We require that L_i be at least mean independent of (M_i, ρ_i, α_i).
- From the schooling equation we can write $S_i = (\rho_i - L_i\gamma - M_i)/k$ and k is identified since we know γ.

Estimating the mean growth rate of earnings when r_i is observed

- Observe that we can estimate the distribution of ρ_i since $\rho_i = r_i + kS_i$, k is identified and (r_i, S_i) are known.
- This is true even if there are no instruments L, $(\gamma = 0)$, provided that $r_i \perp\!\!\!\perp (\rho_i, \alpha_i)$.
- With the instruments that satisfy at least the mean independence condition, we can allow $r_i \not\perp\!\!\!\perp \rho_i$ and all parameters and distributions are still identified.
- The model is fully identified provided r_i is observed and $L_i \perp\!\!\!\perp (M_i, \rho_i, \alpha_i)$.
- Thus, we can identify the mean return to schooling.

Estimating the mean growth rate when r_i is not observed

- If r_i is not observed and so cannot be used as an instrument, but we know that r_i depends on observed factors L_i and M_i, $r_i = L_i\gamma + M_i$ and $L_i \perp\!\!\!\perp (M_i, \alpha_i, \rho_i)$, then our analysis carries over and the mean growth rate $\bar{\rho}$ is identified.
- Recall that $\ln y_i = \alpha_i + \bar{\rho}S_i + (\rho_i - \bar{\rho})S_i$.
- Substitute for S_i to get an expression of y_i in terms of L_i, $\ln y_i = \alpha_i + \rho_i(\rho_i - L_i\gamma - M_i)/k$.

Estimating the mean growth rate when r_i is not observed

- We obtain the vector moment equations:

$$\mathrm{Cov}(\ln y_i, L_i) = \bar{\rho}\ \mathrm{Cov}(S_i, L_i),$$

  so $\overline{\rho}$ is identified from the population moments because the covariances on both sides are available.

- Partition $\gamma = (\gamma_0, \gamma_1)$, where γ_0 is the intercept and γ_1 is the vector of slope coefficients.

- From the schooling equation, we obtain

$$\begin{aligned} S_i &= \frac{\rho_i - L_i\gamma_1 - M_i}{k} - \frac{\gamma_0}{k} \\ &= -L_i\frac{\gamma_1}{k} + \frac{\rho_i - M_i}{k} - \frac{\gamma_0}{k}. \end{aligned}$$

Estimating the mean growth rate when r_i is not observed

- We can identify γ_1/k from the schooling equation, as well as the mean growth rate $\bar{\rho}$.
- However, we cannot identify the distribution of ρ_i or r_i unless further assumptions are invoked.
- We also cannot separately identify γ_0, γ_1 or k.
- Heckman and Vytlacil (1998) show how to define and identify a version of "treatment on the treated" for growth rates in the Becker-Card-Rosen model.

Adding selection bias

- Selection bias can arise in two distinct ways in the Becker-Card-Rosen model: through dependence between α_i and ρ_i and through dependence between α_i and r_i.
- Allowing for selection bias,

$$\begin{aligned} E(\ln y_i \mid S_i) &= E(\alpha_i \mid S_i) + E(\rho_i S_i \mid S_i) \\ &= E(\alpha_i \mid S_i) + E(\rho_i \mid S_i) S_i. \end{aligned}$$

- If there is an L_i that affects r_i but not ρ_i and is independent of (α_i, M_i), i.e., $L_i \perp\!\!\!\perp (\alpha_i, \rho_i, M_i)$, and $E(r_i|L_i)$ is a nontrivial function of L_i, in the special case of a linear schooling model,

$$\begin{aligned} E\left(\ln y_i \mid L_i\right) &= E(\alpha_i \mid L_i) + E(\rho_i S_i \mid L_i) \\ &= \eta + \bar{\rho} E(S_i \mid L_i). \end{aligned}$$

Adding selection bias

- Since we can identify $E(S_i \mid L_i)$ we can identify $\bar{\rho}$.
- Thus, under the stated conditions, the instrumental variable (IV) method identifies $\bar{\rho}$ when there is selection bias.
- In a more general nonparametric case for the schooling equation, which we develop in the next section of this chapter, this argument breaks down and $\bar{\rho}$ is not identified when ρ_i determines S_i in a general way.

Adding selection bias

- The sensitivity of the *IV* method to assumptions about special features of Card's model is a simple demonstration of the fragility of the method.
- We return to this model later and use it to motivate recent developments in the literature on identifying information available to agents when they make their schooling decisions.

Summary

- Card's version of the Becker (1967)-Rosen (1977) model is a useful introduction to the modern literature on heterogeneous "returns to schooling.
- " ρ_i is, in general, a person-specific growth rate of log earnings with schooling and not a rate of return.
- There is a distribution of ρ_i and no scalar measure is an adequate summary of this distribution.
- Recent developments in this literature, to which we now turn, demonstrate that standard instrumental variable methods are blunt tools for recovering economically interpretable parameters.

Part VIII

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- To understand what *IV* estimates in a more general setting, this section analyzes a simple version of (14) in which there are only two levels of schooling.
- Our discussion can be generalized (see Heckman and Vytlacil, 2005 and Heckman, Urzua, and Vytlacil, 2006), but for purposes of exposition it is fruitful to focus on a two outcome model.
- It links *IV* to the analysis of Willis and Rosen (1979) and Willis (1986), who focus on a two outcome model of schooling in which the ρ_i of equation (17) varies in the population.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Recent research on instrumental variables in the correlated coefficient model establishes a close link between *IV* and the selection model (Heckman, 1976) that Willis and Rosen apply to obtain their estimates.
- As shown in Heckman, Urzua, and Vytlacil (2006), the contrast between *IV* and selection methods emphasized by Angrist and Krueger (1999) and echoed throughout the literature is not valid once the *IV* method for the correlated random coefficient model is correctly understood.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Because schooling is usually received in integer amounts, and most well posed models of schooling choice are based on nonlinear discrete choice frameworks, the simple Card model abstracts from key features of the schooling choice – earnings outcome model which can be captured in a simple way by a discrete outcome model.
- Heckman (1997) and Heckman and Vytlacil (1998) show how models of schooling that capture key features of economic theory are intrinsically nonlinear.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- In this model, the mean growth rate of earnings with schooling, $\bar{\rho}$, was assumed to be the parameter of interest without any good justification.
- While statisticians sometimes call such averages the "average causal effect" (ACE), there is no reason to focus on this parameter to the exclusion of other parameters that can be derived from the distribution of ρ_i.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Moreover, as we shall show in this section, the instrumental variable estimators set forth in the recent literature do not in general estimate ACE or any of the other standard treatment effects of schooling on earnings when schooling choices are discrete.
- They do not estimate rates of return to schooling, nor are they designed to.
- Instead, they estimate certain weighted averages of individual growth rates where the weights can sometimes be negative.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Following Heckman and Vytlacil (2000, 2005, 2007b), Heckman, Urzua, and Vytlacil (2006), and Carneiro, Heckman, and Vytlacil (2005), consider the following generalized Roy model of schooling and its "return.
- " A version of it is applied by Willis and Rosen to the problem of choice of college using tools developed in the econometrics of selection bias.
- Our analysis of this model links the modern *IV* literature to the classical selection literature.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Let Y_1 denote the present value of earnings from college.
- Y_0 is the present value of earnings from high school.
- There is a distribution of $G = Y_1 - Y_0$ and another distribution of $G - C$ in the population where C denotes the cost of schooling and G denotes earnings gains from college.
- No single number summarizes either distribution, although much of the literature focuses on one conditional mean or some other single number as the object of economic and econometric interest.

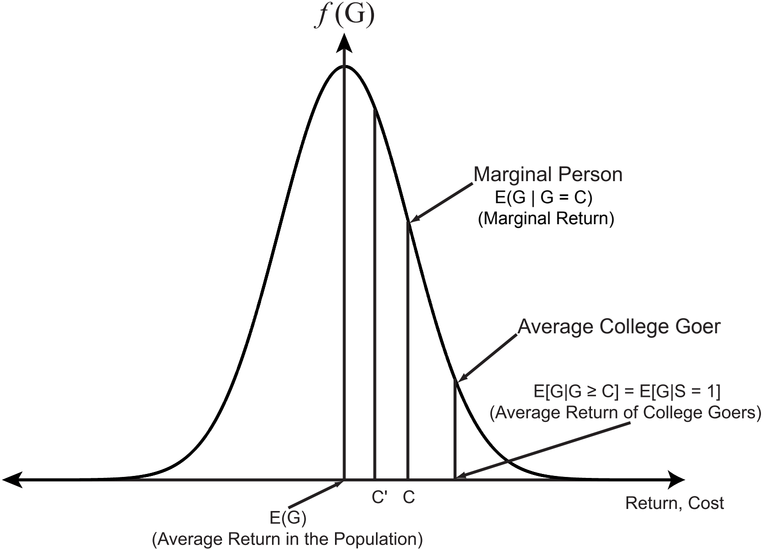
Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Attention has focused in recent years on *IV* estimates of the coefficient of schooling in a regression of log earnings on schooling.
- In special cases, *IV* can sometimes identify the mean growth rate in earnings ($E(\rho_i)$) which is usually not the same as the rate of return.
- But more generally, *IV* does not even identify this parameter.
- This section considers what *IV* estimates in general cases.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- If G varies in the population but everyone faces the same C, individuals decide to enroll in school $(S = 1)$ if $G - C > 0$.
- Figure 9 plots the hypothetical density of G in this example, $f(G)$, and also presents the cost that everyone faces, C.
- Individuals who have values of G to the right of C choose to enroll in school, while those to the left choose not to enroll.

Figure 9
Density of Absolute Returns



C' = Old Cost; C = New Cost

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- The gross gain for the individuals who choose to go to school, $E(G \mid G \geq C)$, is computed with respect to the normalized density of $f(G)$ that is to the right of C.
- The marginal return (the return for individuals at the margin) is exactly equal to C.
- Figure 9 presents both the average and the marginal return for this example.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Suppose that we want to estimate the effect on earnings of compulsory college attendance.
- Those individuals who are induced to enroll in school by this policy have G below C (they were not enrolled in school before the policy), and the average "return" for these individuals is $E(G|G \leq C)$.
- Alternatively, one might be interested in analyzing the effect of a tuition subsidy that changes the cost of attending school from C to C' for everyone in the economy.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Those individuals who are induced to enroll in school by this policy have G below C (they were not enrolled in school before the policy) and G above C' (they decide to enroll after the policy), and the average "return" for these individuals is $E(G|C' < G \leq C)$.
- One needs different parameters to evaluate each of these two different policies ($E\left(G|G \leq C\right)$ vs. $E(G|C' < G \leq C)$).

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Neither is estimated by the average growth rate, and hence by the *IV* method.
- In this example, the marginal entrant into college has a lower return than the average entrant, and the return for the average student is not the relevant return to evaluate either policy.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Standard estimates of the returns to schooling, such as the ones obtained using the method of least squares, as in the vast literature surveyed by Katz and Autor (1999), or using the method of instrumental variables, as surveyed by Card (1999), are not designed to produce either of the policy parameters just described.
- It is unusual in the recent literature on the "returns to schooling" for researchers estimating "the effect" of schooling to specify a policy or economic question of interest and address it directly.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Following Card (1999) and Angrist and Krueger (1999), many define the probability limits of instrumental variable estimators (*LATE*, defined below) as "the" return to schooling without stating what economic questions these statistical objects address.
- Different instruments define different parameters.
- These parameters answer different, implicitly defined, economic questions.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Moreover the commonly accepted interpretation of LATE – that it estimates the returns for those induced to change their schooling status by the change in the instrument – assumes that everyone responds to the instrument in the same direction (i.e., all increase their schooling or all decrease it).
- This is a strong assumption that rules out heterogeneity in the response of schooling choices to instruments.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- In this section we distinguish between policy parameters of interest, conventional evaluation parameters and standard estimates of the "returns to schooling."
- We show how these parameters answer different questions, and how we can recover each of them from the data.
- We illustrate the empirical importance of accounting for heterogeneity and the fragility of instruments even in an ideal data set with far richer instruments than are available in the widely used CPS or Census data analyzed in earlier sections of this survey.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- This section draws from Heckman and Vytlacil (2005, 2007b), Heckman, Urzua, and Vytlacil (2006) and Carneiro, Heckman, and Vytlacil (2005).
- They estimate the growth rate of earnings in schooling relevant for evaluating a particular education policy such as a tuition subsidy (in a partial equilibrium framework) and find that it is very different from the conventional program evaluation parameters usually defined in the literature, such as the "return to schooling" for the average person, or the "return to schooling" for the average student in college.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- It also differs from the estimates obtained by applying least squares or instrumental variables methods, the two methods most often used to estimate "returns" to schooling.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- We clarify the interpretation of what is usually labeled "ability bias" and "selection bias" in this literature.
- Standard intuitions break down in a model of heterogeneous returns.
- They can be very misleading when comparing *OLS* and *IV* estimates of the growth rates of earnings with respect to schooling (See Heckman and Vytlacil, 2005).

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Instrumental variables estimates of the "return to schooling" (really growth rates of earnings with schooling) are usually interpreted as estimating an average "return" to schooling for individuals induced to go to school by changes in the values of the instrument, following the *LATE* (local average treatment effect) interpretation of Imbens and Angrist (1994).
- Angrist and Krueger (1999) are ardent and influential proponents of this approach.
- We discuss the relationship of *LATE* to treatment effects and rates of return below.

Accounting Systematically for Heterogeneity in Returns to Schooling: What Does IV Estimate?

- Intuitions about ability bias break down in a particularly serious way if individuals have multiple skills and sort across schooling levels in such a way that the best individuals in one schooling level are the worst in the other, and vice versa.
- Heckman and Robb (1985, 1986) make the point that *IV* does not identify interpretable parameters in a selection model or a generalized Roy model.

The Generalized Roy Model of Schooling

- To focus the discussion and motivate the empirical literature, we consider a two outcome model. Heckman and Vytlacil (2005, 2007b) and Heckman, Urzua, and Vytlacil (2006) extend this discussion to ordered choice and general unordered choice models with multiple outcomes.

The Generalized Roy Model of Schooling

- From its inception, the modern literature on the "returns to schooling" has recognized that returns may vary across schooling levels and across persons of the same schooling level.
- The early literature was not clear about the sources of variation in returns.
- The Roy model (1951) and its extensions (see Heckman, 1976, 1979), as applied by Willis and Rosen (1979), gives a more precise notion of why returns vary and how they depend on S.

The Generalized Roy Model of Schooling

- In the generalized Roy framework, the potential outcomes associated with two different schooling levels are generated by two random variables (U_0, U_1):

$$\ln Y_0 = \alpha + U_0 \tag{18a}$$

$$\ln Y_1 = \alpha + \bar{\beta} + U_1 \tag{18b}$$

  where $E(U_0) = 0$ and $E(U_1) = 0$ so α $(= E(\ln Y_0))$ and $\alpha + \bar{\beta}$ $(= E(\ln Y_1))$ are the mean potential outcomes for $\ln Y_0$ and $\ln Y_1$ respectively.

- The common coefficient model assumes $U_0 = U_1$.

- We implicitly condition on X, the regressors determining potential outcomes.

The Generalized Roy Model of Schooling

- Let $C(Z)$ denote costs of schooling measured in proportional terms.
- The Z are the variables determining costs.
- The individual level "causal effect" of educational choice $S = 1$ is
$$\beta = \ln Y_1 - \ln Y_0 = \bar{\beta} + U_1 - U_0.$$
- In general, this is not a rate of return but a growth rate of earnings with schooling.
- There is a distribution of β in the population.

The Generalized Roy Model of Schooling

- Observed earnings are written in a “switching regression” form,

$$\begin{aligned}\ln Y &= S \ln Y_1 + (1 - S) \ln Y_0 \\ &= \alpha + \beta S + U_0 \\ &= \alpha + \bar{\beta} S + \{U_0 + S(U_1 - U_0)\}. \qquad (19)\end{aligned}$$

- Persons live once and we only observe them in one or the other education state (recall $S = 0$ or 1).
- This equation captures the literature on counterfactual states that was developed by Roy (1951).
- It is also a version of Quandt’s (1958, 1972) switching regression model.

The Generalized Roy Model of Schooling

- It is equivalent to the familiar semilog specification of the earnings-schooling equation popularized by Mincer (1974), given in equation (14), which in the current notation writes log earnings $\ln Y$ as a function of S,

$$\ln Y = \alpha + \bar{\beta} S + U, \tag{20}$$

  where $U = U_0 + S(U_1 - U_0)$.

- In terms of the notation used above, $U_0 = \varepsilon_\alpha$, $U_1 - U_0 = \varepsilon_\rho$.

The Generalized Roy Model of Schooling

- In the generalized Roy framework, the choice of schooling is explicitly modeled.
- In its simplest form

$$S = \begin{cases} 1 \text{ if } \ln Y_1 - \ln Y_0 \geq C \iff \beta \geq C \\ 0 \text{ otherwise.} \end{cases} \tag{21}$$

- If agents know or can partially predict β at the time they make their schooling decisions, there is dependence between β and S in equation (19).
- This produces the "correlated random coefficient model" that is often applied to general versions of (19).
- Decision rules similar to (21) characterize many other economic choices.

The Generalized Roy Model of Schooling

- The conventional approach to estimating selection models postulates normality of (U_0, U_1) in equations (18a) and (18b), writes $\bar{\beta}$ and α as linear functions of X and postulates independence between X and (U_0, U_1).
- Parallel normality and independence assumptions are made for the unobservables and observables in selection equation (21).
- From estimates of the structural model, it is possible to answer a variety of economic questions and to construct the various treatment parameters and distributions of treatment parameters.
- However in recent years these assumptions have often been viewed as unacceptably strong by empirical labor economists (See, e.g. Angrist and Krueger, 1999).

The Generalized Roy Model of Schooling

- A major advance in the recent literature in econometrics is the development of frameworks that relax conventional linearity, normality and separability assumptions to estimate various economic parameters.
- Heckman and Vytlacil (2000, 2005, 2007b) develop a framework for estimating rates of return to schooling (mean growth rates of earnings with schooling) that do not depend on normality, independence of the conditioning variables with the regressors, separability or linearity of the estimating equations.

The Generalized Roy Model of Schooling

- Their work unites *IV* and selection models and presents a new local *IV* approach as a way to estimate selection models.
- Heckman, Urzua, and Vytlacil (2006) and Heckman and Vytlacil (2005) present extensive discussions of the relationship between the two approaches.

The Generalized Roy Model of Schooling

- Heckman and Vytlacil work with general nonseparable models,

$$\ln Y_1 = \mu_1(X, U_1) \text{ and } \ln Y_0 = \mu_0(X, U_0). \tag{22}$$

- The growth rate of earnings due to schooling is $\ln Y_1 - \ln Y_0 = \beta = \mu_1(X, U_1) - \mu_0(X, U_0)$, which is a general nonseparable function of (U_1, U_0).
- It is not assumed that $X \perp\!\!\!\perp (U_0, U_1)$, so X may be correlated with the unobservables in potential outcomes.

The Generalized Roy Model of Schooling

- As demonstrated by Heckman and Vytlacil (2000, 2005, 2007a), one needs exogeneity of X only if one is seeking to make out of sample projections.
- Like virtually the entire microeconomic literature, they ignore any general equilibrium effects of policies on Y_1, Y_0 or β.

The Generalized Roy Model of Schooling

- A latent variable model that captures decision rule (21) in a general way is:

$$\begin{aligned} S^* &= \mu_S(Z) - U_S \\ S &= 1 \text{ if } S^* \geq 0. \end{aligned} \tag{23}$$

- In this notation the Z can include all of the variables in the outcome equations plus the variables in the cost function which are a source of exclusion restrictions.
- $\mu_S(Z)$ is a general function of the observables where U_S is an unobservable arising from Y_1, Y_0 and C.
- A person goes to school $(S = 1)$ if $S^* \geq 0$.
- Otherwise $S = 0$.

The Generalized Roy Model of Schooling

- In this notation, (Z, X) are observed and (U_1, U_0, U_S) are unobserved.
- U_S may depend on U_1 and U_0 and the unobservables in C in a general way.
- The Z vector may include some or all of the components of X.

The Generalized Roy Model of Schooling

- The separability between Z and U_S in (23) plays a crucial role in the entire modern instrumental variables literature based on *LATE* and its extensions.
- It produces the "monotonicity" or "uniformity" condition of Imbens and Angrist (1994).
- Without the separability, changes in the instruments in Z can induce two-way flows into and out of treatment and cause *IV* to break down as a method for estimating treatment effects.
- See Heckman and Vytlacil (2005, 2007b) and Heckman, Urzua, and Vytlacil (2006).

The Generalized Roy Model of Schooling

- Those authors explore the consequences of a simple random coefficient choice model $\mu_S(Z) = Z\gamma$, where γ is a random coefficient that is statistically independent of Z and U_S.
- If γ can assume both positive and negative values, then monotonicity can be violated.
- But it can also be violated if γ is a non-negative random vector since different components of γ would differ across persons experiencing the same change in Z.

The Generalized Roy Model of Schooling

- The separability that is required to justify (23) and that underlies the entire *LATE*-based literature cannot be justified in many choice-theoretic models of schooling including dynamic discrete choice models.
- The Bellman equation producing the value function in multiperiod settings generates nonseparability between observables and unobservables in the choice equation in early stage decisions even if final stage choices are separable in those variables (see Cunha, Heckman, and Urzua, 2006).
- The method of *IV* applied to a heterogeneous outcome model is fundamentally asymmetric.

The Generalized Roy Model of Schooling

- It allows for heterogeneity in responses to schooling (i.e., it imposes no restrictions on β which may be general random variables).
- At the same time, it restricts the heterogeneity in responses of schooling choices to changes in Z.
- Consider the special case $\mu_s(Z) = Z\gamma$.

The Generalized Roy Model of Schooling

- The "monotonicity conditions" invoked in the recent literature to justify *IV* as estimating the return to schooling for people induced into schooling by a change in instrument rules out a random coefficient model for γ except for very special cases.
- Thus it does not allow for heterogeneity in choices, but it allows for heterogeneity in outcomes.
- See the discussion in Heckman and Vytlacil (2005, 2007b) and Heckman, Urzua, and Vytlacil (2006).

The Generalized Roy Model of Schooling

- Heckman and Vytlacil (2001a, 2007b) assume that
  1. Z has some variables that shift $\mu_S(Z)$ given X (the other variables) – an exclusion condition that is standard in the *IV* literature;
  2. The unobservables (U_0, U_1, U_S) are independent of Z given X (a standard instrumental variables condition) and
  3. $0 < \Pr(S = 1 \mid X) < 1$, so in large samples there are some people who have $S = 1$ and some who have $S = 0$, so comparisons between treated and untreated persons can be made for those values of X.
- They make additional mild regularity assumptions.

The Generalized Roy Model of Schooling

- Under these conditions it is possible to interpret *IV* as a weighted average of willingness to pay measures called the marginal treatment effect (*MTE*).
- A version of this treatment effect was introduced into the econometrics literature by Björklund and Moffitt (1987) for a linear-in-parameters model.

The Generalized Roy Model of Schooling

- Let $P(z)$ be the probability of receiving schooling level 1, $S = 1$ conditional on $Z = z$, $P(z) \equiv \Pr(S = 1 \mid Z = z) = F_{U_S}(\mu_S(z))$ where F_{U_S} is the distribution of U_S.
- Without loss of generality, one may write $U_S \sim$ Unif[0,1] so $\mu_S(z) = P(z)$.
- If $S^* = \nu(Z) - V_S$, and V_S is a continuous random variable, one can always reparameterize the model using simple transformation of variable rules so $\mu_S(Z) = F_{V_S}(\nu(Z))$, where F_{V_S} is the distribution of V and $U_S = F_{V_S}(V_S)$.
- The propensity score $P(z)$ is a monotonic transformation of the mean utility of attending school and we will refer to it as the mean utility.

The Generalized Roy Model of Schooling

- When β varies in the population, the growth rate of earnings with schooling is a random variable and there is a distribution of "causal effects."
- There are various ways to summarize this distribution and, in general, no single statistic will capture all aspects of the distribution.

The Generalized Roy Model of Schooling

- Many summary measures of the distribution of β are used in the recent literature.
- Among them are

$$\begin{aligned} E(\beta \mid X = x) &= E(\ln Y_1 - \ln Y_0 \mid X = x) \\ &= \bar{\beta}(x) \end{aligned}$$

the return to the population average person given characteristics $X = x$.

- This quantity is sometimes called "the" causal effect of S.

The Generalized Roy Model of Schooling

- Others report the "return" for those who attend school:

$$
\begin{aligned}
E\left(\beta \mid S=1, X=x\right) &= E(\ln Y_1 - \ln Y_0 \mid S=1, X=x) \\
&= \bar{\beta}(x) + E(U_1 - U_0 \mid S=1, X=x).
\end{aligned}
$$

- This is the parameter emphasized by Willis and Rosen (1979) where $E(U_1 - U_0 \mid S=1, X=x)$ is the sorting gain—how people who take $S=1$ differ from randomly sampled persons.

The Generalized Roy Model of Schooling

- Another parameter is "the return" for those who are currently not going to school:

$$
\begin{aligned}
E(\beta \mid S=0, X=x) &= E(\ln Y_1 - \ln Y_0 \mid S=0, X=x) \\
&= \bar{\beta}(x) + E(U_1 - U_0 \mid S=0, X=x).
\end{aligned}
$$

- Angrist and Krueger (1991) and Meghir and Palme (2001) estimate this parameter and we discuss it further below. In addition to these "effects" is the effect for persons indifferent between the two levels of schooling, which in the simple Roy model without costs ($C=0$) is $E(\ln Y_1 - \ln Y_0 \mid \ln Y_1 - \ln Y_0 = 0) = 0$.

The Generalized Roy Model of Schooling

- Depending on the conditioning sets and the summary statistics desired, a variety of "causal effects" can be defined. Different causal effects answer different economic questions. As noted by Heckman and Robb (1986), Heckman (1997), and Heckman and Vytlacil (2005, 2007b), under one of two conditions,

I

$$U_1 = U_0 \qquad \text{(common effect model).}$$

The Generalized Roy Model of Schooling

- Or more generally,

II

$\Pr(S = 1 \mid X = x, \beta) = \Pr(S = 1 \mid X)$ (conditional on X, β does not affect choices).

- all of the mean treatment effects conditional on X collapse to the same parameter.
- The second condition is the one implicitly used by Mincer (1974).

The Generalized Roy Model of Schooling

- It assumes that schooling decisions are not made on the basis of any component of the growth rate β.
- If neither condition is satisfied, there are many candidates for the title of causal effect.
- This ambiguity has produced considerable confusion in the empirical literature as different analysts use different definitions in reporting empirical results and many of the estimates are not strictly comparable.

The Generalized Roy Model of Schooling

- Which, if any, of these effects should be designated as "the" causal effect? We have already noted that conventional "causal effects" are not estimates of a marginal internal rate of return, but instead are estimates of some average growth rate of earnings with schooling.
- Instead of hoping that a treatment effect or estimator answers an interesting economic question, a better approach is to state an economic question and find the answer to it.
- This obvious and traditional approach is not pursued in the recent literature.

The Generalized Roy Model of Schooling

- Heckman and Vytlacil (2001c, 2005, 2007b) develop this approach using a standard welfare framework.
- They introduce the notion of a policy relevant treatment effect.

- Aggregate per capita outcomes under one policy are compared with aggregate per capita outcomes under another.
- One of the policies may be no policy at all.
- For utility criterion $V(Y)$, a standard welfare analysis compares an alternative policy with a baseline policy.

The Generalized Roy Model of Schooling

- The Policy Relevant Treatment Effect (*PRTE*) is

$$E(V(Y) \mid \text{Alternative Policy}) - E(V(Y) \mid \text{Baseline Policy}). \tag{24}$$

- Adopting the common coefficient model, so $\beta = \bar{\beta}$, a log utility specification ($V(Y) = \ln Y$) and ignoring general equilibrium effects, where β is a constant, $\bar{\beta}$, the mean change in welfare is

$$\begin{aligned} E(\ln Y \mid \text{Alternative Policy}) - E(\ln Y \mid \text{Baseline Policy}) \\ = \bar{\beta}(\Delta P), \end{aligned}$$

  where (ΔP) is the change in the proportion of people induced to attend school by the policy.

- This can be defined conditional on $X = x$ or overall for the population.

The Generalized Roy Model of Schooling

- In terms of gains per capita to recipients, the effect is $\bar{\beta}$.
- This is also the mean change in log income if β is a random variable but independent of S if conditions I or II apply.
- In the general case, when agents partially anticipate β, and comparative advantage dictates schooling choices, none of the traditional treatment parameters plays the role of $\bar{\beta}$ in (24) or answers the stated economic question.

The Generalized Roy Model of Schooling

- Heckman and Vytlacil (2001c, 2005, 2007b) show how the policy relevant treatment effect can be represented as a weighted average of the MTE.
- The weights are given in Table 9b.
- See Heckman, Urzua, and Vytlacil (2006) for further examples.

Table 9a: Treatment Effects and Estimands as Weighted Averages of the Marginal Treatment Effect

$ATE(x) = \int_0^1 MTE(x, u_S) du_S$	(Average Treatment Effect)
$TT(x) = \int_0^1 MTE(x, u_S) h_{TT}(x, u_S) du_S$	(Treatment on the Treated)
$TUT(x) = \int_0^1 MTE(x, u_S) h_{TUT}(x, u_S) du_S$	(Treatment on the Untreated)
$PRTE(x) = \int_0^1 MTE(x, u_S) h_{PRT}(x, u_S) du_S$	(Policy Relevant Treatment Effect)
$\text{IV}(x) = \int_0^1 MTE(x, u_S) h_{IV}(x, u_S) du_S$	
$\text{OLS}(x) = \int_0^1 MTE(x, u_S) h_{OLS}(x, u_S) du_S$	

Source: Heckman and Vytlacil (2001a,b; 2005, 2006)

Table 9b: Weights*

$$h_{TT}(x,u_S) = \left[\int_{u_S}^{1} f(p \mid X=x)dp\right] \frac{1}{E(P|X=x)}$$

$$h_{TUT}(x,u_S) = \left[\int_{0}^{u_S} f(p \mid X=x)dp\right] \cdot \frac{1}{E((1-P)|X=x)}$$

$$h_{PRT}(x,u_S) = \left[\frac{F_{P^*,X}(u_S)-F_{P,X}(u_S)}{\Delta P}\right]$$

$$h_{IV}(x,u_S) = \left[\int_{u_S}^{1} (p-E(P \mid X=x))f(p \mid X=x)dp\right] \frac{1}{Var(P|X=x)}$$

$$h_{OLS} = \frac{E(U_1|X=x,U_S=u_S)h_1(x,u_D)-E(U_0|X=x,U_S=u_S)h_0(x,u_S)}{MTE(x,u_S)}$$

$$h_1(x,u_S) = \left[\int_{u_S}^{1} f(p \mid X=x)dp\right] \frac{1}{E(P|X=x)}$$

$$h_0(x,u_S) = \left[\int_{0}^{u_S} f(p \mid X=x)dp\right] \frac{1}{E((1-P)|X=x)}$$

* $f(p|X=x)$ is the density of $P(Z)$ given $X=x$.
Source: Heckman and Vytlacil (2001a,b; 2005, 2006)

The Generalized Roy Model of Schooling

- In the empirical literature on the returns to schooling the aim is often to estimate $E(\beta \mid X = x)$, although this is unlikely to be the answer to many relevant policy questions.
- The standard estimation method is instrumental variables.
- However, in the presence of heterogeneity and self-selection, we cannot identify $E(\beta \mid X = x)$ by using standard instrumental variables methods.
- Instead, we identify *LATE* (Imbens and Angrist, 1994), or a weighted average of *LATE* parameters, which is an instrument dependent parameter.

The Generalized Roy Model of Schooling

- It is usually broadly defined as the "average 'return' to schooling for individuals induced to change their schooling by the observed change in the instrument".
- The economic interpretation of this parameter is unclear.
- In general, *LATE* does not correspond to a policy relevant parameter or a rate of return.
- The *LATE* parameter of Imbens and Angrist (19940 is often invoked by empirical analysts to justify an instrumental variable estimate, without providing any precise definition of the economic question it addresses.

The Generalized Roy Model of Schooling

- One way to make this general point is to explore what is estimated by using compulsory schooling as an instrument.
- Compulsory schooling is sometimes viewed as an ideal instrument (see Angrist and Krueger, 1991).
- But when "returns" are heterogeneous, and agents act on that heterogeneity in making schooling decisions, compulsory schooling used as an instrument identifies only one of many possible treatment parameters and in general does not estimate a rate of return to schooling.

The Generalized Roy Model of Schooling

- Compulsory schooling selects at random persons who ordinarily would not be schooled ($S = 0$) and forces them to be schooled.
- It is straightforward to establish that it identifies treatment on the untreated:

$$E(\ln Y_1 - \ln Y_0 \mid X = x, S = 0) = E(\beta \mid X = x, S = 0)$$

but not $ATE = E(\ln Y_1 - \ln Y_0) = \bar{\beta}$, treatment on the treated $TT = E(\ln Y_1 - \ln Y_0 \mid X = x, S = 1) = E(\beta | X = x, S = 1)$, or the marginal internal rate of return.

The Generalized Roy Model of Schooling

- Treatment on the untreated answers an interesting policy question.
- It is informative about the earnings gains for a policy directed toward those who ordinarily would not attend school and who are selected into school at random from this pool.
- If the policy the analyst seeks to evaluate is compulsory schooling then the instrumental variable estimand and the policy relevant treatment effect coincide.
- More generally, if the instrumental variable we use is exactly the policy we want to evaluate, then the *IV* estimand and the policy relevant parameter are the same.

The Generalized Roy Model of Schooling

- But whenever that is not the case, the *IV* estimand does not identify the effect of the policy when returns vary among people and they make choices of treatment based on those returns.
- For example, if the policy we want to consider is a tuition subsidy directed toward the very poorest within the pool of nonattenders, then an instrumental variable estimate based on compulsory schooling will not be the relevant return to evaluate the policy.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- The index model (21) and (23) can be used to define the marginal treatment effect (*MTE*),

$$\Delta^{MTE}(x, u_S) \equiv E(\beta \mid X = x, U_S = u_S).$$

- This is the mean gain to schooling for individuals with characteristics $X = x$ and with unobservable $U_S = u_S$.
- It is a willingness to pay measure for an additional year of schooling for persons indifferent between attending or not attending college at a mean utility $P(Z) = u_S$.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- Under their assumptions, Heckman and Vytlacil (1999, 2001b, 2005, 2007b) establish that all of the conventional treatment parameters used in the program evaluation literature are different weighted averages of the *MTE* where the weights integrate to one.
- The conventional treatment parameters are the average treatment effect or *ATE*, $E(Y_1 - Y_0 \mid S = 1, x)$, and treatment on the untreated or *TUT*, $E(Y_1 - Y_0 \mid S = 0, x)$.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- See Table 9a (from Heckman and Vytlacil, 2000, 2005, 2007b) for the treatment parameters expressed in terms of MTE and Table 9b for the weights.
- The analysis of Heckman and Vytlacil (2001b, 2005, 2007b) unites the selection literature and the modern IV literature using a common analytical framework.
- Heckman, Urzua, and Vytlacil (2006) discuss how to construct the weights.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- These tables also show how one can write the *IV* and *OLS* estimates and the Policy Relevant Treatment Effect as weighted averages of the *MTE*.
- The crucial observation to extract from this table is that the weights on *MTE* are different for *IV* and for the treatment parameters.
- Thus, not only is it true that the treatment parameters are *not* rates of return, but *IV* does not in general estimate the treatment parameters.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- Figure 10a plots the marginal treatment effect (MTE) derived from a generalized normal Roy model using the parameterization of (20) and (21) shown at the base of figure 10b.
- It displays the prototypical pattern that the returns to schooling decline for those persons who have higher costs of schooling (higher U_S), i.e., for persons less likely to attend school.

Figure 10a
Weights for the Marginal Treatment Effect for Different Parameters

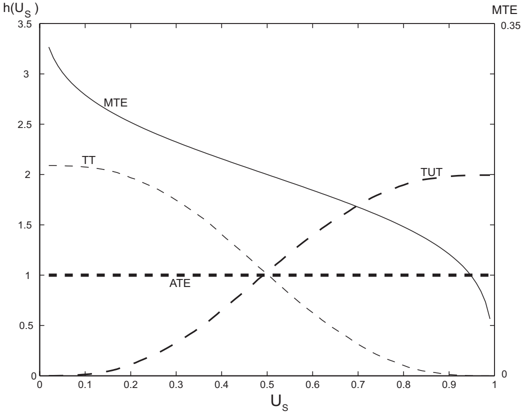
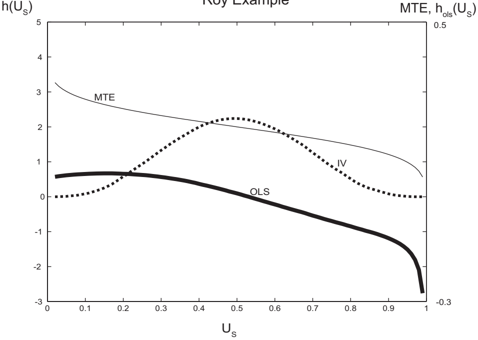


Figure 10b
Marginal Treatment Effect vs. Linear Instrumental Variables
and Ordinary Least Squares Weights
Roy Example



Source: Heckman and Vytlacil (2005b).

$$
\begin{aligned}
lnY_1 &= \alpha + \bar{\beta} + U_1 \\
lnY_0 &= \alpha + U_0 \\
S &= 1 \text{ if } Z - U_S > 0 \\
\alpha &= 0.67 \\
\bar{\beta} &= 0.2
\end{aligned}
$$

$$
\begin{aligned}
U_1 &= \sigma_1 \varepsilon \\
U_0 &= \sigma_0 \varepsilon \\
U_S &= \sigma_S \varepsilon
\end{aligned}
$$

$$
\begin{aligned}
\sigma_1 &= 0.012 \\
\sigma_0 &= -0.050 \\
\sigma_S &= -1 \\
\varepsilon &\sim N(0,1) \\
Z &\sim N(-0.0026, 0.27
\end{aligned}
$$

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- The same figure is implicit in the analysis of Willis and Rosen but they do not develop or exposit it.
- The treatment effect parameters generated from this model are presented in Table 10.
- It also presents *IV* and *OLS* estimates as well as the sorting gain and selection bias terms for this model.

Table 10: Treatment Parameters in the Generalized Roy Example

Ordinary Least Squares	0.1735
Treatment on the Treated	0.2442
Treatment on the Untreated	0.1570
Average Treatment Effect	0.2003
Sorting Gain*	0.0402
Selection Bias†	-0.0708
Linear Instrumental Variables‡	0.2017

*$E[U_1 - U_0 \mid S = 1] = TT - ATE$

†$E[U_0 \mid S = 1] - E[U_0 \mid S = 0] = OLS - TT$

‡Using propensity score as the instrument.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- Figure 10a also displays the weights on MTE used to form ATE (Average Treatment Effect), TT (Treatment on the Treated) and TUT (Treatment on the Untreated) for a generalized Roy model (with tuition costs).
- TT overweights the MTE for persons with low values of U_S who, *ceteris paribus,* are more likely to attend school.
- TUT overweights the MTE for persons with high values of U_S who are less likely to attend school.
- ATE weights MTE uniformly.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- The decline in *MTE* reveals that the "gross return" (β) declines with U_S.
- Those more likely to attend school (based on lower U_S) have higher "gross returns" or higher growth rates of earnings with schooling.
- Not surprisingly, in light of the shape of *MTE* and the shapes of the weights, $TT > ATE > TUT$.
- There is a positive sorting gain $(E(U_1 - U_0 \mid X = x, S = 1) > 0)$ and a negative selection bias $(E(U_0 \mid X = x, S = 1) - E(U_0 \mid X = x, S = 0) < 0)$.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- Figure 10b displays the *MTE* and the weights for *OLS* and for *IV* using $P(Z)$ as the instrument.
- *IV* weights the *MTE* more symmetrically and in a different fashion than *ATE*, *TUT* or *TT*.
- The shape of the *IV* weight is prototypical when $P(Z)$ is the instrument.
- However, for other instruments, including individual components of Z, the shapes of the weights are different (see Heckman, Urzua, and Vytlacil, 2006, for further analysis and examples).

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- We present examples of these weights below.
- *OLS* weights *MTE* very differently.
- The contrast between the *OLS* weight and the *IV* weight conveys the contrast between the CPS/Census literature and the modern *IV* literature.
- In general, neither identifies *ATE* or the other treatment effects, and the conventional treatment effects are not rates of return.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- To estimate *ex post* rates of return, it is necessary to account for foregone earnings and direct costs.
- The treatment effect literature typically accounts for neither and reports differences in labor market payments to different schooling levels.
- To cast the discussion into the framework of this section, let $Y_{1,t}$ be the earnings of a college-educated person at age t.
- Let $Y_{0,t}$ be the earnings for a high school-educated person at age t.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- To this point in this section we have abstracted from age-dependent growth rates of earnings.
- Suppose that it takes τ periods to complete college and that direct costs are C_t per period while in college.
- The interest rate is r, assumed to be constant.
- Assume that while in school persons receive no earnings.
- If they did, they could help offset costs C.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- College educated persons retire at age T_1.
- High school educated persons retire at age T_0.
- One definition of the return to college R is

$$R = \frac{\sum_{t=\tau}^{T_1} \frac{Y_{1,t}}{(1+r)^{t-\tau}} - \sum_{t=0}^{T_0} \frac{(Y_{0,t}+C_t)}{(1+r)^t}}{\sum_{t=0}^{T_0} \frac{Y_{0,t}+C_t}{(1+r)^t}}.$$

- This is a version of the Becker (1964) formula.
- It compares the present values of two earnings streams realized τ periods apart.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- As discussed above, in the special case assumed by Mincer, log earnings are parallel in experience across schooling categories.
- For the case of geometric growth and defining $\bar{Y}_0 = Y_{0,0}$ and $\bar{Y}_1 = Y_{1,\tau}$, earnings may be written as:

$$\begin{aligned} Y_{0,t} &= \bar{Y}_0(1+g)^t \\ Y_{1,t} &= \bar{Y}_1(1+g)^{t-\tau} \qquad t \geq \tau, \end{aligned}$$

  where g is the growth rate of earnings with age.
- Mincer further assumes that $T_1 - T_0 = \tau$ so working lives are the same for both schooling classes.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- The discounted growth rate of earnings with experience, e, is

$$e = \sum_{j=0}^{T_0} \left(\frac{1+g}{1+r} \right)^j .$$

- Assume that direct costs (psychic and tuition) are the same per period during the schooling years and define

$$A(\tau) = \sum_{j=0}^{\tau} \left(\frac{1}{1+r} \right)^j .$$

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- The return in this case is

$$R = \frac{\bar{Y}_1 e - \bar{Y}_0 e - CA(\tau)}{CA(\tau) + \bar{Y}_0 e}.$$

- The growth rate of earnings with schooling is

$$\phi = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{Y}_0} \approx \ln \bar{Y}_1 - \ln \bar{Y}_0.$$

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- This is the "Mincer return" to schooling.
- An alternative expression for the return is

$$R = \frac{\phi - \frac{CA(\tau)}{\bar{Y}_0 e}}{1 + \frac{CA(\tau)}{\bar{Y}_0 e}}.$$

- This shows that the Mincer return ϕ, is greater than the true return, R, whenever costs are positive.
- When costs are zero ($C = 0$), R equals the Mincer return, ϕ.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- Thus, the Mincer assumptions justify the conventional practice of equating growth rates to rates of returns, the implicit assumption in the recent literature on estimating rates of return.
- In general, if $1 + R > (1 + r)^{\tau}$, it pays to go to college; otherwise, it does not.
- An alternative way to state this criterion is that it pays to go to college if
$$1 + \phi > (1 + r)^{\tau}.$$
- When $\tau = 1$, this simplifies to the conventional criterion that $\phi > r$.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- The evidence argues strongly against the practice of equating growth rates with rates of return.
- Mincer's parallelism assumption across schooling levels (i.e., that growth rates of earnings with experience, g, are the same for all schooling levels) is not accurate for earnings profiles from more recent data.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- Additionally, the evidence presented below points to the existence of substantial psychic cost components and an adjustment for psychic cost components substantially reduces the rate of return to schooling.
- The current literature on estimating rates of return makes none of these adjustments and instead reports the growth rate of earnings as a "return."
- While the growth rate of ϕ is an ingredient of returns, it is not in general a return, as the expression for R reveals.

Defining Treatment Effects in the Generalized Roy Model and Relating them to True Rates of Return

- We can use the modern literature to identify growth rates of earnings for persons at different margins of choice.
- Costs, discount rates and horizons need to be adjusted appropriately to get true rates of return.
- To our knowledge, this has not been done in the vast *IV* literature on computing rates of return.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- In the generalized Roy model, there are three sources of potential econometric problems;
  (a) S is correlated with U_0;
  (b) β is correlated with S (i.e., $U_1 - U_0$ is correlated with S);
  (c) β is correlated with U_0.
- The relative importance of the problems depends on what question the analyst seeks to answer.
- Source (a) arises in ability bias or measurement error models.
- Source (b) arises if agents partially anticipate β when making schooling decisions so that $\Pr(S = 1 \mid X, \beta) \neq \Pr(S = 1 \mid X)$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- In this framework, β is an *ex post* "causal effect," which may not be known to agents *ex ante*.
- In the case where decisions about S are made in the absence of information about β, β is independent of S.
- Source (c) arises from the possibility that the gains to schooling (β) may be dependent on the level of potential earnings in the unschooled state (Y_0) as in the Roy model.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- When $U_1 = U_0$, β is a constant for all persons (conditional on X), and we obtain the conventional *IV* model as analyzed by Griliches (1977).
- In this framework, because β is a constant, there is a unique effect of schooling.
- Indeed, β is "the" effect of schooling, and the marginal effect is the same as the average effect (conditional on X).

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- In the notation of equation (20), the usual assumption in the literature is that $\text{Cov}\,(S, U_0) > 0$.
- Measured schooling S may be correlated with unmeasured U_0 because of omitted ability factors.
- Therefore, when β is constant across individuals, the *OLS* estimate of the "return" is an upward biased estimate of β:

$$\text{plim } \hat{\beta}_{OLS} = \beta + \frac{\text{Cov}\,(S, U_0)}{V\,(S)} > \beta.$$

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- Following Griliches (1977) and the scholars who preceded him, many advocate using instrumental variable estimators for β to correct for this problem.
- If there is an instrument Z such that $\text{Cov}(Z, S) \neq 0$ and $\text{Cov}(Z, U_0) = 0$, then:

$$\text{plim } \hat{\beta}_{IV} = \beta + \frac{\text{Cov}(Z, U_0)}{\text{Cov}(Z, S)} = \beta.$$

- Therefore we expect that $\hat{\beta}_{IV} < \hat{\beta}_{OLS}$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- However, as noted by Griliches (1977) and Card (1995, 1999, 2001), almost all of the empirical literature on the returns to schooling shows precisely the opposite pattern: $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$.
- How can one rationalize this finding?
- One standard explanation is that schooling is measured with error.
- This would induce a downward bias in the schooling coefficient, which would be corrected by the use of *IV*.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- This simple explanation has been questioned in two different ways.
- Kane, Rouse, and Staiger (1999) claim that measurement error in schooling is nonclassical and therefore we might not expect the standard attenuation bias that results from nonclassical measurement error.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- Card (1999, 2001) argues that, if measurement error is classical, the amount of measurement error in schooling that would have to exist to justify the large gaps between *OLS* and *IV* estimates is unreasonably large.
- He argues that, in fact, schooling is relatively well measured in the U.S., so that the measurement error explanation for the empirical regularity is likely to be of second order importance.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- The explanation for the empirical regularity that Card (1999, 2001) favors is that there is heterogeneity in the returns to schooling so β is a random variable and it is correlated with schooling.
- For a model with two levels of schooling, this is just the generalized Roy model.
- In this case, it is possible that *IV* estimates of returns to schooling exceed *OLS* estimates.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- Implicitly, his argument has three steps:
  1. *OLS* is an upward biased estimate of the average "return to schooling" (this is the standard ability bias intuition in a model in which β is the same for everyone);
  2. *IV* corresponds to an estimate of the returns to schooling for individuals at the margin; and therefore,
  3. if the *IV* estimate of the "return" exceeds the *OLS* estimate of the "return," then individuals at the margin have higher "returns" than the average individual in the economy.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- In our notation, the probability limits of the least squares and *IV* estimators are

$$\begin{aligned}
\text{plim } \hat{\beta}_{OLS} &= \bar{\beta} + \frac{\text{Cov}(S, U_0)}{V(S)} + \frac{\text{Cov}[S, S(U_1 - U_0)]}{V(S)} \quad (25) \\
\text{plim } \hat{\beta}_{IV} &= \bar{\beta} + \frac{\text{Cov}[Z, S(U_1 - U_0)]}{\text{Cov}(Z, S)} + \frac{\text{Cov}(Z, U_0)}{\text{Cov}(Z, S)} \\
&= \bar{\beta} + \frac{\text{Cov}[Z, S(U_1 - U_0)]}{\text{Cov}(Z, S)}. \quad (26)
\end{aligned}$$

In general, plim $\hat{\beta}_{OLS}$ can be larger than, smaller than or equal to plim $\hat{\beta}_{IV}$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- We can rewrite (25) and (26) as:

$$\begin{aligned}\text{plim } \hat{\beta}_{OLS} &= \bar{\beta} + E(U_0 \mid S=1) - E(U_0 \mid S=0) \\ &\quad + E(U_1 - U_0 \mid S=1) \\ &= E(\beta \mid S=1) + E(U_0 \mid S=1) - E(U_0 \mid S=0)\end{aligned}$$

$$\begin{aligned}\text{plim } \hat{\beta}_{IV} &= \bar{\beta} + E(U_1 - U_0 \mid S=1) \\ &\quad + \frac{\text{Cov}[Z, (U_1 - U_0) \mid S=1] \Pr(S=1)}{\text{Cov}(Z,S)} \\ &= E(\beta \mid S=1) \\ &\quad + \frac{\text{Cov}[Z, (U_1 - U_0) \mid S=1] \Pr(S=1)}{\text{Cov}(Z,S)}.\end{aligned}$$

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- Therefore, plim $\hat{\beta}_{IV}$ >plim $\hat{\beta}_{OLS}$ if $\frac{\text{Cov}[Z,(U_1-U_0)|S=1]\Pr(S=1)}{\text{Cov}(Z,S)} > E(U_0|S=1) - E(U_0|S=0)$.
- The assumption implicit in Card's argument, and in the standard ability bias literature, is that $E(U_0|S=1) - E(U_0|S=0) > 0$.
- This condition is satisfied if persons who go to college are above average in high school.
- In such a case, current college graduates would be at the top of the high school wage distribution if they chose to become high school graduates.
- If this model generates the data, the only way that plim $\hat{\beta}_{IV}$ >plim $\hat{\beta}_{OLS}$ is if $\frac{\text{Cov}[Z,(U_1-U_0)|S=1]\Pr(S=1)}{\text{Cov}(Z,S)} > 0$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- How plausible is this condition?
- Recall that Z is a determinant of the cost of schooling $C(Z)$ and satisfies the standard instrumental variable assumptions.
- Assume that C is increasing in Z which is assumed to be scalar.
- As a consequence of these two conditions,

$$\text{Cov}\,(Z, S) < 0 \text{ and } \text{Cov}\,(Z, U_1) = \text{Cov}\,(Z, U_0) = 0. \qquad (27)$$

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- In the simple two outcome model of schooling, individuals enroll in school if benefits are higher than costs as is clear from equation (21) ($S = 1$ if $\beta - C(Z) = \bar{\beta} + (U_1 - U_0) - C(Z) > 0$).
- In such a model the average individual who attends school has a higher return than the marginal individual ($E(\beta | S = 1) > E(\beta | \beta = C(Z))$).
- Furthermore, even though $\text{Cov}(Z, U_1 - U_0) = 0$, $\text{Cov}(Z, U_1 - U_0 | S = 1) > 0$ (if an individual has a high cost, or high Z, he or she will only attend school if he or she also has a high $U_1 - U_0$).

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- But in that case, because $\text{Cov}(Z, S) < 0$, plim $\hat{\beta}_{IV} <$ plim $\hat{\beta}_{OLS}$.
- Implicit in Card's analysis is the assumption that it is not possible for the average student to have a higher return than the marginal student and still find that $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$.
- Card rationalizes $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$ by assuming that the marginal student with a higher return than the average student is out of school because of some external constraint, such as a liquidity constraint so $E\left(\beta | S = 1\right) < E(\beta | \beta = C\left(Z\right))$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- The less able (lower β) people are excluded from school.
- In Card's original model, the "returns " to schooling decrease with the amount of schooling for each individual ($k_1 < 0$), and those individuals whose schooling decision is more sensitive to changes in the instrument have relatively little schooling and, as a consequence, relatively high returns.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- Drawing on the generalized Roy model, Carneiro and Heckman (2002) and Carneiro, Heckman and Vytlacil (2005) argue instead that the reason why $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$ is not that the marginal student has a higher return than the average student $(E(\beta|S=1) < E(\beta|\beta = C(Z)))$, but instead that $E(U_0|S=1) - E(U_0|S=0) < 0$.
- They show empirically, for a nationally representative sample of U.S. white males (NLSY79), that the marginal "return" is below the average for college goers while, simultaneously, $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- In their setup, $\frac{\text{Cov}[Z,(U_1-U_0)|S=1]\Pr(S=1)}{\text{Cov}(Z,S)} < 0$, $E\left(U_0|S=1\right) - E\left(U_0|S=0\right) < 0$ and $\frac{\text{Cov}[Z,(U_1-U_0)|S=1]\Pr(S=1)}{\text{Cov}(Z,S)} > E\left(U_0|S=1\right) - E\left(U_0|S=0\right)$.
- *OLS* estimates are downward biased for $E\left(\beta|S=1\right)$ because $E\left(U_0|S=1\right) - E\left(U_0|S=0\right) < 0$.
- For example, if individuals with $S=1$ become teachers and those with $S=0$ become plumbers, then the latter are better plumbers than the average teacher would be if he became a plumber.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- This possibility is featured in Willis and Rosen (1979), who speculate that, contrary to conventional wisdom, $\text{Cov}(U_1, U_0) < 0$, although, with their model, they cannot identify this correlation from the data.
- Carneiro, Heckman, and Vytlacil (2005) and Cunha, Heckman, and Navarro (2005) identify this covariance and find evidence that supports the Willis-Rosen conjecture of a negative correlation.
- When analysts use *OLS*, they compare $E(Y_1|S=1)$ with $E(Y_0|S=0)$ (see equation (25)), and since $E(Y_0|S=0) > E(Y_0|S=1)$, the *OLS* estimate is an underestimate of $E(Y_1 - Y_0|S=1)$.

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- To summarize, an important lesson from the recent literature is that in a model of heterogeneous returns, intuitions about ability bias are no longer as simple as in the standard homogeneous returns model with a single measure of ability (Griliches, 1977).
- In such a model, the most able people enroll in school.
- In a more general Roy-type model, there can be multiple abilities (in this case, U_1 and U_0), which can be arbitrarily correlated (positively or negatively).

Understanding Why IV Estimates Exceed OLS Estimates of the Schooling Coefficient

- The idea that individuals with "high ability" are more likely to enroll in school is no longer obvious.
- Recent evidence supports the claim that the most able persons in the U_0 distribution (high school skills) do not go on to college.
- This is true not only in models of schooling, but also in many other models in economics where returns to an activity are heterogeneous and people sort into different activities based on those returns.

Estimating the MTE

- We now show how the local *IV* methods of Heckman and Vytlacil (1999, 2001b, 2005, 2007b) can be used to estimate average returns to school for any population of interest.
- Heckman, Urzua, and Vytlacil (2006) show how to estimate the *MTE* and generate all of the weights shown in Table 9b.
- They also provide software for doing so.

Estimating the MTE

- Using equation (19) the conditional expectation of log Y ($= \ln Y_0(1-S) + \ln Y_1 S$) is

$$
\begin{aligned}
& E(\ln Y \mid Z = z) \\
& \quad = E(\ln Y_0 \mid Z = z) \\
& \qquad + E(\ln Y_1 - \ln Y_0 \mid Z = z, S = 1) \Pr(S = 1 \mid Z = z)
\end{aligned}
$$

  where we keep the conditioning on X implicit.

- From the index structure generated by decision rules (21) and (23), we may write this expectation as

$$
\begin{aligned}
& E(\ln Y \mid Z = z) \\
& \quad = E(\ln Y_0) + E(\beta \mid P(z) \geq U_S, P(Z) = P(z))P(z).
\end{aligned}
$$

Estimating the MTE

- Observe that the instruments enter the model through the probability of selection or the propensity score ($P(z)$).
- Using $P(z)$ as the instrument, and applying the Wald estimator for two different values of Z, z and z', assuming $P(z) < P(z')$, we obtain the *IV* formula:

$$
\begin{aligned}
& \frac{E(\ln Y \mid P(Z) = P(z)) - E(\ln Y \mid P(Z) = P(z'))}{P(z) - P(z')} \\
= \;& \bar{\beta} + \frac{\left\{ \begin{array}{l} E(U_1 - U_0 \mid P(z) \geq U_S)P(z) \\ \quad - E(U_1 - U_0 \mid P(z') \geq U_S)P(z') \end{array} \right\}}{P(z) - P(z')} \\
= \;& E\left(\beta \mid P\left(z\right) < U_S \leq P\left(z'\right)\right) \\
= \;& \Delta^{LATE}(P(z), P(z')),
\end{aligned}
$$

where Δ^{LATE} is the *LATE* parameter.

Estimating the MTE

- This is the average return to schooling for individuals who have U_S between $P(z)$ and $P(z')$ $(P(z) < U_S \leq P(z'))$.
- As we make z and z' closer to each other, we identify β for a narrower group of individuals defined in terms of their U_S.
- The *MTE* can therefore be estimated by taking a limit of *LATE* when z and z' are arbitrarily close to each other.
- When $U_1 \equiv U_0$ or $(U_1 - U_0) \perp\!\!\!\perp U_S$, corresponding to the two special cases in the literature, *IV* based on $P(Z)$ estimates *ATE* $(= \bar{\beta})$ because the second term on the right hand side (second line) of this expression vanishes.

Estimating the MTE

- Otherwise *IV* estimates an economically difficult-to-interpret combination of *MTE* parameters with weights given in Table 9b.

Estimating the MTE

- Another representation of $E(\ln Y \mid P(Z) = P(z))$ reveals the index structure underlying this model more explicitly and writes

$$
\begin{aligned}
&E(\ln Y \mid P(Z) = P(z)) \\
&= \alpha + \bar{\beta} P(z) \\
&\qquad + \int_{-\infty}^{\infty} \int_{0}^{P(z)} (U_1 - U_0) f(U_1 - U_0 \mid U_S = u_S)\, du_S \, d(U_1 - U_0).
\end{aligned}
\tag{28}
$$

Estimating the MTE

- Differentiating with respect to $P(z)$, we obtain MTE:

$$\frac{\partial E(\ln Y \mid P(Z) = P(z))}{\partial P(z)}$$
$$= \bar{\beta} + \int_{-\infty}^{\infty} (U_1 - U_0) f(U_1 - U_0 \mid U_S = P(z))\, d(U_1 - U_0)$$
$$= \Delta^{MTE}(P(z)).$$

Estimating the MTE

- IV estimates $\bar{\beta}$ if $\Delta^{MTE}(u_S)$ does not vary with u_S.
- Under this condition $E(\ln Y \mid P(Z) = P(z))$ is a linear function of $P(z)$.
- Thus, under our assumptions, a test of the linearity of the conditional expectation of $\ln Y$ in $P(z)$ is a test of the validity of linear IV for $\bar{\beta}$.
- It is also a test for the validity of conditions I and II.
- Heckman, Urzua, and Vytlacil (2006) elaborate on this point.

Estimating the MTE

- More generally, a test of the linearity of $E(\ln Y \mid P(Z) = P(z))$ in $P(z)$ is a test of whether or not the data are consistent with a correlated random coefficient model and is also a test of comparative advantage in the labor market for educated labor.
- If $E\left(\ln Y|P\left(z\right)\right)$ is linear in $P\left(z\right)$, standard instrumental variables methods identify "the" effect of S on $\ln Y$.

Estimating the MTE

- In contrast, if $E(\ln Y|P(z))$ is nonlinear in $P(z)$, then there is heterogeneity in the return to college attendance, individuals act at least in part on their own idiosyncratic return, and standard linear instrumental variables methods will not in general identify the average treatment effect or any other of the treatment parameters defined earlier.
- This test for nonlinearity in $P(Z)$ as a sign of correlated heterogeneity is simple to execute and interpret.
- Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2006) implement it and find evidence in support of nonlinearity in the data they analyze.

Estimating the MTE

- It is straightforward to estimate the levels and derivatives of $E(\ln Y \mid P(Z) = P(z))$ and standard errors using the methods developed in Heckman, Ichimura, Smith, and Todd (1998).
- The derivative estimator of MTE is the local instrumental variable (LIV) estimator of Heckman and Vytlacil (1999, 2001b, 2005, 2007b).

Estimating the MTE

- This framework can be extended to consider multiple treatments, which in this case can be either multiple years of schooling, or multiple types or qualities of schooling.
- These can be either continuous (see Florens, Heckman, Meghir, and Vytlacil, 2002) or discrete (see Carneiro, Hansen, and Heckman, 2003; Heckman and Vytlacil, 2005, 2007a,b; Heckman, Urzua, and Vytlacil, 2006).

Estimating the MTE

- Heckman, Urzua, and Vytlacil (2006) establish the close relationship between selection models and instrumental variables models when the response of earnings to changes in education varies among persons.
- Essentially, *IV* estimates the derivatives of outcome equations while the control function estimates them in levels.

Evidence From the Instrumental Variables Literature

- Card (1999) surveys empirical estimates from the instrumental variables literature.
- In the case of the general model presented in this chapter, different instruments identify different weighted averages of the *MTE* and in general do not identify any interpretable economic object such as a rate of return to schooling.
- The intensity of the search for instruments Z uncorrelated with (U_0, U_1) and correlated with S has not been matched by an equally intense search for an interpretation of what economic question the instrumental variables estimators answer.

Evidence From the Instrumental Variables Literature

- As noted by Heckman and Vytlacil (2007a,b) and Heckman, Urzua, and Vytlacil (2006), since the question being addressed by the recent literature is not clearly stated, it is not obvious that *IV* is better than *OLS*.
- The estimates produced from many of the commonly used instruments have large standard errors in producing any particular parameter of interest except for parameters defined by instruments.
- On a purely statistical basis there is often little difference between *IV* and *OLS* estimates once sampling variation is accounted for.
- Many of the instruments used in this literature are controversial.

Evidence From the Instrumental Variables Literature

- Parental education and number of siblings have been used as instruments by Willis and Rosen (1979) and Taber (2001).
- They tend to produce estimates of "effects" with small standard errors.
- However, they are controversial.
- It is necessary to assume that potential wages in both the college and high school state are independent of family background, but many studies show that these are determinants of ability.

Evidence From the Instrumental Variables Literature

- See Cunha, Heckman, Lochner, and Masterov, 2006.
- Unless one controls for ability, the quality of the instruments is in question.
- Many data sets lack direct measures of ability.

Evidence From the Instrumental Variables Literature

- Other popular instruments are based on the geographic location of individuals at the college going age.
- If the decision of going to college and the location decision are correlated then these instruments are not valid.
- For example, individuals who are more likely to enroll in college may choose to locate in areas where colleges are abundant and inexpensive.
- Distance to college is used as an instrument for schooling by Card (1993), Kling (2001) and Cameron and Taber (2004).

Evidence From the Instrumental Variables Literature

- Carneiro and Heckman (2002) show that distance to college in the NLSY79 is correlated with a measure of ability and is an invalid exclusion unless the analyst conditions on ability since ability determines outcomes.
- Tuition is used as an instrument by Kane and Rouse (1995).
- Average tuition in the county of residence may also be a problematic instrument since it is correlated with average college quality in the county (see Carneiro and Heckman, 2002).

Evidence From the Instrumental Variables Literature

- Finally, local labor market variables have been used by Cameron and Heckman (1998), Carneiro, Heckman, and Vytlacil (2005) and Cameron and Taber (2004).
- Cameron and Taber use a measure of the local wage.
- Carneiro, Heckman and Vytlacil use a measure of local unemployment.
- They also control for long term wages in the county of residence both in the selection and in the outcome equations, so that the instrument measures business cycle fluctuations orthogonal to the long term quality of the location of residence.

Evidence From the Instrumental Variables Literature

- The CPS and Census data sets lack strong instruments and for that reason few analysts of those data use the method of *IV*.
- The "quarter of birth" instrument used by Angrist and Krueger (1991) to identify treatment on the untreated is notoriously weak (see Staiger and Stock, 1997).

Evidence From the Instrumental Variables Literature

- Rather than reproduce Card's (1999) survey, we present some evidence from Carneiro, Heckman, and Vytlacil (2005) on estimates of the *MTE* using the method of local instrumental variables and some estimates from Heckman, Urzua, and Vytlacil (2006).
- Both sets of authors use the NLSY data set.
- The fragility of the estimates and the large standard errors document the problems that plague the application of the *IV* to data sets with rich instruments that typically have only a few thousand observations.

Evidence From the Instrumental Variables Literature

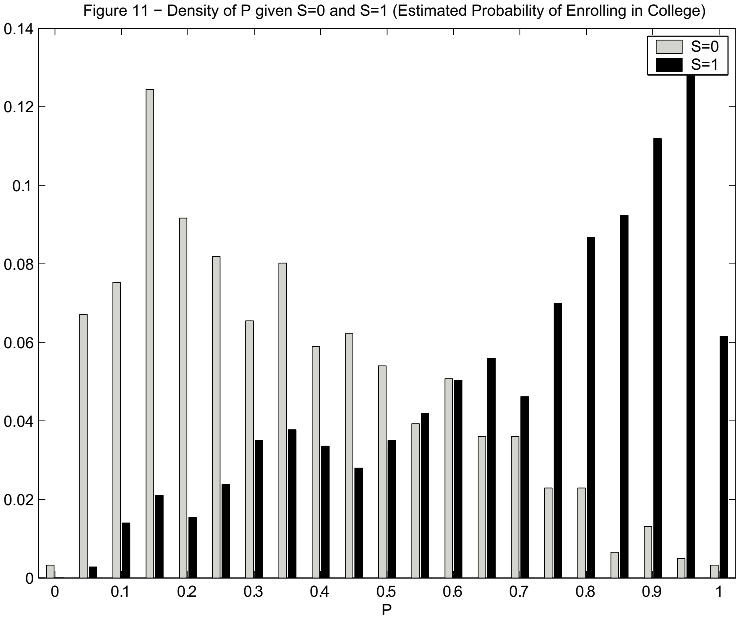
- The details of the estimation procedure used to generate the numbers reported in this section are described in Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2006).
- The regressors in the choice equation are a measure of ability (the Armed Forces Qualifying Test or AFQT), number of siblings, mother's and father's education levels, tuition, distance to college, local unemployment rate, and interaction terms.
- Tuition is a strong predictor of schooling, as are family background and AFQT.

Evidence From the Instrumental Variables Literature

- The exclusions from the earnings equations are tuition, distance to college, local unemployment and opportunity wages in blue collar work at age 17.
- "Distance to College at 14" and "Local Unemployment Rate at 17" have weak effects.
- The only strong exclusions are tuition and opportunity wages, conditioning on ability.

Evidence From the Instrumental Variables Literature

- The density of $P(Z)$ and the support of the estimated propensity score $P(Z)$ (the region over which $P(Z)$ has positive density) is shown in Figure 11 for the Carneiro, Heckman, and Vytlacil (2005) study.
- It is almost the full unit interval, although at the extremes of the interval the cells of data become very thin.
- In their estimation of the *MTE*, Carneiro, Heckman and Vytlacil only use values of P between 0.07 and 0.98.



Note: P is the estimated probability of going to college. It is estimated from a logit regression of college attendance on corrected AFQT, father's education, mother's education, number of siblings, tuition, distance to college and local unemployment.
Source: Carneiro, Heckman and Vytlacil (2005)

Evidence From the Instrumental Variables Literature

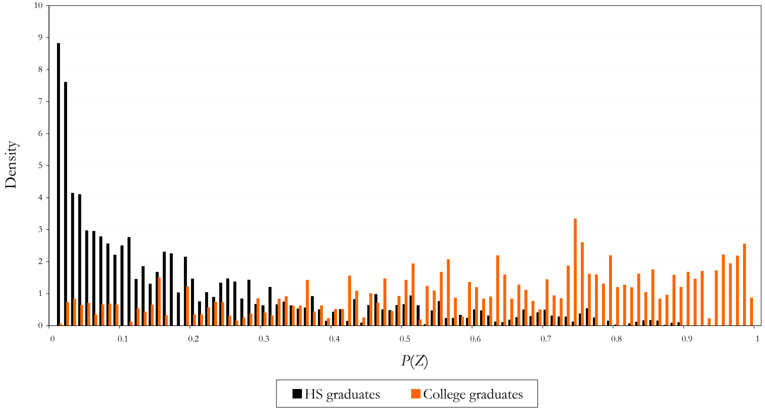
- They trim 5% of the observations in the sample.
- Even after trimming, the sparseness of data in the tails results in a large amount of noise (variability) in the estimation of $E(Y|X, P(Z) = p)$ for values of p close to 0.07 or 0.98, which in turn makes problematic estimation of the parameters defined over the full support of U_S (which require estimation of $E(Y|X, P(Z) = p)$ over the full unit interval).

Evidence From the Instrumental Variables Literature

- The lack of full support of $P(Z)$ means that ATE, TT and TUT are not identified nonparametrically by the method of instrumental variables.
- However the MTE can be estimated pointwise for a wide range of evaluation points without full support.
- This highlights what $LATE$ can and cannot do in these data.
- It can produce a number.
- It cannot produce even a conventional treatment effect, much less a rate of return to schooling.
- The pattern of support of $P(Z)$ is similar in the Heckman, Urzua, and Vytlacil (2006) study.
- See Figure 12 taken from their analysis.

Figure 12. Frequency of the Propensity Score by Final Schooling Decision

HS Graduates and Four Year College Graduates - Males of the NLSY at age 30



Source: Heckman, Urzua and Vytlacil (2004)

Evidence From the Instrumental Variables Literature

- Cognitive ability (as measured by AFQT) is an important determinant of the returns to schooling.
- Simple least squares regressions of log wages on schooling, ability measures, and interactions of schooling and ability (ignoring selection arising from uncontrolled unobservables) have been widely estimated in this and other data sets and generally show that cognitive ability is an important determinant of the returns to schooling.

Evidence From the Instrumental Variables Literature

- Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2006) include AFQT in their model as an observable determinant of the returns to schooling and of the decision to go to college.
- In the absence of such a measure of cognitive ability, selection arising from unobservables should be important.
- Most of the data sets that are used to estimate the returns to education (such as the Current Population Survey or the Census) lack such ability measures.

Evidence From the Instrumental Variables Literature

- The test for selection on the individual returns to attending college checks whether $E(\ln Y|X, P)$ is a linear or a nonlinear function of P.
- Nonlinearity in P means that there is heterogeneity in the returns to college attendance and that individuals select into college based at least in part on their own idiosyncratic return (conditional on X).
- One possible way to implement this test is to approximate $K(P)$ with a polynomial in P and test whether the coefficients in the terms of the polynomial of order higher than one are jointly equal to zero.
- Carneiro, Heckman and Vytlacil test and reject linearity, indicating that a correlated random coefficient model describes the NLSY data.

Evidence From the Instrumental Variables Literature

- Carneiro, Heckman, and Vytlacil (2005) partition the estimated *MTE* into two components, one depending on X and the other on u_S,

$$\begin{aligned} MTE\left(x, u_S\right) &= E\left(\ln Y_1 - \ln Y_0 | X = x, U_S = u_S\right) \\ &= \mu_1\left(X\right) - \mu_0\left(X\right) + E\left(U_1 - U_0 | U_S = u_S\right). \end{aligned}$$

- Figure 13 plots the component of the *MTE* that depends on U_S but not on X where the confidence interval bands are bootstrapped.
- The estimates are obtained using Robinson's (1988) method for estimating partially linear models.

Figure 13 − $E(Y_1-Y_0|X,U_S)$ estimated using Locally Quadratic Regression (averaged over X)

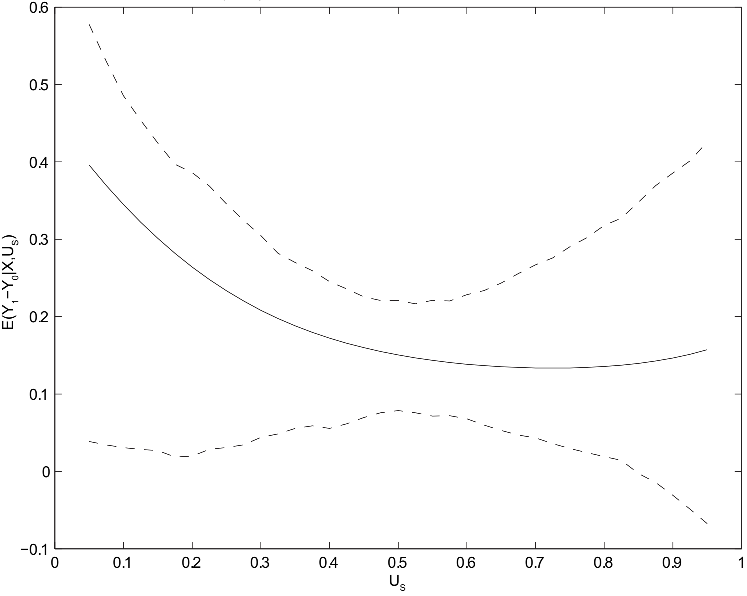


Figure 13: Legend

To estimate the function in this figure $(E(Y_1 - Y_0|X, U_S))$ we use a two step procedure. We first estimate $\mu_0(X)$ and $\mu_1(X)$ from a regression of log wages on polynomials in X, interactions of polynomials in X and P, and a nonparametric function of P (where P is the predicted probability of attending college). We use Robinson's (1988) method for estimating partially linear models. X includes experience, corrected AFQT and local unemployment. Then we compute the residual of this regression by subtracting $\mu_0(X) + P * [\mu_1(X) - \mu_0(X)]$ from log wages. Finally we estimate the nonlinear function in the figure by running a local quadratic regression of this residual on P and taking the coefficients on the linear term. Then we add a constant term to this function which is simply the average of $\mu_1(X) - \mu_0(X)$. $E(Y_1 - Y_0|X, U_S)$ is divided by 3.5 to account for the fact that individuals that attend college have on average 3.5 more years of schooling than those who do not. Therefore these correspond to estimates of returns to one year of college. The confidence interval bands are bootstrapped (250 replications).

Evidence From the Instrumental Variables Literature

- $E(U_1 - U_0 \mid U_S = u_S)$ is declining in u_S for values of u_S below 0.7, and then it is flat and if anything it slightly rises.
- Returns are annualized to reflect the fact that college goers on average attend 3.5 years of college.
- The most college worthy persons in the sense of having high gross returns are more likely to go to college (they have low U_S).
- The magnitude of the heterogeneity in returns is substantial: returns can vary from 13% to 40% per year of college.

Evidence From the Instrumental Variables Literature

- The wide standard error bands are symptomatic of a phenomenon that plagues the entire *IV* literature.
- Estimates are not precisely determined.
- Figure 14 from Heckman, Urzua, and Vytlacil (2004) reveals a similar pattern and a wide band of standard errors.
- Over broad intervals the confidence bands include zero indicating no effect of schooling on earnings.
- If β is independent of S, the *MTE* is flat.
- The evidence clearly rejects this so a correlated random coefficient model describes their data but there is a considerable loss in precision in using instrumental variables.

Figure 14. MTE with Confidence Interval

Sample of HS Graduates and Four Year College Graduates − Males at age 30 − Nonparametric

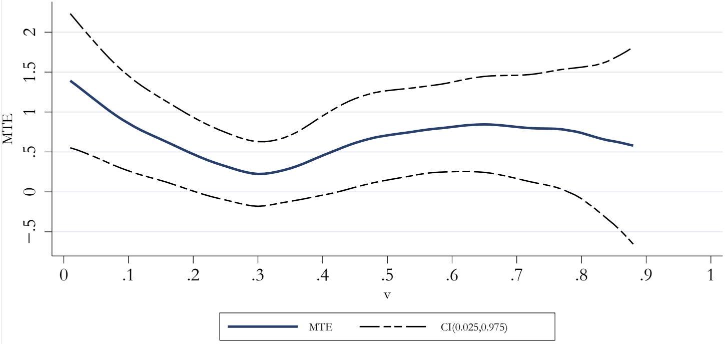


Figure 14: Legend

The dependent variable in the outcome equation is hourly earnings at age 30. The controls in the outcome equations are tenure, tenure squared, experience, corrected AFQT, black (dummy), hispanic (dummy), marital status, and years of schooling. Let $D = 0$ denote dropout status, and $D = 1$ denote GED status. The model for D (choice model) includes as controls the corrected AFQT, number of siblings, fathers education, mothers education, family income at age 17, local GED costs, broken home at age 14, average local wage at age 17 for dropouts and high school graduates, local unemployment rate at age 17 for dropouts and high school graduates, the dummy variables black and hispanics, and a set of dummy variables controlling for the year of birth. The choice model is estimated using a probit model. In computing the MTE, the bandwidth in the first step is selected using the leave-one-out cross-validation method. In the second step, following Heckman, Ichimura and Todd (1998), we set the bandwidth to 0.3. We use biweight kernel functions.

Evidence From the Instrumental Variables Literature

- Table 11 presents estimates of different summary measures of returns to one year of college for two models from Carneiro, Heckman, and Vytlacil (2005).
- In the first column they use family background as an exclusion and in the second they do not.
- The point estimates are similar in both models but they are more precise in the first one, and therefore we focus on those.
- However, this precision in estimation is obtained by using what many would argue are invalid exclusion restrictions.
- These parameters are obtained by using the appropriate weights for each parameter (see Carneiro, Heckman, and Vytlacil, 2005).

Table 11: Estimates of Various Returns to One Year of College

	Family Background is Exclusion	Family Background is not Exclusion
	$0.07 < P < 0.98$	$0.07 < P < 0.98$
Average Treatment Effect	0.2124	0.1638
	(0.0648)	(0.0916)
	[0.0069;0.2641]	[-0.0074;0.2955]
Treatment on the Treated	0.3202	0.2279
	(0.1103)	(0.1171)
	[0.0045;0.4094]	[-0.0036;0.3820]
Treatment on the Untreated	0.1042	0.0897
	(0.0802)	(0.1285)
	[-0.0027;0.2522]	[-0.1400;0.3024]
Policy Relevant Treatment Effect	0.2489	0.1905
($500 Tuition Subsidy)	(0.0854)	(0.1651)
	[0.0024;0.3520]	[-0.1037;0.3602]
Ordinary Least Squares	0.0788	0.0796
	(0.0091)	(0.0114)
	[0.0654;0.0955]	[0.0614;0.0983]
Instrumental Variables	0.1649	0.1530
	(0.0389)	(0.0758)
	[0.0888;0.2166]	[0.0036;0.2479]

Notes: Bootstrapped 5-95% standard errors (in parenthesis) and confidence intervals (in brackets) are presented below the corresponding coefficients (250 replications).
Source: Carneiro, Heckman and Vytlacil (2005).

Evidence From the Instrumental Variables Literature

- The limited support of P near the boundary values of $P = 0$ and $P = 1$ creates a practical problem for the computation of the treatment parameters such as ATE, TT, and TUT, since MTE cannot be estimated for values of U_S outside the support of P.
- The sparseness of the data in the extremes does not allow accurate estimation of the MTE at evaluation points close to 0 or 1.
- The numbers presented in Table 11 are constructed after restricting the weights to integrate over the region $[0.07, 0.98]$.
- These can be interpreted as the parameters defined in the empirical (trimmed) support of $P(Z)$, which is close to the full unit interval.

Evidence From the Instrumental Variables Literature

- The sensitivity of estimates to lack of support in the tails ($P = 0$ or $P = 1$) is important for parameters, such as ATE or TT, that put substantial weight on the tails of the MTE distribution.
- Even with support over most of the interval $[0, 1]$, such parameters cannot be identified unless 0 (for both ATE and TT) and 1 (for ATE) are contained in the support of the distribution of $P(Z)$.

Evidence From the Instrumental Variables Literature

- Estimates of these parameters are highly sensitive to imprecise estimation or extrapolation error for $E(Y|X, P(Z) = p)$ for values of p close to 0 or 1 .
- Even though empirical economists often seek to identify ATE and TT, usually they are not easily estimated nor are they always economically interesting parameters.
- As we have stressed repeatedly, they are not rates of return.

Evidence From the Instrumental Variables Literature

- Integrating only over $P(Z)$ in the interval $[0.07, 0.98]$, Table 11 reports estimates of the average annual return to college for a randomly selected person in the population (*ATE*) of 21.24%, which is between the annual return for the average individual who attends college (*TT*), 32.02%, and the average return for high school graduates who never attend college (*TUT*), 10.42%. Card reports *IV* estimates between 6 and 16% using different instruments but, as previously noted, different instruments weight *MTE* differently and answer different implicit questions.

Evidence From the Instrumental Variables Literature

- None of these numbers corresponds to the average annual return to college for those individuals of poor backgrounds who are induced to enroll in college by a $500 tuition subsidy (*PRTE*), which is 24.89%.
- This is the relevant return for evaluating this specific policy using a Benthamite welfare criterion.
- It is below *TT*, which means that the marginal entrant induced to go to college by this specific policy has an annual return well below (ten log points) that of the average college attendee.

Evidence From the Instrumental Variables Literature

- Carneiro, Heckman, and Vytlacil (2005) compare all of these estimated summary measures of returns with the *OLS* and *IV* estimates of the annual return to college, where the instrument is $\hat{P}(Z)$, the estimated probability of attending college for individuals with characteristics Z.
- *OLS* estimates *ATE* if S and X are orthogonal to $U_0 + S(U_1 - U_0)$.
- Since the returns estimated by *OLS* and by *IV* both depend on X, they evaluate the *OLS* and *IV* returns at the average value of X for individuals induced to enroll in college by a $500 tuition subsidy, so that they can compare these estimates with the policy relevant treatment effect.

Evidence From the Instrumental Variables Literature

- The *OLS* estimate of the return to a year of college is 7.88% while the *IV* estimate is 16.49%.
- Only by accident does *IV* identify policy relevant treatment effects when the *MTE* is not constant in U_S and the instrument is not the policy.
- Carneiro, Heckman, and Vytlacil (2005) display the weights for all the treatment parameters reported in this section.

Evidence From the Instrumental Variables Literature

- Carneiro, Heckman, and Vytlacil (2005) report that $\hat{\beta}_{OLS} < \hat{\beta}_{IV}$.
- This finding is common in the literature (Card, 2001).
- At the same time, the returns to schooling are higher for individuals more likely to enroll in college, which means that the average return for the marginal individual is below the return for the average student in college.
- As previously explained and confirmed in the empirical work of Carneiro, Heckman, and Vytlacil (2005) reported here and in Cunha, Heckman, and Navarro (2005), this is possible because the conventional measure of selection bias $(E(U_0|S=1) - E(U_0|S=0))$ is negative and not positive, as is implicitly assumed in Card (1999, 2001) and in most of the empirical literature.

Evidence From the Instrumental Variables Literature

- In a model of heterogeneous returns, standard intuitions about instrumental variables and ability bias break down.
- Carneiro, Heckman, and Vytlacil (2005) confirm the conjecture of Willis and Rosen (1979).
- The evidence of Cunha, Heckman and Navarro (2005) shows that the single skill or efficiency units representation of the labor market which is implicit in most of the literature is invalid.

Evidence From the Instrumental Variables Literature

- Table 12, taken from the analysis of Heckman, Urzua, and Vytlacil (2004), demonstrates the sensitivity of *IV* estimates to the choice of instruments and to whether or not the estimates are conducted on samples where there is full support.
- As Figure 12 reveals, there are many intervals over which support is less than full, or very thin.
- In Table 12, for the full sample (first column) or the common support sample (second column), the *IV* estimates are all over the map. (Their estimates should be divided by 3.5 to get the annual returns to college reported in Carneiro, Heckman, and Vytlacil (2005).)

Table 12: Instrumental Variables Estimates
NLSY - HS Graduates and Four-Year College Graduates
Males at Age 30*

Instruments	Standard IV		IV-MTE (Common Support)§		
	Full Sample†	Common Support‡	Parametric	Polynomial	Nonparametric
Number of Siblings at 14	0.983	1.122	0.390	0.634	0.634
	(0.512)	(0.591)	(0.121)	(0.163)	(0.160)
Family income in 1979 (thousands)	1.667	1.803	0.416	0.590	0.612
	(0.432)	(0.630)	(0.121)	(0.143)	(0.147)
Local Wage of HS Graduates at County Level at age 17	94.600	41.400	0.407	0.591	0.618
	(1713.300)	(334.000)	(0.141)	(0.269)	(0.190)
Two Year Coll. Grad's local wage at age 17	5.008	5.394	0.426	0.600	0.622
	(4.077)	(4.941)	(0.135)	(0.216)	(0.188)
Four Year Coll. Grad's local wage at age 17	2.742	3.149	0.428	0.614	0.629
	(1.093)	(1.537)	(0.125)	(0.187)	(0.162)
Local Unemp. Rate of HS Graduates at County Level at age 17	0.675	0.612	0.203	0.523	0.526
	(0.604)	(0.675)	(0.442)	(33.651)	(15.377)
Two Year Coll. Grad's local unemployment rate at age 17	0.210	0.187	0.363	0.580	0.588
	(0.579)	(0.727)	(0.260)	(1.824)	(1.927)
Four Year Coll. Grad's local unemployment rate at age 17	3.465	4.480	0.405	0.554	0.588
	(13.476)	(11.586)	(0.144)	(0.294)	(0.220)
Distance to a two year college	3.369	4.603	0.416	0.606	0.621
	(3.223)	(6.282)	(0.139)	(0.186)	(0.165)
Distance to a four year college	4.810	7.440	0.415	0.629	0.634
	(5.180)	(12.150)	(0.120)	(0.170)	(0.161)
Two year college tuition	-2.637	-1.870	0.417	0.798	0.677
	(3.321)	(2.172)	(0.767)	(8.690)	(5.931)
Four Year college tuition	12.500	72.300	0.436	0.650	0.642
	(42.780)	(1465.920)	(1.248)	(0.729)	(0.604)
Propensity Score	0.496	0.505	0.420	0.572	0.604
	(0.093)	(0.103)	(0.121)	(0.138)	(0.143)

*We excluded the oversample of poor whites and the military sample. †The *IV* estimates and the standard deviations (in parentheses) are computed applying the traditional formulae to the full sample. The number of observations in our sample is 982. ‡The *IV* estimates and the standard deviations (in parentheses) are computed applying the traditional formulae to the common support sample. This sample contains only observations for which the estimated propensity score belongs to the common support of the propensity score between the control (HS graduates) and treatment group (4 year college graduates) (912 observations). §In the first column the *IV* estimates are computed by taking the weighted sum of the *MTE* estimated using the parametric approach. In the second column the *IV* estimates are computed by taking the weighted sum of the *MTE* estimated using a polynomial of degree 4 to approximate $E(Y|P)$. The *IV* estimates in the last column are computed by taking the weighted sum of the *MTE* estimated using the nonparametric approach. The propensity score ($\text{Prob}(D=1|Z=z)$) is computed using the instruments presented in the table as well as two dummy variables as controls for the place of residence at age 14 (south and urban), and a set of dummy variables controlling for the year of birth (1958-1963). The standard deviations (in parentheses) are obtained using bootstrapping (100 draws).

Source: Heckman, Urzua and Vytlacil (2004)

Evidence From the Instrumental Variables Literature

- The final three columns show the *IV* based on an estimated *MTE* using
  1. a parametric normal model (third column);
  2. a semiparametric polynomial estimation method and
  3. a nonparametric method based on local linear regression.
- The weights used to produce the *IV* estimates are given in Table 9b and are tailored to each estimation situation.
- There is close agreement between the two semiparametric methods and they are very different from the estimates in the third column that assume normality.
- The instability manifest in the numbers reported in the first two columns is reduced by using the *MTE*.
- But the instability is manifest in a number of studies in the literature.

Evidence From the Instrumental Variables Literature

- Table 13 shows estimates of the various treatment parameters based on the three versions of the *MTE*.
- There is a sharp contrast in the estimates produced from the parametric and nonparametric approaches.
- The different treatment parameters estimate different objects.
- The *LATE* estimators, defined for different points of evaluation $P(Z)$ (given by the arguments in parentheses) estimate very different numbers.

Table 13: Treatment Parameter Estimates
NLSY - HS Graduates and Four-Year College Graduates
Males at Age 30*

Treatment Parameter†	Parametric‡	Polynomial‡	Nonparametric§
Treatment on the Treated	0.362	0.758	0.696
	(0.123)	(0.201)	(0.181)
Treatment on the Untreated	0.509	0.687	0.652
	(0.149)	(0.142)	(0.167)
Average Treatment Effect	0.455	0.713	0.668
	(0.127)	(0.153)	(0.151)
$LATE$ $(0.62, 0.38)$	0.483	0.59	0.659
	(0.138)	(0.185)	(0.192)
$LATE$ $(0.79, 0.55)$	0.555	1.04	0.792
	(0.175)	(0.269)	(0.245)
$LATE$ $(0.45, 0.21)$	0.412	0.157	0.383
	(0.120)	(0.184)	(0.159)

*We excluded the oversample of poor whites and the military sample. †The treatment parameters are estimated by taking the weighted sum of the MTE estimated using a polynomial of degree 4 to approximate $E(Y|P)$. ‡The treatment parameters were estimated by taking the weighted sum of the MTE estimated using the parametric approach. §The treatment parameters were estimated by taking the weighted sum of the MTE estimated using the nonparametric approach. The standard deviations (in parentheses) are computed using bootstrapping (100 draws).
Source: Heckman, Urzua and Vytlacil (2006)

Evidence From the Instrumental Variables Literature

- Figures 15a and 15b from Heckman, Urzua, and Vytlacil (2004) graph the weights for the *MTE* for some of the instruments used to generate the numbers in Table 12.
- The weights for $P(Z)$ as an instrument are very different from the weights for four-year college tuition (Figure 15a) and especially two-year college tuition (Figure 15b).
- This accounts for why different instruments define different parameters in terms of their weighting of a common *MTE* function.
- It is the *MTE* function and not an *IV* estimate that plays the role of a policy invariant parameter in the modern literature on instrumental variables.

Figure 15a. IV Weights

Propensity Score vs Four year college tuition as the Instrument

NLSY − Sample of HS Graduates and Four Year College Graduates − Males at age 30

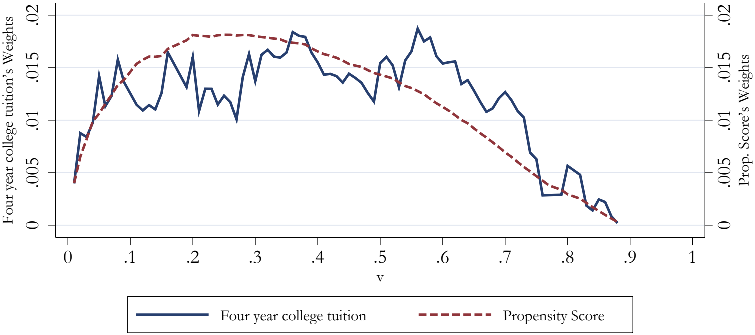


Figure 15b. IV Weights

Propensity Score vs Two year college tuition as the Instrument
NLSY − Sample of HS Graduates and Four Year College Graduates − Males at age 30

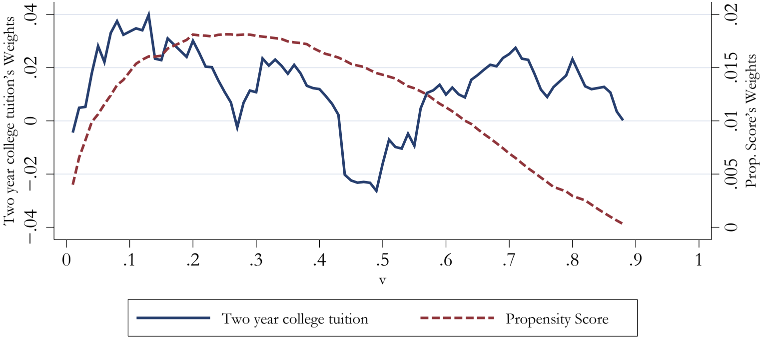


Figure 15: Legend

The dependent variable in the outcome equation is hourly earnings at age 30. The controls in the outcome equations are tenure, tenure squared, experience, corrected AFQT, black (dummy), hispanic (dummy), marital status, and years of schooling. Let $D = 0$ denote dropout status, and $D = 1$ denote GED status. The model for D (choice model) includes as controls the corrected AFQT, number of siblings, fathers education, mothers education, family income at age 17, local GED costs, broken home at age 14, average local wage at age 17 for dropouts and high school graduates, local unemployment rate at age 17 for dropouts and high school graduates, the dummy variables black and hispanics, and a set of dummy variables controlling for the year of birth. The choice model is estimated using a probit model. In computing the MTE, the bandwidth in the first step is selected using the leave-one-out cross-validation method. In the second step, following Heckman, Ichimura and Todd (1998), we set the bandwidth to 0.3. We use biweight kernel functions.

The Validity of the Conventional Instruments

- This section examines the validity of conventional instruments in the NLSY data which is unusually rich.
- Many data sets on earnings and schooling do not possess measures of cognitive ability.
- For example, the CPS and many other data sets used to estimate the returns to schooling surveyed in Katz and Autor (1999) do not report measures of cognitive ability.
- In this case, ability becomes part of U_1, U_0 and U_S instead of being in X.

The Validity of the Conventional Instruments

- The assumption of independence between the instrument and U_1 and U_0 implies that the instruments have to be independent of cognitive ability.
- However, the instruments that are commonly used in the literature are correlated with AFQT, a widely used measure of ability.
- The first column of Table 14a shows the coefficient of a regression of each instrument (Z) on college attendance (S), denoted by $\beta_{S,Z}$.
- With the exception of the local unemployment rate, all candidate instruments are strongly correlated with schooling.

Table 14a: Regression of Instrumental Variables (Z) on Schooling (S) and AFQT (A)

Instrumental Variable	$\beta_{S,Z}$	$\beta_{A,Z}$	F-Stat
Number of Siblings	-0.0302	-0.0468	15.04
	(0.0078)	(0.0141)	
Mother's Education	0.0760	0.1286	157.56
	(0.0060)	(0.0110)	
Father's Education	0.0582	0.0986	201.33
	(0.0041)	(0.0075)	
Average County Tuition at 17	-0.0062	-0.0044	13.32
	(0.0017)	(0.0031)	
Distance to College at 14	-0.0038	-0.0081	8.56
	(0.0013)	(0.0023)	
State Unemployment Rate at 17	-0.0052	-0.0038	0.42
	(0.0081)	(0.0148)	

Source: Carneiro, Heckman and Vytlacil (2005).

Table 14b: Residualized Regression of Instrumental Variables (Z) on Schooling (S) and AFQT (A)

Instrumental Variable	$\beta_{S,Z}$	$\beta_{A,Z}$	F-Stat
Average County Tuition at 17	-0.0041	-0.0009	6.81
	(0.0015)	(0.0029)	
Distance to College at 14	-0.0008	-0.0032	0.53
	(0.0012)	(0.0022)	
State Unemployment Rate at 17	-0.0027	0.0005	0.13
	(0.0075)	(0.0138)	

Source: Carneiro, Heckman and Vytlacil (2005).

The Validity of the Conventional Instruments

- The second column of this table presents the coefficient of a regression of each instrument on AFQT scores (A), denoted by $\beta_{A,Z}$.
- It shows that most of the candidates for instrumental variables in the literature are also correlated with cognitive ability.
- Therefore, in data sets where cognitive ability is not available most of these variables are not valid instruments since they violate the crucial *IV* assumption of independence.

The Validity of the Conventional Instruments

- Since few data sets have measures of cognitive ability, this finding calls into question much of the *IV* literature.
- Notice that the local unemployment rate is not strongly correlated with AFQT.
- However, it is only weakly correlated with college attendance.

The Validity of the Conventional Instruments

- The third column of Table 14a presents the F-statistic for the test of the hypothesis that the coefficient on the instrument is zero in a regression of schooling on the instrument.
- Staiger and Stock (1997) suggest using an F-statistic of 10 as a threshold for separating weak and strong instruments.
- The table shows that the local unemployment variable has an F statistic well below 10 which suggests that it is a weak instrument when used by itself.
- Therefore either the candidate instrumental variable is correlated with ability or it is weakly correlated with schooling.

The Validity of the Conventional Instruments

- Table 14b presents coefficients of regressions of each instrument on schooling and ability, after controlling for family background variables (number of siblings and parental education).
- Conditioning on family background weakens the correlation between AFQT and the instruments.
- However the F-test for a regression of schooling on the residualized instrument is low by Staiger-Stock standards.
- Residualizing on family background attenuates the correlation between the instruments and ability but also between the instruments and schooling.
- The strength of this dependence is reported in the third column of Table 14b.

The Validity of the Conventional Instruments

- The instrument used by Carneiro, Heckman, and Vytlacil (2005) is $P(Z)$.
- Regressing schooling on polynomials in experience, corrected AFQT, number of siblings, mother's education, father's education (the variables we include in the wage regression) and $P(Z)$, the F-statistic of the coefficient on P is 33.76.
- By including AFQT in the wage regression they attenuate the possibility of using invalid instruments.
- By using an index of instruments instead of a single instrument, it is possible to overcome the weak instrument problem.

The Validity of the Conventional Instruments

- Furthermore, using an index of instruments instead of a single instrument tends to reduce support problems for any instrument.
- Even if one instrument has limited support, other instruments can augment the support of P.
- Observe that the IV estimates based on $P(Z)$ are more stable in Table 12 than are the estimates based on the individual components.

Summary of the Modern Literature on Instrumental Variables

- Heckman and Vytlacil (2001a, 2005, 2007b) show how to write different conventional mean parameters and *IV* estimates as weighted averages of the marginal treatment effect (*MTE*).
- In a model with heterogeneous responses, different instruments define different parameters.
- Unless the instruments are the policies being studied, these parameters answer well-posed economic questions only by accident.
- It is possible to identify and estimate the *MTE* using a robust nonparametric selection model.

Summary of the Modern Literature on Instrumental Variables

- Their method allows them to combine diverse instruments into a scalar instrument motivated by economic theory.
- This combined instrument expands the support of any one instrument, and allows the analyst to perform out-of-sample policy forecasts.
- Focusing on a policy relevant question, they construct estimators based on the MTE to answer it, rather than hoping that a particular instrumental variable estimator happens to answer a question of economic interest.

Summary of the Modern Literature on Instrumental Variables

- The approach based on the *MTE* unites the selection and *IV* literatures.
- As noted by Heckman, Urzua and Vytlacil (2006), both methods use $P(Z)$ but one conditions on it (the selection model) while the other (the *IV* literature) does not.

Summary of the Modern Literature on Instrumental Variables

- The recent literature confirms in a semiparametric setting a central claim of the parametric Willis and Rosen (1979) analysis (Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2005).
- Individuals sort into schooling on the basis of both observed and unobserved gains where the observer is the economist analyzing the data.
- Moreover, as noted by Willis and Rosen (1979), it is not possible to rationalize labor market data with the single skill (or efficiency units) model that governs most of the standard intuitions about ability bias in schooling.
- In fact, these intuitions break down in a general model of heterogeneous returns, and lead to potentially wrong interpretations of the data.

Summary of the Modern Literature on Instrumental Variables

- Instrumental variables are not guaranteed to estimate policy relevant treatment parameters or conventional treatment parameters.
- Different instruments define different parameters, and in the empirical analysis of Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2006) they produce wildly different "effects" of schooling on earnings.

Summary of the Modern Literature on Instrumental Variables

- The current practice of reporting *IV* estimates as "returns" to schooling defines the parameter being identified by an econometric method and not by an economic question.
- Our examples show that the *IV* method does not produce an economically interesting or interpretable parameter, and in general does not estimate a rate of return.
- Different *IV* estimators weight the *MTE* differently and are not comparable in their economic content.

Summary of the Modern Literature on Instrumental Variables

- Even granting the validity and the strength of the instruments, the entire recent *IV* enterprise for correlated random coefficient models is premised on a fundamental asymmetry.
- Returns (growth rates) are allowed to be heterogeneous in a general way.
- Schooling may either increase or decrease rates of return.
- However, choices are not permitted to be heterogeneous in a general way (Heckman and Vytlacil, 2005; Heckman, Urzua, and Vytlacil, 2006).

Summary of the Modern Literature on Instrumental Variables

- The monotonicity assumptions (or index structure assumptions embodied in (21) or (23) so that schooling is determined by an index of "net utility" where the observables are separable from the nonobservables) impose the condition that all persons respond in the same way in their schooling choices for any change in Z.
- Thus if increasing a coordinate of Z, say Z_1, increases schooling for one person, the same increase cannot decrease schooling for anyone else.

Summary of the Modern Literature on Instrumental Variables

- This condition rules out heterogeneity in the choice equations.
- These conditions are at odds with a variety of economic models for schooling such as models for dynamic discrete choice (see Heckman and Navarro, 2006).
- See Belzil and Hansen (2005) for an interesting contrast between *IV* and structural estimates of returns to schooling.
- Their structural models and those of Heckman and Navarro do not impose monotonicity conditions on the choice data.

Summary of the Modern Literature on Instrumental Variables

- If the monotonicity conditions are violated, increases in Z_1 may increase participation in schooling for some and decrease it for others.
- In this case, instrumental variables methods do not estimate treatment effects and the local instrumental variable does not identify the marginal treatment effect.
- See Heckman and Vytlacil (2001c, 2005) and Heckman, Urzua, and Vytlacil (2006) for further discussion of this point.

Part IX

Estimating Distributions of Returns to Schooling

Estimating Distributions of Returns to Schooling

- Following the representative agent tradition, economists usually summarize the distribution of the growth rate of earnings with schooling by some mean.
- Earlier, we presented a variety of mean treatment effects which are defined by the conditioning variables used.
- Different means answer different policy questions.

Estimating Distributions of Returns to Schooling

- The research reported in this section (based on Aakvik, Heckman, and Vytlacil, 2005; Heckman, Smith, and Clements, 1997; Carneiro, Hansen, and Heckman, 2001, 2003; Cunha, Heckman, and Navarro, 2005, 2006; and Cunha and Heckman, 2006b) moves beyond means as descriptions of policy outcomes and considers joint counterfactual distributions of outcomes (for example, $F(Y_1, Y_0)$, gains $F(Y_1 - Y_0)$ or $F(Y_1, Y_0|S = 1)$).
- These are *ex post* distributions realized after schooling decisions are completed.
- We analyze *ex ante* distributions in the next section.

Estimating Distributions of Returns to Schooling

- From knowledge of the *ex post* joint distributions of counterfactual outcomes, it is possible to determine the proportion of people who benefit or lose from schooling, the origin and destination outcomes of those who change status because of schooling and the amount of gain (or loss) from various policy choices such as tuition subsidies by persons at different deciles of an initial prepolicy income distribution.

Estimating Distributions of Returns to Schooling

- Using the joint distribution of counterfactuals, it is possible to develop a more nuanced understanding of the distributional impacts of public policies directed toward education, and to move beyond comparisons of aggregate distributions induced by different policies to consider how people in different portions of an initial distribution are affected by public policy.

Estimating Distributions of Returns to Schooling

- From knowledge of the mean treatment effects presented earlier, if $Y_1 - Y_0$ varies in the population, it is not possible to answer the simple question of who benefits from schooling and the proportion of people benefiting, except in the special case where everyone with the same X receives the same benefit.
- Our methods can be used to explain effects of schooling (and other interventions) on earnings, employment and health.
- In this chapter, we focus on earnings measures.

Estimating Distributions of Returns to Schooling

- Under the assumptions presented above, joint distributions of counterfactuals are not identified non-parametrically (see Heckman, 1990).
- We observe Y_1 or Y_0 for the same person but not both.
- Thus it is not possible to use cross section data to tabulate the joint distribution of (Y_0, Y_1) from the raw data.
- However, with additional information, it is possible.

Estimating Distributions of Returns to Schooling

- More precisely, an agent can experience one of two possible counterfactual schooling levels with associated outcomes (Y_0, Y_1).
- As before, we denote X as determinants of the counterfactual outcomes (Y_0, Y_1); $S = 1$ if the agent is in state 1; $S = 0$ otherwise.
- The observed outcome is $Y = SY_1 + (1 - S)Y_0$.
- Let Z be a determinant of S that does not affect Y_1, Y_0.
- The standard treatment effect model analyzed earlier and in this section considers policies that shift Z and that affect choices of treatment but not potential outcomes (Y_0, Y_1).
- It ignores general equilibrium effects.

Estimating Distributions of Returns to Schooling

- The goal is to recover $F(Y_0, Y_1 \mid X)$ and hence $F(Y_1 - Y_0 \mid X)$, and related distributions such as those for gross gains $\left(\frac{1}{1+r} Y_1 - Y_0\right)$ or net gains $\left(\frac{Y_1}{1+r} - Y_0 - C\right)$ assuming one period of foregone earnings is required to move from "0" to "1".

Estimating Distributions of Returns to Schooling

- The problem of recovering joint distributions from cross section data has two aspects.
- The first is the selection problem.
- From data on the distribution of earnings by schooling and characteristics X, $F(Y_1 \mid S = 1, X)$ and $F(Y_0 \mid S = 0, X)$, under what conditions can one recover $F(Y_1 \mid X)$ and $F(Y_0 \mid X)$, respectively?
- The second problem is how to construct the joint distribution $F(Y_0, Y_1 \mid X)$ from the two marginal distributions of earnings for each secondary schooling level.

Estimating Distributions of Returns to Schooling

- If the selection problem can be solved and the marginal distributions of Y_1 and Y_0 are identified, results from probability theory due to Fréchet (1951) and Hoeffding (1940) can be used to bound $F(Y_1, Y_0 \mid S, X)$ from the marginal distributions.
- In practice these bounds are often very wide, and the inferences based on the bounding distributions are often not very helpful.

Estimating Distributions of Returns to Schooling

- A second approach, based on matching, postulates access to variables Q that have the property that conditional on Q, $F(Y_0 \mid S=0, X, Q) = F(Y_0 \mid X, Q)$ and $F(Y_1 \mid S=1, X, Q) = F(Y_1 \mid X, Q)$.
- Matching thus assumes that conditional on observed variables, there is no selection problem.
- If it is further assumed that all of the dependence between (Y_0, Y_1) given X comes through Q, then it follows that $F(Y_1, Y_0 \mid X, Q) = F(Y_1 \mid X, Q)\, F(Y_0 \mid X, Q)$.
- Using these results, it is possible to create the joint distribution $F(Y_0, Y_1 \mid X)$ because

$$F(Y_0, Y_1 \mid X) = \int F(Y_0 \mid X, Q)\, F(Y_1 \mid X, Q)\, d\mu(Q \mid X).$$

Estimating Distributions of Returns to Schooling

- $\mu(Q \mid X)$ is the conditional distribution of Q given X.
- We obtain $F(Y_0 \mid X, Q), F(Y_1 \mid X, Q)$ by matching.
- We know the distribution of Q given X because we observe Q and X.
- Thus we can construct the right hand side of this expression.
- Matching makes the strong assumption that conditional on (Q, X) the marginal return to schooling is the same as the average return.

Estimating Distributions of Returns to Schooling

- One traditional approach in economics assumes that the joint distribution $F(Y_0, Y_1 \mid X)$ is a degenerate one dimensional distribution.
- It assumes that conditional on X, Y_1 and Y_0 are deterministically related,

$$Y_1 = Y_0 + \Delta \tag{29}$$

where Δ is the difference in means between Y_1 and Y_0 for the selection corrected distribution.

- This assumes that schooling has the same effect on everyone (with the same X) and that effect is Δ.

Estimating Distributions of Returns to Schooling

- Heckman and Smith (1998) and Heckman, Smith, and Clements (1997) relax this assumption by assuming perfect ranking in the positions of individuals $F(Y_1 \mid X)$ and $F(Y_0 \mid X)$ distributions.
- The best in one distribution is the best in the other.
- Assuming continuous and strictly increasing marginal distributions, they postulate that quantiles are perfectly ranked so $Y_1 = F_1^{-1}(F_0(Y_0))$ where $F_1 = F_1(y_1 \mid X)$ and $F_0 = F_0(y_0 \mid X)$.

Estimating Distributions of Returns to Schooling

- This assumption generates a deterministic relationship which turns out to be the tight upper bound of the Fréchet bounds.
- An alternative assumption is that people are perfectly inversely ranked so the best in one distribution is the worst in the other: $Y_1 = F_1^{-1}(1 - F_0(Y_0))$.
- This is the tight Fréchet lower bound.

Estimating Distributions of Returns to Schooling

- A perfect ranking (or perfect inverse ranking) assumption generalizes the perfect-ranking, constant-shift assumptions implicit in the conventional literature.
- It allows analysts to apply conditional quantile methods to estimate the distributions of gains.
- However, it imposes a strong and arbitrary dependence across distributions.
- When the perfect ranking assumption is relaxed and tested, it is rejected.

Estimating Distributions of Returns to Schooling

- A more general framework attacks this problem in a different way than does matching or invoking special assumptions about relationships between the ranks of persons in the Y_0 and Y_1 distribution.
- This line of research starts from the analysis of Heckman (1990); Heckman and Smith (1998); Aakvik, Heckman, and Vytlacil (2005); Carneiro, Hansen, and Heckman (2001, 2003); Cunha, Heckman, and Navarro (2005, 2006); and Cunha and Heckman (2006a).

Estimating Distributions of Returns to Schooling

- In this chapter we draw on the analysis of Carneiro, Hansen, and Heckman (2003).
- They start with the marginal distributions of Y_1 and of Y_0 given X.
- They allow for unobservables to generate the joint dependence and do not rely on matching.

Estimating Distributions of Returns to Schooling

- The basic idea is to restrict the dependence among the (U_0, U_1, U_S) by factor models or other restrictions.
- A low dimensional set of random variables generates the dependence across the unobservables.
- Such dimension reduction coupled with use of the choice data and measurements that proxy components of the (U_0, U_1, U_S), provides enough information to identify the joint distribution of (Y_1, Y_0) and of (Y_1, Y_0, S).

Estimating Distributions of Returns to Schooling

- Assume separability between unobservables and observables and that Y_1 and Y_0 are lifetime earnings:

$$\begin{aligned} Y_1 &= \mu_1(X) + U_1 \\ Y_0 &= \mu_0(X) + U_0. \end{aligned}$$

- Denote S^* as the latent variable generating schooling choices:

$$\begin{aligned} S^* &= \mu_S(Z) + U_S \\ S &= \mathbf{1}(S^* \geq 0). \end{aligned}$$

- Recall that we allow any X to be in Z.

Estimating Distributions of Returns to Schooling

- To motivate the approach, assume that (U_0, U_1, U_S) is normally distributed with mean zero and covariance matrix Σ_G ("G" for Generalized Roy).
- If the distributions are normal, they can be fully characterized by means and covariances.
- To simplify the discussion, we focus our exposition on normal models although that is not essential.
- We assume that (U_0, U_1, U_S) are statistically independent of (X, Z).

Estimating Distributions of Returns to Schooling

- Under normality, standard results in the selection bias literature show that from data on Y_1 given $S = 1$, and X, and data on Y_0 for $S = 0$ and X, and data on choices of schooling given Z, one can identify $\mu_1(X)$, $\mu_0(X)$ and $\mu_S(Z)$, the latter up to scale σ_S (where $\sigma_S^2 = \text{Var}(U_S)$).
- See Heckman (1976) or Cunha, Heckman, and Navarro (2005).
- In addition, one can identify the joint densities of $(U_0, U_S/\sigma_S)$ and $(U_1, U_S/\sigma_S)$.
- Without further information, one cannot identify the joint density of $(U_0, U_1, U_S/\sigma_S)$.

Estimating Distributions of Returns to Schooling

- Recent developments in microeconometrics show that analysts can identify these same objects without a normality assumption provided that there are variables Z that generate enough variation in $\mu_S(Z)$.
- The intuition for why variation identifies the model is presented in Heckman and Honoré (1990), Heckman (1990) and Cunha, Heckman, and Navarro (2006).
- If Z has sufficient variation, there are limit sets where $P(Z) = 1$ and other sets where $P(Z) = 0$ so there is no selection problem in those limit sets.

Estimating Distributions of Returns to Schooling

- Formal proofs and general conditions are given in Carneiro, Hansen, and Heckman (2003).
- Normality plays no central role in the analysis of this section.
- We use it because it is familiar in the economics of education due to the application of the Generalized Roy model by Willis and Rosen (1979).

Estimating Distributions of Returns to Schooling

- To get the gist of the method underlying recent work, we adopt a factor structure model for the U_0, U_1, U_S.
- Other restrictions across the unobservables are possible (see Urzua, 2005).
- Factor models are extensively developed by Jöreskog and Goldberger (1975).
- Aakvik, Heckman, and Vytlacil (2005) and Carneiro, Hansen, and Heckman (2001, 2003) apply their analysis to generate counterfactuals.

Estimating Distributions of Returns to Schooling

- For simplicity, we assume a one factor model where θ is the factor that generates dependence across the unobservables:

$$
\begin{aligned}
U_0 &= \alpha_0\theta + \varepsilon_0 \\
U_1 &= \alpha_1\theta + \varepsilon_1 \\
U_S &= \alpha_S\theta + \varepsilon_S .
\end{aligned}
$$

- We assume $E(U_0) = 0, E(U_1) = 0, E(U_S) = 0$.
- In addition, $E(\theta) = 0$, $E(\varepsilon_0) = 0$, $E(\varepsilon_1) = 0$ and $E(\varepsilon_S) = 0$.

Estimating Distributions of Returns to Schooling

- To set the scale of the unobserved factor, we can normalize one "loading" (coefficient on θ) to 1.
- Other normalizations are possible.
- We assume that θ is a scalar factor (say unmeasured ability) and the $(\varepsilon_0, \varepsilon_1, \varepsilon_S)$ are independent of θ and of each other.
- All the dependence across the unobservables arises from θ.

Estimating Distributions of Returns to Schooling

- Under normality or from the general semiparametric identification analysis of Carneiro, Hansen, and Heckman (2003), we can identify

$$\begin{aligned} \operatorname{Cov}\left(U_0, \frac{U_S}{\sigma_S}\right) &= \frac{\alpha_0 \alpha_S}{\sigma_S}\sigma_\theta^2 \\ \operatorname{Cov}\left(U_1, \frac{U_S}{\sigma_S}\right) &= \frac{\alpha_1 \alpha_S}{\sigma_S}\sigma_\theta^2. \end{aligned}$$

- From the ratio of the second covariance to the first we obtain $\frac{\alpha_1}{\alpha_0}$, assuming $\alpha_0 \neq 0$.
- Thus we obtain the sign of the dependence between U_0, U_1 because

$$\operatorname{Cov}(U_0, U_1) = \alpha_0 \alpha_1 \sigma_\theta^2.$$

Estimating Distributions of Returns to Schooling

- From the ratio, we obtain α_1 if we normalize $\alpha_0 = 1$.
- Without further information, we can only identify the variance of U_S up to scale, which can be normalized to 1.
- Alternatively, we could normalize the variance of ε_S to 1.
- Below, we present a condition that sets the scale of U_S.

Estimating Distributions of Returns to Schooling

- Knowledge of the sign of $\dfrac{\alpha_1}{\alpha_0}$ is informative on the sign of the correlation between college and high school skills, a key unanswered question in the analysis of Willis and Rosen (1979).
- They conjecture that $\text{Cov}\left(U_0, U_1\right) < 0$.
- The evidence reported in Carneiro, Hansen, and Heckman (2001, 2003), Cunha, Heckman, and Navarro (2005, 2006), and Cunha and Heckman (2006a) supports their conjecture.
- Those with high levels of U_1 have lower levels of U_0.

Estimating Distributions of Returns to Schooling

- With additional information, we can identify the full joint distribution.
- We now present some examples.
- Cunha, Heckman, and Navarro (2005) present a more comprehensive analysis.

Access to a single test score

- Assume access to data on Y_0 given $S = 0$, X, Z; to data on Y_1 given $S = 1$, X, Z; and data on S given X, Z.
- Suppose that the analyst also has access to a single test score T that is a proxy for θ,

$$T = \mu_T(X) + U_T$$

where $U_T = \alpha_T \theta + \varepsilon_T$ so

$$T = \mu_T(X) + \alpha_T \theta + \varepsilon_T,$$

where ε_T is independent of ε_0, ε_1, ε_S and (X, Z).

Access to a single test score

- We can identify the mean $\mu_T(X)$ from observations on T and X.
- We pick up three additional covariance terms, conditional on X, Z :

$$
\begin{aligned}
\operatorname{Cov}(Y_1, T) &= \alpha_1 \alpha_T \sigma_\theta^2, \\
\operatorname{Cov}(Y_0, T) &= \alpha_0 \alpha_T \sigma_\theta^2, \\
\operatorname{Cov}(S^*, T) &= \frac{\alpha_S}{\sigma_S} \alpha_T \sigma_\theta^2.
\end{aligned}
$$

- To simplify the notation we keep the conditioning on X and Z implicit.
- Suppose that we normalize the loading on the test score to one ($\alpha_T = 1$).

Access to a single test score

- It is no longer necessary to normalize $\alpha_0 = 1$ as in the preceding section.
- From the ratio of the covariance of Y_1 with S^* with the covariance of S^* with T, we obtain the left hand side of

$$\frac{\text{Cov}\left(Y_1, S^*\right)}{\text{Cov}\left(S^*, T\right)} = \frac{\alpha_1 \alpha_S \sigma_\theta^2}{\alpha_S \alpha_T \sigma_\theta^2} = \alpha_1,$$

because $\alpha_T = 1$ (normalization).

Access to a single test score

- From the preceding argument without the test score, we obtain α_0 since
$$\frac{\text{Cov}\,(Y_1, S^*)}{\text{Cov}\,(Y_0, S^*)} = \frac{\alpha_1 \alpha_S \sigma_\theta^2}{\alpha_0 \alpha_S \sigma_\theta^2} = \frac{\alpha_1}{\alpha_0}.$$
- From knowledge of α_1 and α_0 and the normalization for α_T, we obtain σ_θ^2 from $\text{Cov}\,(Y_1, T)$ or $\text{Cov}\,(Y_0, T)$.
- We obtain α_S (up to scale σ_S) from $\text{Cov}\,(S^*, T) = \alpha_S \alpha_T \sigma_\theta^2$ since we know α_T $(= 1)$ and σ_θ^2.
- The model is overidentified.
- We can set the scale of σ_S by a standard argument from the discrete choice literature.
- See the discussion below.

Access to a single test score

- Observe that if we write out the decision rule for schooling in terms of costs, we can characterize the latent variable determining schooling choices as:

$$S^* = Y_1 - Y_0 - C,$$

  where $C = \mu_C(Z) + U_C$ and $U_C = \alpha_C\theta + \varepsilon_C$, where ε_C is independent of θ and the other ε's.
- $E\left(U_C\right) = 0$ and U_C is independent of (X, Z).
- Then,

$$\begin{aligned}
\alpha_S &= \alpha_1 - \alpha_0 - \alpha_C \\
\varepsilon_S &= \varepsilon_1 - \varepsilon_0 - \varepsilon_C \\
\text{Var}\left(\varepsilon_S\right) &= \text{Var}\left(\varepsilon_1\right) + \text{Var}\left(\varepsilon_0\right) + \text{Var}\left(\varepsilon_C\right).
\end{aligned}$$

Access to a single test score

- Identification of α_0, α_1 and α_S implies identification of α_C.
- Identification of the variance of ε_S implies identification of the variance of ε_C since the variances of ε_1 and ε_0 are known.

Access to a single test score

- Observe further that the scale σ_{U_S} is identified if there are variables in X but not in Z (see Heckman, 1976, 1979; Heckman and Robb, 1985, 1986; Willis and Rosen, 1979). From the variance of T given X, we obtain $\text{Var}(\varepsilon_T)$ since we know $\text{Var}(T)$ (conditional on X) and we know $\alpha_T^2\sigma_\theta^2$:

$$\text{Var}(T) - \alpha_T^2\sigma_\theta^2 = \sigma_{\varepsilon_T}^2.$$

- Recall that we keep the conditioning on X implicit.
- By similar reasoning, it is possible to identify $\text{Var}(\varepsilon_0)$, $\text{Var}(\varepsilon_1)$ and the fraction of $\text{Var}(U_S)$ due to ε_S.
- We can thus construct the joint distribution of (Y_0, Y_1, C) since we know $\mu_C(Z)$ and all of the factor loadings.

Access to a single test score

- We have assumed normality because it is convenient.
- Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006a) show that it is possible to nonparametrically identify the distributions of θ, ε_0, ε_1, ε_S and ε_T so these results do not hinge on arbitrary distributional assumptions.

Access to a single test score

- There are other ways to construct the joint distributions that do not require a test score.
- Access to panel data on earnings affords identification.
- One way, which leads into our analysis below of *ex ante* vs. *ex post* returns, is discussed next.

Two (or more) periods of panel data on earnings

- Suppose that for each person we have two periods of earnings data in one counterfactual state or the other.
- We write

$$\begin{aligned} Y_{1t} &= \mu_{1t}(X) + \alpha_{1t}\theta + \varepsilon_{1t} \qquad t = 1, 2 \\ Y_{0t} &= \mu_{0t}(X) + \alpha_{0t}\theta + \varepsilon_{0t} \qquad t = 1, 2. \end{aligned}$$

- We observe one or the other lifecycle stream of earnings for each person, but never both streams for the same person.
- We assume that the interest rate is zero and that agents maximize the present value of their income.

Two (or more) periods of panel data on earnings

- Thus in terms of the index

$$S^{*} = (Y_{12} + Y_{11}) - (Y_{02} + Y_{01}) - C$$

$$S = 1(S^{*} \geq 0)$$

where C was defined previously.

- We assume no test score – just two periods of panel data.

Two (or more) periods of panel data on earnings

- Under normality, application of the standard normal selection model allows us to identify $\mu_{1t}(X)$ for $t = 1, 2$; $\mu_{0t}(X)$ for $t = 1, 2$ and $\mu_{11}(X) + \mu_{12}(X) - \mu_{01}(X) - \mu_{02}(X) - \mu_C(X)$, the latter up to a scalar σ_S, the standard deviation of

$$U_S = (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\theta + \varepsilon_{11} + \varepsilon_{12} - \varepsilon_{01} - \varepsilon_{02} - \varepsilon_C.$$

- Following our discussion of 536, we can recover the scale if there are variables in $(\mu_{11}(X) + \mu_{12}(X) - (\mu_{01}(X) + \mu_{02}(X)))$ not in $\mu_C(Z)$.
- For simplicity we assume that this condition holds.

Two (or more) periods of panel data on earnings

- From normality, we can recover the joint distributions of (S^*, Y_{11}, Y_{12}) and (S^*, Y_{01}, Y_{02}) but not directly the joint distribution of $(S^*, Y_{11}, Y_{12}, Y_{01}, Y_{02})$.
- Thus, conditioning on X and Z we can recover the joint distribution of (U_S, U_{01}, U_{02}) and (U_S, U_{11}, U_{12}) but apparently not that of $(U_S, U_{01}, U_{02}, U_{11}, U_{12})$.
- However, under our factor structure assumptions this joint distribution can be recovered as we next show.

Two (or more) periods of panel data on earnings

- From the available data, we can identify the following covariances:

$$
\begin{aligned}
\text{Cov}(U_S, U_{12}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{12}\sigma_\theta^2 \\
\text{Cov}(U_S, U_{11}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{11}\sigma_\theta^2 \\
\text{Cov}(U_S, U_{01}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{01}\sigma_\theta^2 \\
\text{Cov}(U_S, U_{02}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{02}\sigma_\theta^2 \\
\text{Cov}(U_{11}, U_{12}) &= \alpha_{11}\alpha_{12}\sigma_\theta^2 \\
\text{Cov}(U_{01}, U_{02}) &= \alpha_{01}\alpha_{02}\sigma_\theta^2.
\end{aligned}
$$

Two (or more) periods of panel data on earnings

- If we normalize $\alpha_{01} = 1$ (recall that one normalization is needed to set the scale of θ), we can form the ratios

$$\frac{\text{Cov}(U_S, U_{12})}{\text{Cov}(U_S, U_{01})} = \alpha_{12} \qquad \frac{\text{Cov}(U_S, U_{11})}{\text{Cov}(U_S, U_{01})} = \alpha_{11}$$

$$\frac{\text{Cov}(U_S, U_{02})}{\text{Cov}(U_S, U_{01})} = \alpha_{02}.$$

Two (or more) periods of panel data on earnings

- From these coefficients and the remaining covariances, we identify σ_θ^2 using $\text{Cov}(U_{11}, U_{12})$ and/or $\text{Cov}(U_{01}, U_{02})$.
- Thus if the factor loadings are nonzero,

$$\frac{\text{Cov}(U_{11}, U_{12})}{\alpha_{11}\alpha_{12}} = \sigma_\theta^2$$

and

$$\frac{\text{Cov}(U_{01}, U_{02})}{\alpha_{01}\alpha_{02}} = \sigma_\theta^2.$$

Two (or more) periods of panel data on earnings

- We can recover σ_θ^2 (since we know $\alpha_{11}\alpha_{12}$ and $\alpha_{01}\alpha_{02}$) from $\mathrm{Cov}(U_{11}, U_{12})$ and $\mathrm{Cov}(U_{01}, U_{02})$.
- We can also recover α_C since we know σ_θ^2, $\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C$, and α_{11}, α_{12}, $\alpha_{01,}$ α_{02}.
- We can form (conditional on X) $\mathrm{Cov}(Y_{11}, Y_{01}) = \alpha_{11}\alpha_{01}\sigma_\theta^2$; $\mathrm{Cov}(Y_{12}, Y_{01}) = \alpha_{12}\alpha_{01}\sigma_\theta^2$; $\mathrm{Cov}(Y_{11}, Y_{02}) = \alpha_{11}\alpha_{02}\sigma_\theta^2$ and $\mathrm{Cov}(Y_{12}, Y_{02}) = \alpha_{12}\alpha_{02}\sigma_\theta^2$.
- Thus we can identify the joint distribution of $(Y_{01}, Y_{02}, Y_{11}, Y_{12}, C)$ since we can identify $\mu_C(Z)$ from the schooling choice equation since we know $\mu_{01}(X)$, $\mu_{02}(X)$, $\mu_{11}(X)$, $\mu_{12}(X)$ and we have assumed that there are some Z not in X so that σ_S is identified.

Two (or more) periods of panel data on earnings

- This analysis can be generalized to a general nonnormal setting using the analysis of Carneiro, Hansen, and Heckman (2003).
- For simplicity, we have worked with a one factor model.
- The analyses of Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2005, 2006), Cunha and Heckman (2006a), and Heckman and Navarro (2006) use multiple factors.
- We offer an example in the next section.

Two (or more) periods of panel data on earnings

- The key idea in constructing joint distributions of counterfactuals using the analysis of Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006a) is *not* the factor structure for unobservables although it is convenient.
- The motivating idea is the assumption that a low dimensional set of random variables generates the dependence across outcomes.
- Other low dimensional representations such as the ARMA model or the dynamic factor structure model (see Sargent and Sims, 1977) can also be used.
- Urzua (2005) develops such a model and applies it to estimating rates of returns to schooling.

Two (or more) periods of panel data on earnings

- The factor structure model presented in this section is easy to exposit and has been used to estimate joint distributions of counterfactuals.
- We present some examples in the next section.
- That section reviews recent work that generalizes the analysis of this section to derive *ex ante* and *ex post* outcome distributions, and measure the fundamental uncertainty facing agents in the labor market.
- With these methods it is possible to compute the distributions of both *ex ante* and *ex post* rates of return to schooling.

Part X

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- In computing *ex ante* returns to schooling, it is necessary to characterize what is in the agent's information set at the time schooling decisions are made.
- To do so, the recent literature exploits the key idea that if agents know something and use that information in making their schooling decisions, it will affect their schooling choices.
- With panel data on earnings and other measurements of the factors, which may be test scores or information on other choices, we can assess what components of those outcomes were known at the time schooling choices were made.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- The literature on panel data earnings dynamics (e.g. Lillard and Willis, 1978; MaCurdy, 1982) is not designed to estimate what is in agent information sets.
- It estimates earnings equations of the following type:

$$Y_{i,t} = X_{i,t}\beta + S_i\tau + U_{i,t}, \tag{30}$$

  where $Y_{i,t}, X_{i,t}, S_i, U_{i,t}$ denote (for person i at time t) the realized earnings, observable characteristics, educational attainment, and unobservable characteristics, respectively, from the point of view of the observing economist.
- The variables generating outcomes realized at time t may or may not have been known to the agents at the time they made their schooling decisions.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- The error term $U_{i,t}$ is usually decomposed into two or more components.
- For example, it is common to specify that

$$U_{i,t} = \phi_i + \delta_{i,t}. \tag{31}$$

- The term ϕ_i is a person-specific effect.
- The error term $\delta_{i,t}$ is often assumed to follow an ARMA (p, q) process (see Hause, 1980; MaCurdy, 1982) such as $\delta_{i,t} = \rho\delta_{i,t-1} + m_{i,t}$, where $m_{i,t}$ is a mean zero innovation independent of $X_{i,t}$ and the other error components.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- The components $X_{i,t}, \phi_i$, and $\delta_{i,t}$ all contribute to measured *ex post* variability across persons.
- However, the literature is silent about the difference between heterogeneity and uncertainty, the unforecastable part of earnings as of a given age.
- The literature on income mobility and on inequality measures all variability *ex post* as in Chiswick (1974), Mincer (1974) and Chiswick and Mincer (1972).

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- An alternative specification of the error process postulates a factor structure for earnings that uses the representation:

$$U_{i,t} = \theta_i \alpha_t + \varepsilon_{i,t}, \tag{32}$$

where θ_i is a vector of skills (e.g. ability, initial human capital, motivation, and the like), α_t is a vector of skill prices, and the $\varepsilon_{i,t}$ are mutually independent mean zero shocks independent of θ_i.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- Hause (1980) and Heckman and Scheinkman (1987) analyze such earnings models.
- Any process in the form of equation (31) can be written in terms of (32).
- The latter specification is more directly interpretable as a pricing equation than (31).

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- Depending on the available market arrangements for coping with risk, the predictable components of $U_{i,t}$ will have a different effect on choices and economic welfare than the unpredictable components, if people are risk averse and cannot fully insure against uncertainty.
- Statistical decompositions based on (30), (31), and (32) or versions of them describe *ex post* variability but tell us nothing about which components of (30) or (32) are forecastable by agents *ex ante*.
- Is ϕ_i unknown to the agent? $\delta_{i,t}$? Or $\phi_i + \delta_{i,t}$? Or $m_{i,t}$?
- In representation (32), the entire vector θ_i, components of the θ_i, the $\varepsilon_{i,t}$, or all of these may or may not be known to the agent at the time schooling choices are made.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- The methodology developed in Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2005) and Cunha and Heckman (2006a,b) provides a framework within which it is possible to identify components of life cycle outcomes that are forecastable and acted on at the time decisions are taken from ones that are not.
- In order to choose between high school and college, agents forecast future earnings (and other returns and costs) for each schooling level.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- Using information about educational choices at the time the choice is made, together with the *ex post* realization of earnings and costs that are observed at later ages, it is possible to estimate and test which components of future earnings and costs are forecast by the agent.
- This can be done provided we know, or can estimate, the earnings of agents under both schooling choices and provided we specify the market environment under which they operate as well as their preferences over outcomes.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- For market environments where separation theorems are valid, so that consumption decisions are made independently of wealth maximizing decisions, it is not necessary to know agent preferences to decompose realized earnings outcomes in this fashion.
- Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2005) and Cunha and Heckman (2006a,b) use choice information to extract *ex ante* or forecast components of earnings and to distinguish them from realized earnings under different market environments.

Ex Ante and Ex Post Returns: Distinguishing Heterogeneity from Uncertainty

- The difference between forecast and realized earnings allows them to identify the distributions of the components of uncertainty facing agents at the time they make their schooling decisions.

A Generalized Roy Model

- To state these issues more precisely, consider a version of the generalized Roy (1951) economy with two sectors.
- This builds on the second example in the section on Estimating Distributions of Returns to Schooling.
- Let S_i denote different schooling levels.
- $S_i = 0$ denotes choice of the high school sector for person i, and $S_i = 1$ denotes choice of the college sector.
- Each person chooses to be in one or the other sector but cannot be in both.

A Generalized Roy Model

- Let the two potential outcomes be represented by the pair $(Y_{0,i}, Y_{1,i})$, only one of which is observed by the analyst for any agent.
- Denote by C_i the direct cost of choosing sector 1, which is associated with choosing the college sector (e.g., tuition and non-pecuniary costs of attending college expressed in monetary values).
- We have used this framework throughout this chapter.

A Generalized Roy Model

- $Y_{1,i}$ is the *ex post* present value of earnings in the college sector, discounted over horizon T for a person choosing at a fixed age, assumed for convenience to be zero,

$$Y_{1,i} = \sum_{t=0}^{T} \frac{Y_{1,i,t}}{(1+r)^t},$$

and $Y_{0,i}$ is the *ex post* present value of earnings in the high school sector at age zero,

$$Y_{0,i} = \sum_{t=0}^{T} \frac{Y_{0,i,t}}{(1+r)^t},$$

where r is the one-period risk-free interest rate.

A Generalized Roy Model

- $Y_{1,i}$ and $Y_{0,i}$ can be constructed from time series of *ex post* potential earnings streams in the two states: $(Y_{0,i,0}, \ldots, Y_{0,i,T})$ for high school and $(Y_{1,i,0}, \ldots, Y_{1,i,T})$ for college.
- A practical problem is that we only observe one or the other of these streams.
- This partial observability creates a fundamental identification problem which can be solved using the methods described above.

A Generalized Roy Model

- The variables $Y_{1,i}$, $Y_{0,i}$, and C_i are *ex post* realizations of returns and costs, respectively.
- At the time agents make their schooling choices, these may be only partially known to the agent, if at all.
- Let $\mathcal{I}_{i,0}$ denote the information set of agent i at the time the schooling choice is made, which is time period $t = 0$ in our notation.

A Generalized Roy Model

- Under a complete markets assumption with all risks diversifiable (so that there is risk-neutral pricing) or under a perfect foresight model with unrestricted borrowing or lending but full repayment, the decision rule governing sectoral choices at decision time '0' is

$$S_i = \begin{cases} 1, & \text{if } E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right) \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

- Under perfect foresight, the postulated information set would include $Y_{1,i}$, $Y_{0,i}$, and C_i.

A Generalized Roy Model

- Under either model of information, the decision rule is simple: one attends school if the expected gains from schooling are greater than or equal to the expected costs.
- Thus under either set of assumptions, a separation theorem governs choices.
- Agents maximize expected wealth independently of their consumption decisions over time.

A Generalized Roy Model

- The decision rule is more complicated in the absence of full risk diversifiability and depends on the curvature of utility functions, the availability of markets to spread risk, and possibilities for storage.
- See Cunha and Heckman, 2006a, and Navarro, 2005, for a more extensive discussion.
- In these more realistic economic settings, the components of earnings and costs required to forecast the gain to schooling depend on higher moments than the mean.
- In this chapter we use a model with a simple market setting to motivate the identification analysis of a more general environment analyzed elsewhere (Carneiro, Hansen, and Heckman, 2003, and Cunha and Heckman, 2006b).

A Generalized Roy Model

- Suppose that we seek to determine $\mathcal{I}_{i,0}$.
- This is a difficult task.
- Typically we can only partially identify $\mathcal{I}_{i,0}$ and generate a list of candidate variables that belong in the information set.
- We can usually only estimate the distributions of the unobservables in $\mathcal{I}_{i,0}$ (from the standpoint of the econometrician) and not individual person-specific information sets.
- Before describing the analysis of Cunha, Heckman, and Navarro (2005), we consider how this question might be addressed in the linear-in-the-parameters Card model.

Identifying Information Sets in the Card Model

- We seek to decompose the "returns" coefficient or the gross gains from schooling in an earnings-schooling model into components that are known at the time schooling choices are made and components that are not known.
- For simplicity assume that, for person i, returns are the same at all levels of schooling.
- Write the log of discounted lifetime earnings of person i as

$$Y_i = \alpha + \rho_i S_i + U_i, \tag{34}$$

where ρ_i is the person-specific *ex post* return, S_i is years of schooling, and U_i is a mean zero unobservable.

Identifying Information Sets in the Card Model

- We seek to decompose ρ_i into two components $\rho_i = \eta_i + \nu_i$, where η_i is a component known to the agent when he/she makes schooling decisions and ν_i is revealed after the choice is made.
- Schooling choices are assumed to depend on what is known to the agent at the time decisions are made, $S_i = \lambda\left(\eta_i, Z_i, \tau_i\right)$, where the Z_i are other observed determinants of schooling and τ_i represents additional factors unobserved by the analyst but known to the agent.
- Both of these variables are in the agent's information set at the time schooling choices are made.
- We seek to determine what components of *ex post* lifetime earnings Y_i enter the schooling choice equation.

Identifying Information Sets in the Card Model

- If η_i is known to the agent and acted on, it enters the schooling choice equation.
- Otherwise it does not.
- Component ν_i and any measurement errors in $Y_{1,i}$ or $Y_{0,i}$ should not be determinants of schooling choices.
- Neither should future skill prices that are unknown at the time agents make their decisions.

Identifying Information Sets in the Card Model

- If agents do not use η_i in making their schooling choices, even if they know it, η_i would not enter the schooling choice equation.
- Determining the correlation between realized Y_i and schooling choices based on *ex ante* forecasts enables economists to identify components known to agents and acted on in making their schooling decisions.
- Even if we cannot identify ρ_i, η_i, or ν_i for each person, under conditions specified in this chapter we can identify their distributions.

Identifying Information Sets in the Card Model

- If we correctly specify the X and the Z that are known to the agent at the time schooling choices are made, local instrumental variable estimates of the MTE as described above (in the section on accounting for heterogeneity) identify *ex ante* gross gains.
- Any dependence between U_S and $Y_1 - Y_0$ arises from information known to the agent at the time schooling choices are made.

Identifying Information Sets in the Card Model

- If the conditioning set is misspecified by using information on X and Z that accumulates after schooling choices are made and that predicts realized earnings (but not *ex ante* earnings), the estimated MTE identifies an *ex post* return relative to that information set.
- Thus, it is important to specify the conditioning set correctly to obtain the appropriate *ex ante* return.
- How to pick the information set?

Identifying Information Sets in the Card Model

- Suppose that the model for schooling can be written in linear in parameters form, as in the Card model:

$$S_i = \lambda_0 + \lambda_1 \eta_i + \lambda_2 \nu_i + \lambda_3 Z_i + \tau_i, \tag{35}$$

  where τ_i has mean zero and is assumed to be independent of Z_i.

- The Z_i and the τ_i proxy costs and may be correlated with U_i and ρ_i in (34).
- In this framework, the goal of the analysis is to determine the η_i and ν_i components.
- By definition, $\lambda_2 = 0$ if ν_i is not known when agents make their schooling choices.

Identifying Information Sets in the Card Model

- As a simple example, consider the model of the section on mean growth rate of earnings.
- We drop "*i*" subscripts unless they clarify notation.
- We observe the cost of funds, r, and assume $r \perp\!\!\!\perp (\rho, \alpha)$.
- This assumes that the costs of schooling are independent of the "return" ρ and the payment to raw ability, α.
- We established identification of $\bar{\rho}$.
- (If there are observed regressors X determining the mean of $\bar{\rho}$, we identify $\bar{\rho}(X)$, the conditional mean of ρ).

Identifying Information Sets in the Card Model

- Suppose that agents do not know ρ at the time they make their schooling decisions but instead know $E(\rho) = \bar{\rho}$.
- If agents act on this expected return to schooling, decisions are given by

$$S = \frac{\bar{\rho} - r}{k}$$

and *ex post* earnings observed after schooling are

$$Y = \bar{\alpha} + \bar{\rho}S + \{(\alpha - \bar{\alpha}) + (\rho - \bar{\rho})S\}.$$

- In the notation introduced in the Card model, $\eta = \bar{\rho}$ and $\nu = \rho - \bar{\rho}$.

Identifying Information Sets in the Card Model

- In this case,
$$\text{Cov}\,(Y, S) = \bar{\rho}\,\text{Var}\,(S)$$
because $(\rho - \bar{\rho})$ is independent of S.

- Note further that $(\bar{\alpha}, \bar{\rho})$ can be identified by least squares because $S \perp\!\!\!\perp [(\alpha - \bar{\alpha}), (\rho - \bar{\rho})\,S]$.

Identifying Information Sets in the Card Model

- If, on the other hand, agents know ρ at the time they make their schooling decisions, *OLS* breaks down for identifying $\bar{\rho}$ because ρ is correlated with S.
- We can identify $\bar{\rho}$ and the distribution of ρ using the method of instrumental variables presented.
- Under our assumptions, r is a valid instrument for S.

Identifying Information Sets in the Card Model

- In this case

$$\text{Cov}(Y, S) = \bar{\rho}\,\text{Var}(S) + \text{Cov}(S, (\rho - \bar{\rho})\, S) .$$

- Since we observe S, we can identify $\bar{\rho}$ and can construct $(\rho - \bar{\rho})$ for each S, we can form both terms on the right hand side.
- Under the assumption that agents do not know ρ but forecast it by $\bar{\rho}$, ρ is independent of S so we can test for independence directly.
- In this case the second term on the right hand side is zero and does not contribute to the explanation of $\text{Cov}(\ln Y, S)$.

Identifying Information Sets in the Card Model

- Note further that a Durbin (1954) – Wu (1973) – Hausman (1978) test can be used to compare the *OLS* and *IV* estimates, which should be the same under the model that assumes that ρ is not known at the time schooling decisions are made and that agents base their choice of schooling on $E(\rho) = \bar{\rho}$.
- If the economist does not observe r, but instead observes determinants L satisfying the conditions discussed above, then we can still conduct the Durbin – Wu – Hausman test to discriminate between the two hypotheses, but we cannot form $\text{Cov}(\rho, S)$ directly.

Identifying Information Sets in the Card Model

- If we add selection bias to the Card model (so $E(\alpha \mid S)$ depends on S), we can identify $\bar{\rho}$ by *IV* as shown above, but *OLS* is no longer consistent even if, in making their schooling decisions, agents forecast ρ using $\bar{\rho}$.
- Selection bias can occur, for example, if fellowship aid is given on the basis of raw ability.
- Thus the Durbin-Wu-Hausman test is not helpful in assessing what is in the agent's information set.

Identifying Information Sets in the Card Model

- Even ignoring selection bias, if we misspecify the information set, in the case where r is not observed, the proposed testing approach based on the Durbin-Wu-Hausman test breaks down.
- Thus if we include in L variables that predict *ex post* gains $(\rho - \bar{\rho})$ and are correlated with S, we do not identify $\bar{\rho}$.
- The Durbin-Wu-Hausman test is not informative on the stated question.

Identifying Information Sets in the Card Model

- For example, if local labor market variables proxy the opportunity cost of school (the r), and also predict the evolution of *ex post* earnings $(\rho - \bar{\rho})$, they are invalid.
- The question of determining the appropriate information set is front and center and cannot in general be inferred using *IV* methods.

Identifying Information Sets in the Card Model

- The method developed by Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006a,b) exploits the covariance between S and the realized Y_t to determine which components of Y_t are known at the time schooling decisions are made.
- It explicitly models selection bias and allows for measurement error in earnings.
- It does not rely on linearity of the schooling relationship in terms of $\rho - r$.
- Their method recognizes the discrete nature of the schooling decision.
- It builds on the modern literature on constructing counterfactual schooling models.

Identifying Information Sets

- Cunha, Heckman, and Navarro (2005, 2006) henceforth CHN, exploit covariances between schooling and realized earnings that arise under different information structures to test which information structure characterizes the data.
- To see how the method works, simplify the model to two schooling levels.
- Heckman and Navarro (2006) analyze models with multiple schooling levels, but do not present empirical estimates of their model.

Identifying Information Sets

- Suppose, contrary to what is possible, that the analyst observes $Y_{0,i}$, $Y_{1,i}$, and C_i.
- Such information would come from an ideal data set in which we could observe two different lifetime earnings streams for the same person in high school and in college as well as the costs they pay for attending college.
- From such information, we could construct $Y_{1,i} - Y_{0,i} - C_i$.
- If we knew the information set $\mathcal{I}_{i,0}$ of the agent that governs schooling choices, we could also construct $E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right)$.

Identifying Information Sets

- Under a given model of expectations, we could form the residual

$$V_{\mathcal{I}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right),$$

  and from the *ex ante* college choice decision, we could determine whether S_i depends on $V_{\mathcal{I}_{i,0}}$.

- It should not if we have specified $\mathcal{I}_{i,0}$ correctly.

- In terms of the model of equations (34) and (35), if there are no direct costs of schooling, $E\left(Y_{1,i} - Y_{0,i} \mid \mathcal{I}_{i,0}\right) = \eta_i$, and $V_{\mathcal{I}_{i,0}} = \nu_i$.

Identifying Information Sets

- A test for correct specification of candidate information set $\widetilde{\mathcal{I}}_{i,0}$ is a test of whether S_i depends on $V_{\widetilde{\mathcal{I}}_{i,0}}$, where
$$V_{\widetilde{\mathcal{I}}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E\left(Y_{1,i} - Y_{0,i} - C_i \mid \widetilde{\mathcal{I}}_{i,0}\right).$$
- More precisely, the information set is valid if $S_i \perp\!\!\!\perp V_{\widetilde{\mathcal{I}}_{i,0}} \mid \widetilde{\mathcal{I}}_{i,0}$.
- In terms of the simple linear schooling model of equations (34) and (35), this condition says that ν_i should not enter the schooling choice equation ($\lambda_2 = 0$).
- A test of misspecification of $\widetilde{\mathcal{I}}_{i,0}$ is a test of whether the coefficient of $V_{\widetilde{\mathcal{I}}_{i,0}}$ is statistically significantly different from zero in the schooling choice equation.

Identifying Information Sets

- More generally, $\widetilde{\mathcal{I}}_{i,0}$ is the correct information set if $V_{\widetilde{\mathcal{I}}_{i,0}}$ does not help to predict schooling.
- One can search among candidate information sets $\widetilde{\mathcal{I}}_{i,0}$ to determine which ones satisfy the requirement that the generated $V_{\widetilde{\mathcal{I}}_{i,0}}$ does not predict S_i and what components of $Y_{1,i} - Y_{0,i} - C_i$ (and $Y_{1,i} - Y_{0,i}$) are predictable at the age schooling decisions are made for the specified information set.

Identifying Information Sets

- There may be several information sets that satisfy this property.
- For a properly specified $\widetilde{\mathcal{I}}_{i,0}$, $V_{\widetilde{\mathcal{I}}_{i,0}}$ should not cause (predict) schooling choices.
- The components of $V_{\widetilde{\mathcal{I}}_{i,0}}$ that are unpredictable are called intrinsic components of uncertainty, as defined in this chapter.

Identifying Information Sets

- It is difficult to determine the exact content of $\mathcal{I}_{i,0}$ known to each agent.
- If we could, we would perfectly predict S_i given our decision rule.
- More realistically, we might find variables that proxy $\mathcal{I}_{i,0}$ or their distribution.
- Thus, in the example of equations (34) and (35) we would seek to determine the distribution of ν_i and the allocation of the variance of ρ_i to η_i and ν_i rather than trying to estimate ρ_i, η_i, or ν_i for each person.

Identifying Information Sets

- This strategy is pursued in Cunha, Heckman, and Navarro (2005, 2006) for a two-choice model of schooling, and generalized by Cunha and Heckman (2006a).
- To implement such a test requires overcoming the problem of missing counterfactual earnings equations.
- We now describe a method for doing so developed in Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005, 2006, 2007).

An Approach Based on Factor Structures

- The essence of the idea underlying the method of testing for what is in an agent's information set at the time schooling decisions are made is communicated by adapting the model presented in the section on Estimating Distributions of Returns to Schooling.
- Suppose that at the time they make their schooling decisions, agents do not know θ or the future $\varepsilon_{1t}, \varepsilon_{0t}, t = 1, 2$.
- Recall that period 1 is the initial period in that example.
- θ is realized after schooling choices are made.

An Approach Based on Factor Structures

- The agents know X, Z and ε_C.
- Thus $\mathcal{I}_{i,0} = \{X_i, Z_i, \varepsilon_C\}$.
- Suppose that θ is independent of X, Z, ε_C and $E(\theta \mid X, Z, \varepsilon_C) = 0$.
- Under rational expectations U_S is independent of all future earnings disturbances so that $\text{Cov}(U_S, U_{11}) = 0$, $\text{Cov}(U_S, U_{12}) = 0$, $\text{Cov}(U_S, U_{01}) = 0$, $\text{Cov}(U_S, U_{02}) = 0$.
- However realized earnings are correlated with each other through the realized θ.

An Approach Based on Factor Structures

- Under the assumptions discussed earlier, we can test for the zero covariances.
- If nonzero covariances are found, then θ is a component of heterogeneity.
- Otherwise θ contributes to *ex ante* uncertainty.
- By design, this example is overly simplistic.
- It is more likely that there are multiple sources of unobserved heterogeneity (θ is a vector) and that they may only partially know the X that are realized after schooling decisions are made (e.g. macro shocks or new trends in skill prices).
- A more general procedure is required to account for those possibilities which we now describe.

An Approach Based on Factor Structures

- Consider the following linear in parameters model for a full T periods.
- This analysis generalizes the example just presented.
- Write earnings in each counterfactual state as

$$\begin{aligned} Y_{0,i,t} &= X_{i,t}\beta_{0,t} + U_{0,i,t}, \\ Y_{1,i,t} &= X_{i,t}\beta_{1,t} + U_{1,i,t}. \qquad t = 0, \ldots, T. \end{aligned}$$

- We let costs of college be defined as

$$C_i = Z_i\gamma + U_{i,C}.$$

- Assume that the life cycle of the agent ends after period T.
- Linearity of outcomes in terms of parameters is convenient but not essential to the method of CHN.

An Approach Based on Factor Structures

- Suppose that there exists a vector of factors $\theta_i = (\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,L})$ such that $\theta_{i,k}$ and $\theta_{i,j}$ are mutually independent random variables for $k, j = 1, \ldots, L, k \neq j$.
- They represent the error term in earnings at age t for agent i in the following manner:

$$\begin{aligned} U_{0,i,t} &= \theta_i \alpha_{0,t} + \varepsilon_{0,i,t}, \\ U_{1,i,t} &= \theta_i \alpha_{1,t} + \varepsilon_{1,i,t}, \end{aligned}$$

where $\alpha_{0,t}$ and $\alpha_{1,t}$ are vectors and θ_i is a vector distributed independently across persons.

An Approach Based on Factor Structures

- The $\varepsilon_{0,i,t}$ and $\varepsilon_{1,i,t}$ are mutually independent of each other and independent of the θ_i.
- We can also decompose the cost function C_i in a similar fashion:

$$C_i = Z_i\gamma + \theta_i\alpha_C + \varepsilon_{i,C}.$$

An Approach Based on Factor Structures

- All of the statistical dependence across potential outcomes and costs is generated by θ, X, and Z.
- Thus, if we could match on θ_i (as well as X and Z), we could use matching to infer the distribution of counterfactuals and capture all of the dependence across the counterfactual states through the θ_i.
- Thus we could use θ as the Q used above if we could observe it.
- However, in general, CHN allow for the possibility that not all of the required elements of θ_i are observed.

An Approach Based on Factor Structures

- The parameters α_C and $\alpha_{s,t}$ for $s = 0, 1$, and $t = 0, \ldots, T$ are the factor loadings.
- $\varepsilon_{i,C}$ is independent of the θ_i and the other ε components.
- In this notation, the choice equation can be written as:

$$\begin{aligned} S_i^* &= E\left(\sum_{t=0}^{T} \frac{(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}) - (X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t})}{(1+r)^t} \right. \\ &\qquad \left. - (Z_i\gamma + \theta_i\alpha_C + \varepsilon_{iC}) \mid \mathcal{I}_{i,0} \right) \\ S_i &= 1 \text{ if } S^* \geq 0;\ S_i = 0 \text{ otherwise.} \end{aligned} \tag{36}$$

- The sum inside the parentheses is the discounted earnings of agent i in college minus the discounted earnings of the agent in high school.
- The second term is the cost of college.

An Approach Based on Factor Structures

- Constructing (36) entails making a counterfactual comparison.
- Even if the earnings of one schooling level are observed over the lifetime using panel data, the earnings in the counterfactual state are not.
- After the schooling choice is made, some components of the $X_{i,t}$, the θ_i, and the $\varepsilon_{i,t}$ may be revealed (e.g. unemployment rates, macro shocks) to both the observing economist and the agent, although different components may be revealed to each and at different times.

An Approach Based on Factor Structures

- For this reason, application of *IV* even in the linear schooling model is problematic.
- If the wrong information set is used, the *IV* method will not identify the true *ex ante* returns.

An Approach Based on Factor Structures

- Examining alternative information sets, one can determine which ones produce models for outcomes that fit the data best in terms of producing a model that predicts date $t = 0$ schooling choices and at the same time passes the CHN test for misspecification of predicted earnings and costs.
- Some components of the error terms may be known or not known at the date schooling choices are made.
- The unforecastable components are intrinsic uncertainty as CHN define it.
- The forecastable information is called heterogeneity.

An Approach Based on Factor Structures

- To formally characterize the CHN empirical procedure, it is useful to introduce some additional notation.
- Let $\odot$ denote the Hadamard product ($a \odot b = (a_1 b_1, \ldots, a_L b_L)$) for vectors a and b of length L.
- This is a componentwise multiplication of vectors to produce a vector.
- Let Δ_{X_t}, $t = 0, \ldots, T$, Δ_Z, Δ_θ, Δ_{ε_t}, Δ_{ε_C}, denote coefficient vectors associated with the X_t, $t = 0, \ldots, T$, the Z, the θ, the $\varepsilon_{1,t} - \varepsilon_{0,t}$, and the ε_C, respectively.
- These coefficients will be estimated to be nonzero in a schooling choice equation if a proposed information set is not the actual information set used by agents.

An Approach Based on Factor Structures

- For a proposed information set $\widetilde{\mathcal{I}}_{i,0}$ which may or may not be the true information set on which agents act, CHN define the proposed choice index $\tilde{S}_i^*$ in the following way:

$$
\begin{aligned}
\tilde{S}_i^* = & \sum_{t=0}^{T} \frac{E\left(X_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} (\beta_{1,t} - \beta_{0,t}) \\
& + \sum_{t=0}^{T} \frac{\left[X_{i,t} - E\left(X_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t} (\beta_{1,t} - \beta_{0,t}) \odot \Delta_{X_t} \\
& + E(\theta_i \mid \widetilde{\mathcal{I}}_{i,0}) \left[\sum_{t=0}^{T} \frac{(\alpha_{1,t} - \alpha_{0,t})}{(1+r)^t} - \alpha_C\right] \\
& + \left[\theta_i - E\left(\theta_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \left\{ \left[\sum_{t=0}^{T} \frac{(\alpha_{1,t} - \alpha_{0,t})}{(1+r)^t} - \alpha_C\right] \odot \Delta_\theta \right\} \\
& + \sum_{t=0}^{T} \frac{E\left(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} \\
& + \sum_{t=0}^{T} \frac{\left[(\varepsilon_{1,i,t} - \varepsilon_{0,i,t}) - E\left(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t} \Delta_{\varepsilon_t} \\
& - E\left(Z_i \mid \widetilde{\mathcal{I}}_{i,0}\right) \gamma - \left[Z_i - E\left(Z_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \gamma \odot \Delta_Z \\
& - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right) - \left[\varepsilon_{iC} - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \Delta_{\varepsilon_C}.
\end{aligned}
\tag{37}
$$

An Approach Based on Factor Structures

- To conduct their test, CHN fit a schooling choice model based on the proposed model (37).
- They estimate the parameters of the model including the Δ parameters.

An Approach Based on Factor Structures

- This decomposition for $\tilde{S}_i^*$ assumes that agents know the β, the γ, and the α.
- If that is not correct, the presence of additional unforecastable components due to unknown coefficients affects the interpretation of the estimates.
- A test of no misspecification of information set $\widetilde{\mathcal{I}}_{i,0}$ is a joint test of the hypothesis that the Δ are all zero.
- That is, when $\widetilde{\mathcal{I}}_{i,0} = \mathcal{I}_{i,0}$ then the proposed choice index $\widetilde{S}_i^* = S_i^*$.

An Approach Based on Factor Structures

- In a correctly specified model, the components associated with zero Δ_j are the unforecastable elements or the elements which, even if known to the agent, are not acted on in making schooling choices.
- To illustrate the application of the method of CHN, we elaborate on an example discussed previously, and assume for simplicity that the $X_{i,t}$, the Z_i, the $\varepsilon_{i,C}$, the $\beta_{1,t}, \beta_{0,t}$, the $\alpha_{1,t}, \alpha_{0,t}$, and α_C are known to the agent, and the $\varepsilon_{j,i,t}$ are unknown and are set at their mean zero values.

An Approach Based on Factor Structures

- We can infer which components of the θ_i are known and acted on in making schooling decisions if we postulate that some components of θ_i are known perfectly at date $t = 0$ while others are not known at all, and their forecast values have mean zero given $\mathcal{I}_{i,0}$.

An Approach Based on Factor Structures

- If there is an element of the vector θ_i, say $\theta_{i,2}$ (factor 2), that has nonzero loadings (coefficients) in the schooling choice equation and a nonzero loading on one or more potential future earnings, then one can say that at the time the schooling choice is made, the agent knows the unobservable captured by factor 2 that affects future earnings.
- If $\theta_{i,2}$ does not enter the choice equation but explains future earnings, then $\theta_{i,2}$ is unknown (not predictable by the agent) at the age schooling decisions are made.

An Approach Based on Factor Structures

- An alternative interpretation is that the second component of $\left[\sum_{t=0}^{T} \frac{(\alpha_{1,t}-\alpha_{0,t})}{(1+r)^t} - \alpha_C\right]$ is zero, i.e., that even if the component is known, it is not acted on.
- CHN can only test for what the agent knows and acts on.

An Approach Based on Factor Structures

- One plausible case is that for their model $\varepsilon_{i,C}$ is known (since schooling costs are incurred up front), but the future $\varepsilon_{1,i,t}$ and $\varepsilon_{0,i,t}$ are not, have mean zero, and are insurable.
- If there are components of the $\varepsilon_{j,i,t}$ that are predictable at age $t = 0$, they will induce additional dependence between S_i and future earnings that will pick up additional factors beyond those initially specified.
- The CHN procedure can be generalized to consider all components of (37).
- With it, the analyst can test the predictive power of each subset of the overall possible information set at the date the schooling decision is being made.

An Approach Based on Factor Structures

- In the context of the factor structure representation for earnings and costs, the contrast between the CHN approach to identifying components of intrinsic uncertainty and the approach followed in the literature is as follows.
- The traditional approach as exemplified by Keane and Wolpin (1997) assumes that the θ_i are known to the agent while the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are not.
- The CHN approach allows the analyst to determine which components of θ_i and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known and acted on at the time schooling decisions are made.

An Approach Based on Factor Structures

- Statistical decompositions do not tell us which components of (32) are known at the time agents make their schooling decisions.
- A model of expectations and schooling is needed.
- If some of the components of $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known to the agent at the date schooling decisions are made and enter (37), then additional dependence between S_i and future $Y_{1,i} - Y_{0,i}$ due to the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$, beyond that due to θ_i, would be estimated.

An Approach Based on Factor Structures

- It is helpful to contrast the dependence between S_i and future $Y_{0,i,t}$, $Y_{1,i,t}$ arising from θ_i and the dependence between S_i and the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$.
- Some of the θ_i in the *ex post* earnings equation may not appear in the choice equation.
- Under other information sets, some additional dependence between S_i and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ may arise.

An Approach Based on Factor Structures

- The contrast between the sources generating realized earnings outcomes and the sources generating dependence between S_i and realized earnings is the essential idea in the analysis of CHN.
- The method can be generalized to deal with nonlinear preferences and imperfect market environments.
- A central issue, discussed next, is how far one can go in identifying income information processes without specifying preferences, insurance, and market environments.

More general preferences and market settings

- To focus on the main ideas in the literature, we have used the simple market structures of complete contingent claims markets.
- What can be identified in more general environments?
- In the absence of perfect certainty or perfect risk sharing, preferences and market environments also determine schooling choices.
- The separation theorem allowing consumption and schooling decisions to be analyzed in isolation of each other that we have used thus far breaks down.

More general preferences and market settings

- If we postulate information processes *a priori*, and assume that preferences are known up to some unknown parameters as in Flavin (1981), Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2004) we can identify departures from specified market structures.
- Cunha and Heckman (2006b) postulate an Aiyagari (1994) – Laitner (1992) economy with one asset and parametric preferences to identify the information processes in the agent's information set.
- They take a parametric position on preferences and a nonparametric position on the economic environment and the information set.

More general preferences and market settings

- An open question, not yet fully resolved in the literature, is how far one can go in nonparametrically jointly identifying preferences, market structures and information sets.
- See Cunha, Heckman, and Navarro, 2005.
- Navarro (2005) adds consumption data to the schooling choice and earnings data to secure identification of risk preference parameters (within a parametric family) and information sets, and to test among alternative models for market environments.

More general preferences and market settings

- Alternative assumptions about what analysts know produce different interpretations of the same evidence.
- The lack of full insurance interpretation given to the empirical results by Flavin (1981) and Blundell, Pistaferri, and Preston (2004) may be a consequence of their misspecification of the agent's information set generating process.
- We now present some evidence on *ex ante* vs. *ex post* returns presented by Cunha and Heckman (2006b), henceforth CH.

Evidence on Uncertainty and Heterogeneity of Returns

- Few data sets contain the full life cycle of earnings along with the test scores and schooling choices needed to directly estimate the CHN model and extract components of uncertainty.
- It is necessary to pool data sets.
- CHN (2005) combine NLSY and PSID data sets.
- We summarize the analysis of CH in this subsection.
- See their paper for their exclusions and identification conditions.

Evidence on Uncertainty and Heterogeneity of Returns

- Following the preceding theoretical analysis, they consider only two schooling choices: high school and college graduation.
- For simplicity and familiarity, we focus on their results that are based on assuming that complete contingent claims markets characterize the data.
- We consider evidence from other market settings below.
- Because they assume that all shocks are idiosyncratic and that complete markets operate, schooling choices are made on the basis of expected present value income maximization.

Evidence on Uncertainty and Heterogeneity of Returns

- Carneiro, Hansen, and Heckman (2003) assume the absence of any credit markets or insurance.
- Navarro (2005) checks whether the CHN and CH findings about components of uncertainty are robust to different assumptions about the availability of credit markets and insurance markets.
- He estimates an Aiyagari-Laitner economy with a single asset and borrowing constraints and discusses risk aversion and the relative importance of uncertainty.
- We summarize the evidence from alternative assumptions about market structures below.
- We now summarize the model of CH (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- Suppose that the error term for $Y_{s,t}$ is generated by a two factor model,

$$Y_{s,t} = X\beta_{s,t} + \theta_1\alpha_{s,t,1} + \theta_2\alpha_{s,t,2} + \varepsilon_{s,t}. \tag{38}$$

- We omit the "i" subscripts to eliminate notational burden.
- Cunha and Heckman (2006b) report that two factors are all that is required to fit the data.

Evidence on Uncertainty and Heterogeneity of Returns

- They use a test score system of K ability tests:

$$T_{jT} = X_T \omega_{jT} + \theta_1 \alpha_{jT} + \varepsilon_{jT}, \quad j = 1, \ldots, K. \tag{39}$$

- Thus factor 1 is identified as an ability component.
- The cost function C is specified by:

$$C = Z\gamma + \theta_1 \alpha_{C,1} + \theta_2 \alpha_{C,2} + \varepsilon_C. \tag{40}$$

Evidence on Uncertainty and Heterogeneity of Returns

- They assume that agents know the model coefficients and X, Z, ε_C and some, but not necessarily all, components of θ.
- Let the components known to the agent be $\bar{\theta}$.
- The decision rule for attending college is based on:

$$\begin{aligned} S^* &= E\left(Y_{1,0} + \frac{Y_{1,1}}{1+r} - Y_{0,0} - \frac{Y_{0,1}}{1+r} \mid X, \bar{\theta}\right) \\ &\quad - E\left(C \mid Z, X, \bar{\theta}, \varepsilon_C\right) \\ S &= \mathbf{1}\left(S^* \geq 0\right). \end{aligned} \tag{41}$$

- Cunha and Heckman (2006b) report evidence that the estimated factors are highly nonnormal.

Evidence on Uncertainty and Heterogeneity of Returns

- Table 15 presents the conditional distribution of *ex ante* potential college earnings given *ex ante* potential high school earnings, decile by decile, as reported by Cunha and Heckman (2006b).
- The table displays positive dependence between the relative positions of individuals in the two distributions, but it is not especially strong.
- In all high school deciles save the highest, almost 90% of persons are in a different college decile.
- Observe that this comparison is not made in terms of positions in the overall distribution of earnings.

Table 15: Ex-Ante Conditional Distributions for the NLSY79
(College Earnings Conditional on High School Earnings)
$\Pr(d_i < Y_c < d_i + 1 | d_j < Y_h < d_j + 1)$
where d_i is the i^{th} decile of the College Lifetime Ex-Ante Earnings Distribution
and d_j is the j^{th} decile of the High School Ex-Ante Lifetime Earnings Distribution
Individual fixes known q at their means, so Information Set $= \{q_1 = 0, q_2 = 0\}$
Correlation$(Y_C, Y_H) = 0.4083$

High School	College									
	1	2	3	4	5	6	7	8	9	10
1	0.1833	0.1631	0.1330	0.1066	0.0928	0.0758	0.0675	0.0630	0.0615	0.0535
2	0.1217	0.1525	0.1262	0.1139	0.1044	0.0979	0.0857	0.0796	0.0683	0.0498
3	0.1102	0.1263	0.1224	0.1198	0.1124	0.0970	0.0931	0.0907	0.0775	0.0506
4	0.0796	0.1083	0.1142	0.1168	0.1045	0.1034	0.1121	0.1006	0.0953	0.0652
5	0.0701	0.0993	0.1003	0.1027	0.1104	0.1165	0.1086	0.1112	0.1043	0.0768
6	0.0573	0.0932	0.1079	0.1023	0.1110	0.1166	0.1130	0.1102	0.1059	0.0825
7	0.0495	0.0810	0.0950	0.1021	0.1101	0.1162	0.1202	0.1174	0.1134	0.0950
8	0.0511	0.0754	0.0770	0.1006	0.1006	0.1053	0.1244	0.1212	0.1297	0.1147
9	0.0411	0.0651	0.0841	0.0914	0.1039	0.1117	0.1162	0.1216	0.1442	0.1206
10	0.0590	0.0599	0.0622	0.0645	0.0697	0.0782	0.0770	0.1028	0.1181	0.3087

Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- CH can determine where individuals are located in the distribution of population potential high school earnings and the distribution of potential college earnings although in the data we only observe them in either one or the other state.
- Their evidence shows that the assumption of preservation of ranks across counterfactual distributions that is maintained in much of the recent literature (e.g. Juhn, Murphy, and Pierce, 1993) is far too strong.
- They also report evidence of less than perfect sorting on *ex post* earnings.

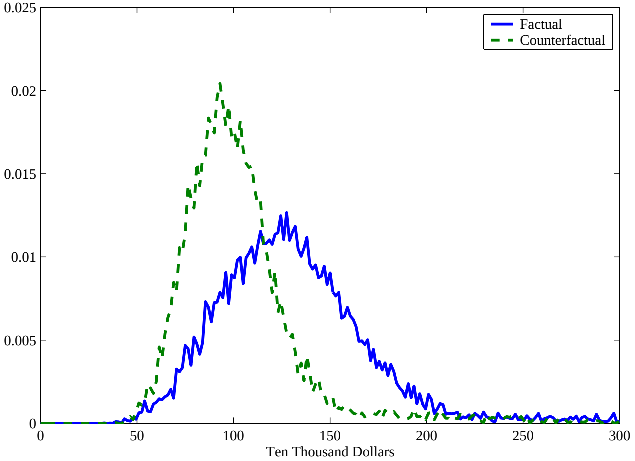
Evidence on Uncertainty and Heterogeneity of Returns

- Figure 16 presents the marginal density of predicted (actual) present value of earnings for college students and the counterfactual density of the present value of their earnings if they were high school students.
- When we compare the densities of present value of earnings in the college sector for persons who choose college against the densities of counterfactual high school earnings for college graduates, the density of the present value of earnings for college graduates in college is to the right of the counterfactual density of the present value of high school earnings for college graduates.

Figure 16

Densities of present value of earnings for College Graduates

Factual and Counterfactual NLSY/1979 Sample



Present Value of Lifetime Earnings from age 18 to 65 for high school graduates using a discount rate of 3%. Let Y_0 denote present value of earnings in high school sector. Let Y_1 denote present value of earnings in college sector. In this graph we plot the counterfactual density function $f(y_1 \mid S=0)$ (the dashed line), against the factual density function $f(y_1 \mid S=0)$. Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- A parallel analysis for high school graduates reveals that the density of college earnings for high school graduates is to the right of the distribution of their high school earnings so that many high school graduates would earn more by going to college.

Evidence on Uncertainty and Heterogeneity of Returns

- Table 16 from CHN reports the fitted and counterfactual present value of earnings for agents who choose high school.
- The typical high school student would earn \$968.5 thousand over the life cycle.
- He would earn \$1,125.8 thousand if he had chosen to be a college graduate.
- This implies a return of 20.5% to a college education over the whole life cycle (i.e., a monetary gain of \$157.3 thousand).

Table 16: Average Present Value of *Ex Post* Earnings[1] for High School Graduates
Fitted and Counterfactual[2]
White males from NLSY79

	High School (Fitted)	College (Counterfactual)	Returns[3]
Average	968.5100	1125.7870	0.2055
Standard Error	7.9137	9.4583	0.0113

[1]Thousands of dollars. Discounted using a 3% interest rate

[2]The counterfactual is constructed using the estimated college outcome equation applied to the population of persons selecting high school.

[3]As a fraction of the base state, i.e., $\frac{(\mathrm{PV}_{\mathrm{earnings}}(\mathrm{Col})-\mathrm{PV}_{\mathrm{earnings}}(\mathrm{HS}))}{\mathrm{PV}_{\mathrm{earnings}}(\mathrm{HS})}$.

Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- In Table 17, CHN note that the typical college graduate earns \$1,390.3 thousand if he goes to college (above the counterfactual earnings of what a typical high school student would earn in college), and would make only \$1,033.7 thousand over his lifetime if he chose to be a high school graduate instead.
- The returns to college education for the typical college graduate (which in the literature on program evaluation is referred to as the effect of Treatment on the Treated) is almost double that of the return for a high school graduate.
- In monetary terms a college graduate has a gain of going to college almost \$175 thousand higher over his lifetime than does the typical high school graduate.

Table 17: Average Present Value Of *Ex Post* Earnings[1] for College Graduates
Fitted and Counterfactual[2]
White males from NLSY79

	High School (Counterfactual)	College (fitted)	Returns[3]
Average	1033.721	1390.321	0.374
Standard Error	14.665	30.218	0.280

[1]Thousands of dollars. Discounted using a 3% interest rate

[2] The counterfactual is constructed using the estimated high school outcome equation applied to the population of persons selecting college.

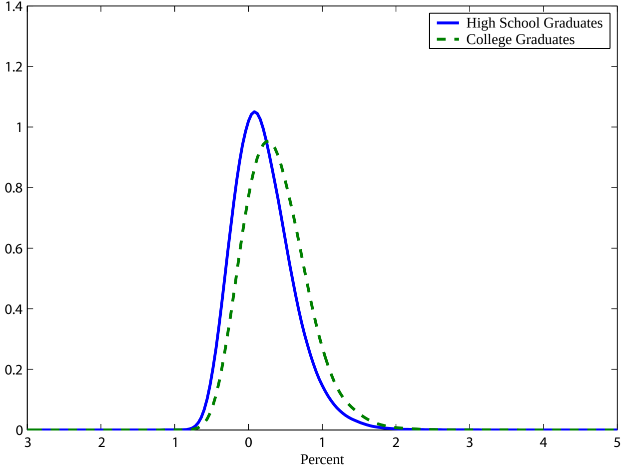
[3]As a fraction of the base state, i.e., $\frac{(\text{PV}_{\text{earnings}}(\text{Col})-\text{PV}_{\text{earnings}}(\text{HS}))}{\text{PV}_{\text{earnings}}(\text{HS})}$.

Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- Figure 17 plots the density of *ex post* gross returns to education excluding direct costs and psychic costs for agents who are high school graduates (the solid curve), and the density of returns to education for agents who are college graduates (the dashed curve).
- In reporting our estimated returns, CH follow conventions in the literature and actually present growth rates in terms of present values, and not true rates of return (ignoring option values).
- Thus these figures report the growth rates in present values $\left(\frac{PV(1)-PV(0)}{PV(0)}\right)$ where "1" and "0" refer to college and high school and all present values are discounted to a common benchmark level.

Figure 17
Densities of Returns to College
NLSY/1979 Sample



Let Y_0 denote present value of earnings in high school sector. Let Y_1 denote present value of earnings in college sector. Let $R = (Y_1 - Y_0)/Y_0$ denote the gross rate of return to college. In this graph we plot the density function of the returns to college conditional on being a high school graduate, $f(r \mid S=0)$ (the solid line), against the density function of returns to college conditional on being a college graduate, $f(r \mid S=1)$. We use kernel density estimation to smooth these functions. Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- Tuition and psychic costs are ignored.
- College graduates have returns distributed somewhat to the right of high school graduates, so the difference is not only a difference for the mean individual but is actually present over the entire distribution.
- Agents who choose a college education are the ones who tend to gain more from it.

Evidence on Uncertainty and Heterogeneity of Returns

- With their methodology, CHN can also determine returns to the marginal student.
- This could also be estimated by the *MTE* method discussed earlier.
- Under rational expectations, mean *ex post* and *ex ante* returns are the same although the distributions may differ.
- Table 18 reveals that the average individual who is just indifferent between a college education and a high school diploma earns $976.04 thousand as a high school graduate or $1,208.26 thousand as a college graduate.

Table 18: Average Present Value Of *Ex Post* Earnings[1] for Individuals at the Margin
Fitted and Counterfactual[2]
White males from NLSY79

	High School	College	Returns[3]
Average	976.04	1208.26	0.2828
Std. Err.	21.503	33.613	0.0457

[1]Thousands of dollars. Discounted using a 3% interest rate

[2] The counterfactual is defined as the result of taking a person at random from the population regardless of his schooling choice.

[3]As a fraction of the base state, i.e., $\frac{(\mathrm{PV}_{\mathrm{earnings}}(\mathrm{Col})-\mathrm{PV}_{\mathrm{earnings}}(\mathrm{HS}))}{\mathrm{PV}_{\mathrm{earnings}}(\mathrm{HS})}$.

Source: Cunha and Heckman (2006a).

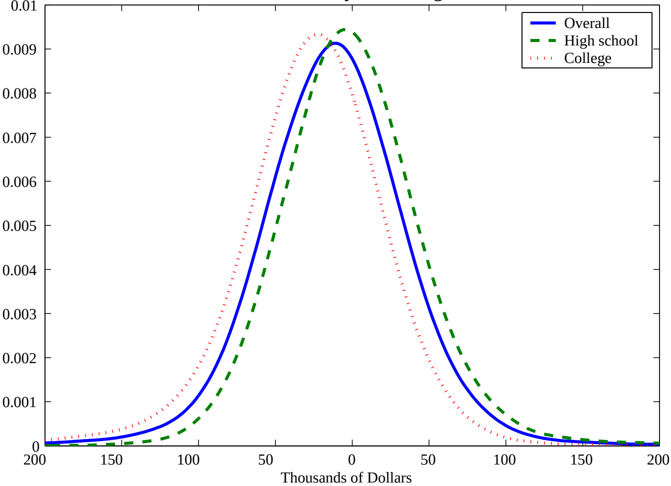
Evidence on Uncertainty and Heterogeneity of Returns

- This implies a return of 28%.
- The returns to people at the margin are above those of the typical high school graduate, but below those for the typical college graduate.
- Since persons at the margin are more likely to be affected by a policy that encourages college attendance, their returns are the ones that should be used in order to compute the marginal benefit of policies that induce people into schooling.

Evidence on Uncertainty and Heterogeneity of Returns

- A major question that emerges from the analyses of CHN and CH is, why, if high school graduates have such positive returns to attending college, don't more attend? People do not pick schooling levels based only on monetary returns.
- Recall that their choice criterion (equation 41) also includes both pecuniary and non-pecuniary costs of attending college.
- Figure 18 shows the estimated density of the monetary value of this cost both overall and by schooling level.
- Fewer high school graduates than college graduates perceive a benefit (negative cost) from attending college.

Figure 18
Densities of monetary value of psychic cost
both overall and by schooling level



Let C denote the monetary value of psychic costs. Let f(c) denote the density function of psychic costs in monetary terms. The dashed line shows the density of psychic costs for high school graduates, that is, f(c|S=0). The dotted line shows the density of psychic costs for college graduates, that is, f(c|S=1) The solid line is the unconditional density of the monetary value of psychic costs, f(c).
Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- Table 19 explores this point in more detail by presenting the mean total cost of attending college (first row) and the mean cost that is due to ability (i.e., factor 1), given in the second row.
- The mean cost of attending college is negative for the average college graduate and positive for the average high school graduate.
- Costs are substantially smaller for college graduates.
- Average college graduates have higher ability.
- The average contribution of ability to costs is positive for high school graduates (a true cost).
- It is negative for college graduates, so it is perceived as a benefit.

Table 19: Monetary Costs of Schooling Levels

	High School	College	Overall
Mean Monetary Value of Total Cost of Attending College	0.4393	-26.8651	-11.9223
Mean Monetary Value of Ability Cost of Attending College	12.7152	-14.8924	0.0000

Values in thousands of dollars (2000).

Let C denote the psychic costs in monetary terms. Then C is given by

$$C = Z\gamma + \theta_1 \alpha_{C_1} + \theta_2 \alpha_{C_2} + \varepsilon_C$$

The contribution of ability to the costs of attending college is monetary value is $\theta_1 \alpha_C$. Recall that, on average, the ability is different between those attending college and those attending high school.

Source: Cunha and Heckman (2006a)

Evidence on Uncertainty and Heterogeneity of Returns

- This is one answer to the stated puzzle.
- People do not only (or even mainly) make their schooling decisions by looking at their monetary returns in terms of earnings.
- Psychic costs play a very important role.
- More able people have lower psychic costs of attending college.
- The high estimated psychic cost is one reason why the rates of return that ignore psychic costs (and tuition) are so high.

Evidence on Uncertainty and Heterogeneity of Returns

- This high psychic cost is one explanation why the college attendance rate is so low when the monetary returns are so high.
- One convention in the classical human capital literature—that tuition and psychic costs are negligible—is at odds with this evidence.
- The evidence against strict income maximization is overwhelming.

Evidence on Uncertainty and Heterogeneity of Returns

- However, explanations based on psychic costs are intrinsically unsatisfactory.
- One can rationalize any economic choice data by an appeal to psychic costs.
- Heckman, Stixrud, and Urzua (2006) show the important role played by noncognitive skills as well as cognitive skills in explaining schooling (and other decisions).
- They show how, in principle, conventional risk aversion, time preference and leisure preference parameters can be related to psychometric measures of cognitive and noncognitive skills.
- Establishing this link will provide a better foundation for understanding what "psychic costs" actually represent.

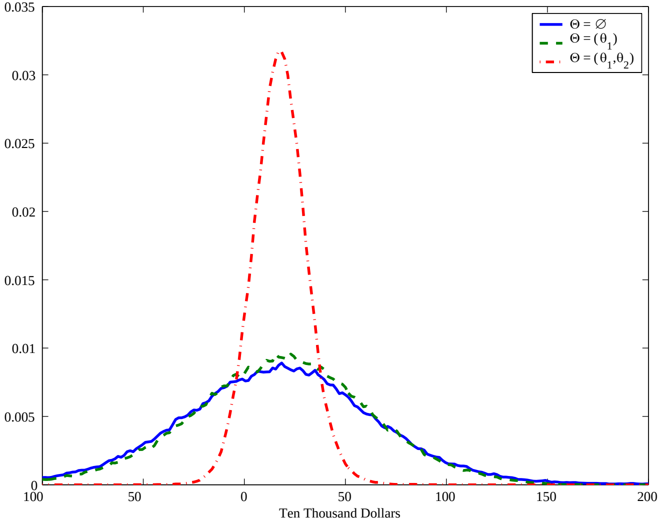
Evidence on Uncertainty and Heterogeneity of Returns

- "Psychic costs" may also be a stand in for credit constraints and risk aversion.
- However, the evidence on psychic costs in schooling choice equations is more sturdy than this discussion might suggest.
- Carneiro, Hansen, and Heckman (2003) obtain similar conclusions on the importance of psychic costs from a model where people are not allowed to borrow or lend and there is risk aversion.
- In Cunha, Heckman, and Navarro (2005), on the other hand, there are no constraints on borrowing or lending, and they also show sizeable components of psychic costs.

Evidence on Uncertainty and Heterogeneity of Returns

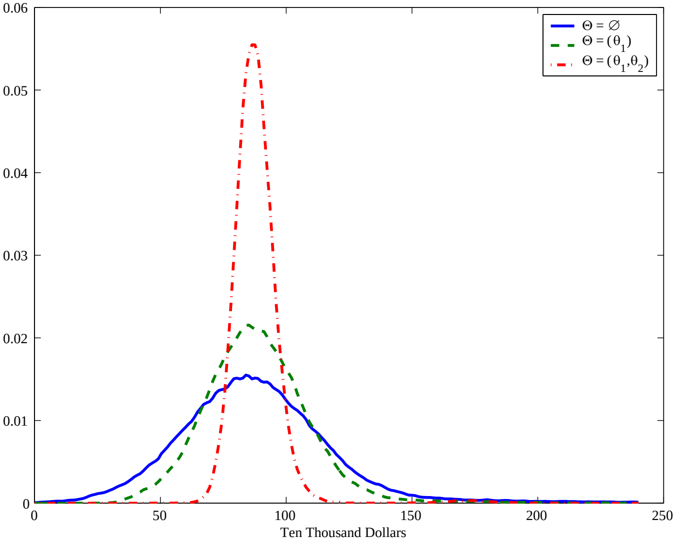
- Figures 19 through 21, from Cunha and Heckman (2006b) separate the effect of heterogeneity from uncertainty in earnings.
- The figures plot the distribution of *ex ante* and *ex post* outcomes under different information sets.
- The information set of the agent is $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$, Θ contains some or all of the factors.
- In this paper and in the analyses of Cunha, Heckman, and Navarro (2005, 2006, 2007) and Carneiro, Hansen, and Heckman (2003), the various information sets consist of different components of θ.

Figure 19
Densities of present value of returns - NLSY/1979
under different information sets for the agent calculated
for the entire population irregardless of schooling choice



Let Θ denote the information set of the agent. Let Y_0 denote the present value of returns (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of Y_0 conditional on information set Θ. The solid line plots the density of Y_0 when $\Theta=\phi$. The dashed line plots the density of Y_0 when $\Theta=\{\theta_1\}$.The dotted and dashed line plots the density of of Y_0 when $\Theta=\{\theta_1, \theta_2\}$. The X variables are in the information

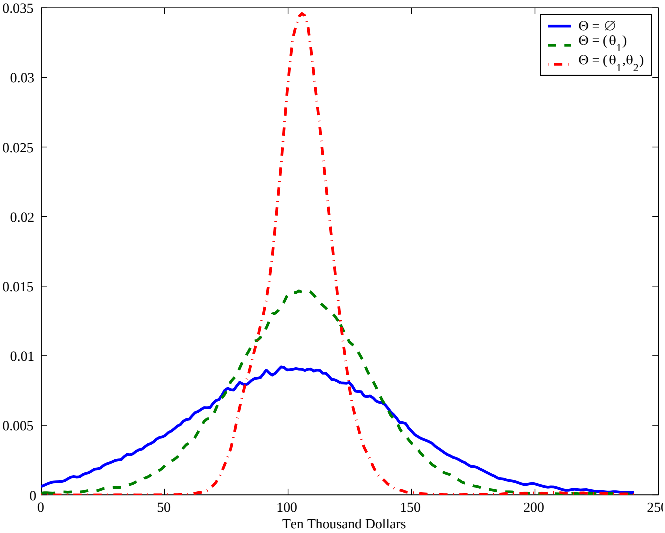
Figure 20
Densities of present value of high school earnings - NLSY/1979
under different information sets for the agent calculated
for the entire population irregardless of schooling choice



Let Θ denote the information set of the agent. Let Y_0 denote the present value of earnings in the high school sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of Y_0 conditional on information set Θ. The solid line plots

Figure 21

Densities of present value of college earnings - NLSY/1979 under different information sets for the agent calculated for the entire population irregardless of schooling choice



Let Θ denote the information set of the agent. Let Y_0 denote the present value of earnings in the college sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of Y_0 conditional on information set Θ. The solid line plots the density of Y_0 when $\Theta=\phi$. The dashed line plots the density of Y_0 when $\Theta=\{\theta_1\}$.

Evidence on Uncertainty and Heterogeneity of Returns

- First consider Figure 19.
- It presents results for a variety of information sets.
- First assume that agents do not know their factors; consequently, $\Theta = \varnothing$.
- This is the *ex post* or realized distribution of the variation that is observed in the data.
- If the agents learn about factor 1, so that, $\Theta = \{\theta_1\}$, the reduction in the forecast variance is very small.

Evidence on Uncertainty and Heterogeneity of Returns

- Factor 1, which is associated with cognitive ability, is important for forecasting educational choices, but does not do a very good job in forecasting earnings.
- If the agent is given knowledge of factor 2, but not factor 1, so that $\Theta = \{\theta_2\}$, then the agent is able to substantially reduce the forecast variance of earnings in high school.
- Factor 2 does not greatly affect college choices, but it greatly informs the agent about his future earnings.
- When the agent is given knowledge of both factors 1 and 2, that is, $\Theta = \{\theta_1, \theta_2\}$, he can forecast earnings marginally better.

Evidence on Uncertainty and Heterogeneity of Returns

- Figure 20 reveals much the same story about the college earnings distribution.
- These results suggest that selection into college is not based primarily on expected economic returns to education.
- Cost factors play an important role.

Evidence on Uncertainty and Heterogeneity of Returns

- Table 20 presents agent forecast variances of the present value of future earnings and their return.
- CH establish that agents know (θ_1, θ_2), the (X, Z) and the coefficient vectors of the model at the time that agents make college enrollment decisions.
- Agents do not know the ε's.

Table 20: Agent's Forecast Variance of Present Value of Earnings[1]
Under Different Information Sets
NLSY79

For Lifetime:	$\text{Var}(Y_c)$	$\text{Var}(Y_h)$	$\text{Var}(Y_c - Y_h)$
Total Residual Variance	290.84	103.13	334.02
Share of Total Variance Due to Forecastable Components	65.13%	55.94%	56.04%
Share of Total Variance Due to Unforecastable Components	34.87%	44.06%	43.94%

[1]We use a discount rate ρ of 3% to calculate the present value of earnings.

Source: Cunha and Heckman (2006a).

Evidence on Uncertainty and Heterogeneity of Returns

- They forecast 65% of the variance of the present value of college earnings, 56% of the variance of the present value of high school earnings and 56% of the gross lifetime gain to attending college.
- The unforecastable components are due to uncertainty by the agent at the time schooling decisions are made.
- There are substantial roles for both heterogeneity and uncertainty.

Evidence on Uncertainty and Heterogeneity of Returns

- Once the distinction between heterogeneity and uncertainty is made, it is possible to be precise about the distinction between *ex ante* and *ex post* decision making.
- From their analysis, CH conclude that, at the time agents pick their schooling, the ε's in their earnings equations are unknown to them.
- These are the components that correspond to “luck.”

Evidence on Uncertainty and Heterogeneity of Returns

- It is clear that decision making would be different, at least for some individuals, if the agent knew these chance components when choosing schooling levels, since the decision rule would now be

$$\begin{aligned} S^* &= Y_{1,0} + \frac{Y_{1,1}}{1+r} - Y_{0,0} - \frac{Y_{0,1}}{1+r} - C > 0 \\ S &= 1 \text{ if } S^* > 0;\ S = 0 \text{ otherwise,} \end{aligned}$$

where no expectation is taken to calculate S^* since *ex post* all terms on the right hand side of the top equation are known with certainty by the agent.

Evidence on Uncertainty and Heterogeneity of Returns

- In their empirical model, if individuals could pick their schooling level using their *ex post* information (i.e., after learning their luck components in earnings) 13.81% of high school graduates would rather be college graduates and 17.15% of college graduates would have stopped their schooling at the high school level.
- Using the estimated counterfactual distributions, it is possible to consider a variety of policy counterfactuals on distributions of outcomes locating persons in pre- and post-policy distributions.

Evidence on Uncertainty and Heterogeneity of Returns

- They analyze how tuition subsidies move people from one quantile of a Y_0 distribution to another quantile of a Y_1 distribution.
- See Carneiro, Hansen, and Heckman (2001, 2003), Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006a).

Extensions and Alternative Specifications

- Carneiro, Hansen, and Heckman (2003) estimate a version of this model with complete autarky.
- Individuals have to live within their means each period.
- Navarro (2005) estimates a version of this model with restriction on intertemporal trade as in the Aiyagari-Laitner economy.

Extensions and Alternative Specifications

- Different assumptions about credit markets and preferences produce a range of estimates of the proportion of the total variability of returns to schooling that are unforecastable, ranging from 37% (Carneiro, Hansen, and Heckman, 2003) for complete autarky and log preferences, to 53 – 56% (Cunha, Heckman, and Navarro, 2005; Cunha and Heckman, 2006b) for complete markets, to 44% (Cunha and Heckman, 2006a) for another complete market economy.

Extensions and Alternative Specifications

- This line of research has just begun.
- It shows what is possible with rich panel data.
- The empirical evidence on the importance of uncertainty is still emerging.
- Yet most of the papers suggest a substantial role for uncertainty in producing *ex post* returns.

Extensions and Alternative Specifications

- Accounting for uncertainty and psychic costs may help to explain the high *ex post* rates of return presented in Section 41.
- It may also account for the sluggish response of schooling enrollment rates to rising returns to schooling that is documented in Ellwood and Kane (2000) and Card and Lemieux (2001) because of the wedge between utility and money returns.

Models with Sequential Updating of Information

- We have thus far discussed one shot models of schooling choice.
- In truth, schooling is a sequential decision process made with increasingly richer information sets at later stages of the choice process.

Models with Sequential Updating of Information

- We established that with independent shocks across schooling levels, dynamic selection leads to a downward bias of OLS estimators for rates of return because people with good shocks to their earnings at lower levels of schooling drop out of school.
- We also established that the internal rate of return is not, in general, a valid guide to rational economic decision making.

Models with Sequential Updating of Information

- Keane and Wolpin (1997) and Eckstein and Wolpin (1999) pioneered the estimation of dynamic discrete choice models for analyzing schooling choices.
- They assume a complete market environment and do not entertain a range of alternative market structures facing agents.
- In the notation of this chapter, they assume a one (discrete) factor model with factor loadings that are different across different counterfactual states, but are constant over time ($\alpha_{s,t} = \alpha_s$, $s = 1, \ldots, \bar{S}$ where there are $\bar{S}$ states).

Models with Sequential Updating of Information

- At a point in time, t, $\varepsilon_{s,t}$, $s = 1, \ldots, \bar{S}$ are assumed to be multivariate normal random variables.
- Over time the $\varepsilon_t = \left(\varepsilon_{1,t}, \ldots, \varepsilon_{\bar{S},t}\right)$ are assumed to be independent and identically distributed.
- They assume agents know θ but not the ε_t, $t = 0, \ldots, T$.
- The unobservables are thus equicorrelated over time (age) because the factor loadings are assumed equal over time and ε_t is independent and identically distributed over time.

Models with Sequential Updating of Information

- They make parametric normality assumptions in estimating their models.
- They impose and do not test the particular information structure that they use.
- In their model, about 90% of the variance in lifetime returns is predictable at age 16.
- Their estimate of the predictability of future earnings is much higher than any of the estimates surveyed in the previous section.

Models with Sequential Updating of Information

- Heckman and Navarro (2006) formulate and identify semiparametric sequential schooling models based on the factor structures exposited previously.
- They present a new semiparametric identification analysis for this class of models.
- Like Keane and Wolpin, they assume a complete contingent claims market.
- In preliminary research, they report substantial effects on empirical estimates from relaxing normality assumptions.

Models with Sequential Updating of Information

- Instead of assuming a particular information structure, they test among alternative information structures about the arrival of information on the components of vector θ at different stages of the lifecycle.
- Their work supports the analysis of CHN and CH in finding a sizable role for heterogeneity (predictable variability) in accounting for measured variability.
- They estimate the sequential reduction of uncertainty as information is acquired.

Part XI

Summary and Conclusions

Summary and Conclusions

- Since the seminal work of Becker (1964), economists have sought to estimate the rate of return to schooling to determine whether there is underinvestment or overinvestment in education.
- The quest continues to this day, and the data available to estimate it have greatly improved.
- This chapter reviews the body of literature that has emerged on estimating returns to schooling over the past 40 years, and how access to better data has improved estimates of the rate of return.

Summary and Conclusions

- Mincer's early efforts suggested one way of estimating mean rates of return and distributions of rates of return on widely available Census and CPS cross section data.
- Mincer's earnings equation still serves as the point of departure for most empirical studies of the returns to school.
- His analysis provides a basic theoretical underpinning for estimating the internal rate of return to education using regressions of log earnings on schooling and a separable quadratic function in experience.

Summary and Conclusions

- A number of strong assumptions must hold in order to interpret the "Mincer coefficient" (i.e., the coefficient on schooling in a log earnings equation) as an internal rate of return.
- While many of these assumptions turn out to hold in the 1960 data for the U.S. labor market that he analyzed (e.g. separability in education and experience, log-linearity of earnings in schooling, negligible tuition costs of school, and negligible taxes), this chapter shows that in recent U.S. data they no longer hold.
- After documenting evidence against Mincer's assumptions, we consider alternative approaches to estimating the marginal internal rate of return to schooling across different schooling levels.

Summary and Conclusions

- We estimate general nonparametric earnings functions and generate from them marginal internal rates of return that account for taxes and tuition.
- The levels and time series patterns of marginal internal rates of return differ dramatically from those produced by a Mincer model.
- Deviations from parallelism and linearity in schooling in log earnings equations—keystones of the Mincer approach—are quantitatively important in determining internal rates of return, as are the effects of taxes and tuition.

Summary and Conclusions

- Economists cannot continue to pretend that violations of the required assumptions are innocuous when using Mincer regressions to estimate ‘returns to schooling’.
- Although we report estimates based on U.S. data, we conjecture that similar problems with Mincer’s assumptions apply to many other countries.
- Replication of our study on data from other countries would be highly desirable.

Summary and Conclusions

- Our analysis shows how to use nonparametric earnings profiles reported in the recent literature to estimate rates of return.
- The recent literature surveyed in Katz and Autor (1999) establishes that the payment to college graduates has gone up relative to that of high school graduates in the past two decades.
- It does not determine whether rates of return have increased.

Summary and Conclusions

- We show that using the Mincer estimate of the rate of return misrepresents trends in actual rates of return, because of misspecification of the earnings-schooling-experience relationship and because of neglecting components of the return such as tuition costs and taxes.
- It also leads to inaccurate estimates of earnings associated with different schooling levels.

Summary and Conclusions

- The standard representative agent income maximizing model that serves as the foundation for many analyses of returns to schooling motivated by economic theory suggests that marginal internal rates of return should be the same across observed schooling choices and should equal the common real interest rate faced by students.
- Yet, our reported estimates of the return to high school and college completion for recent years are substantially larger than the real interest rates faced by consumers, even on credit card debt.

Summary and Conclusions

- One possible explanation for this disparity is the failure of the income maximizing concept, rather than the utility maximizing concept, to represent schooling decisions.
- Psychic costs or distaste for schooling may explain why more than fifteen percent of new cohorts of American youth do not receive a high school degree despite its high estimated financial return.

Summary and Conclusions

- Results from Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006a,b) discussed above show high psychic cost components estimated under different assumptions about the economic environments facing agents.
- Although in theory substantial credit constraints could explain the patterns of college-going decisions, recent research finds them to be quantitatively unimportant in the U.S. economy (see the survey by Cunha, Heckman, Lochner, and Masterov, 2006), and the estimates of high psychic costs are robust to alternative assumptions about credit markets.

Summary and Conclusions

- Heckman, Stixrud, and Urzua (2006) establish the importance of both cognitive and noncognitive skills in explaining schooling decisions, wages and a variety of risky behaviors.
- Psychic costs are related to both cognitive and noncognitive skills.
- They discuss, but do not definitively establish, the link between psychometric measures of cognitive and noncognitive skills and conventional measures of risk aversion, preference for leisure and time preference that would be a more satisfactory foundation for explaining "psychic costs."

Summary and Conclusions

- Mincer and many other researchers use cross sections of earnings to estimate life-cycle earnings of the various cohorts sampled in the cross-section, the so-called synthetic cohort approach.
- This practice is problematic when labor markets are nonstationary as in recent years.
- The use of repeated cross-section or panel data that follow the experience of actual cohorts is essential for accurately measuring rates of return to schooling.

Summary and Conclusions

- However, use of repeated cross section data does not produce lower estimated returns.
- If anything, the return from repeated cross section data is higher, leaving the puzzle of high estimated returns to schooling in place.

Summary and Conclusions

- If analysts seek to estimate *ex post* returns, a cohort analysis is clearly preferred to a cross-section approach.
- However, if analysts are interested in estimating *ex ante* returns in a changing economic environment, the choice is less clear cut.
- Expectations about the future need to be specified or, better, estimated or measured.

Summary and Conclusions

- We summarize an emerging literature that moves beyond estimating mean growth rates of earnings with schooling or treatment effects to estimate distributions of growth rates and rates of return.
- This approach is based on the principle that dependence across counterfactual distributions is generated by low dimensional unobservables.
- The new methods can be implemented using panel data on earnings and schooling.
- Access to test scores or other proxies for the latent factors facilitates identification of the distribution of rates of return.

Summary and Conclusions

- Application of the new methods to rich panel data allows analysts to disentangle uncertainty from measured variability.
- We review evidence from Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006a,b), who develop and implement an approach for empirically distinguishing *ex ante* from *ex post* returns to schooling using rich panel data.
- They find that uncertainty about the future is empirically important for understanding schooling decisions.

Summary and Conclusions

- To the extent that individuals are risk averse, the evidence on uncertainty helps to explain some of the high estimated returns to schooling reported in Section 41 and in the entire literature (see Navarro, 2005; Cunha and Heckman, 2006b).
- At the same time, a substantial amount of observed variability in earnings is predictable at the date schooling decisions are made.

Summary and Conclusions

- In a dynamic setting, uncertainty about future earnings and schooling outcomes creates a wedge between *ex post* average rates of return and real interest rates due to the option value of continuing on in school and updating information.
- For example, some individuals may attend college, knowing that the expected returns to only a few years of college are low but the expected returns from finishing college are quite high.

Summary and Conclusions

- Even if college graduation is not certain, many individuals may be willing to take the gamble of attending with the hope that they will finish successfully.
- Our estimates of low returns to college attendance and high returns to college completion are consistent with this story.

Summary and Conclusions

- Our analysis of option values raises questions about the internal rate of return—a pillar of classical human capital theory—as a useful measure of returns to schooling.
- In a model with uncertainty and sequential decision making, there may be many discount rates that equate theoretically correct value functions across different schooling choices.
- The validity of internal rate of return measures depends crucially on the amount of uncertainty in future earnings associated with different education levels.

Summary and Conclusions

- The recent literature finds a substantial amount of predictability in future earnings and empirical estimates of option values that are relatively small.
- This mitigates concerns about using internal rates of return as a criterion for evaluating educational policy.
- However, work on this topic has just begun, so any conclusion about the empirical importance of option values has to be tempered with caution.

Summary and Conclusions

- The most common criticism directed against the Mincerian approach questions the strong assumption that individuals making different schooling choices are *ex ante* identical (see, e.g. Griliches, 1977; Willis, 1986; Willis and Rosen, 1979; Card, 1995, 1999; Heckman and Vytlacil, 1999, 2005; Carneiro, Hansen, and Heckman, 2003; Carneiro, Heckman, and Vytlacil, 2005).
- The recent literature that attempts to address the consequences of heterogeneity on estimated rates of return focuses on mean growth rates of earnings with schooling and not on true rates of return.

Summary and Conclusions

- Card's (1995, 1999) version of Becker's Woytinsky lecture offers a useful framework for analyzing growth rates in earnings in a heterogeneous world.
- Under strong assumptions that schooling choice equations are linear in growth rates and in costs of schooling, instrumental variable methods can be used to identify the average effect of schooling on earnings.

Summary and Conclusions

- However, researchers are often interested in other treatment parameters that can be directly linked to the effects of a particular policy intervention.
- These parameters are not typically estimated by instrumental variable estimators.
- Since schooling is a discrete outcome, traditional instrumental variables methods produce parameters that are instrument-dependent and are rarely economically interpretable.

Summary and Conclusions

- The empirical debate on the importance of accounting for the endogeneity of schooling in estimating rates of return is far from settled.
- Much of this literature does not estimate rates of return but instead focuses on various treatment effects.
- An entire recent literature has directed attention away from estimating rates of return, or other economically interpretable parameters, toward estimating the probability limits of *IV* estimators which often lack any economic interpretation.
- Many of the popular instruments are weak and the *IV* literature has lost sight of estimating distributions of returns.

Summary and Conclusions

- Much of the recent literature has focused on the rising returns to college.
- The estimates presented in this chapter suggest a substantially greater increase in the returns to high school, raising the obvious questions: why do so many individuals continue to drop out of high school and why is the correctly measured high school dropout rate increasing? The answer may rely on high "psychic" costs of school, credit constraints, risk and uncertainty, or unobserved differences in ability between dropouts and graduates.

Summary and Conclusions

- It remains to be established whether the enormous increase in the returns to high school in recent decades estimated using the internal rates of return implicit in the recent Census-CPS literature can be explained by changes in ability differences between high school dropouts and graduates.
- The relatively slow growth in high school dropout rates since 1970 and the continued increase in rates of return to high school (as measured by cross-section or cohort-based estimates) since that time poses a serious challenge to simple explanations based on this premise.

Summary and Conclusions

- The new literature is beginning to sort out these competing explanations.
- Recent developments in the literature employ new methods to take advantage of rich longitudinal microdata in order to begin distinguishing among the many possibilities.

Summary and Conclusions

- With better tools and better data, the conventions of 1960s labor economics should no longer guide estimation of rates of return to schooling in the 21st century.
- The Mincer model is no longer a valid guide to estimating the returns to schooling or accounting for heterogeneity in returns.
- The modern *IV* literature aims to recover growth rates of earnings with schooling, allowing for heterogeneity, but has lost sight of the economic questions posed by Mincer.
- Recent developments in econometrics and the economics of education coupled with rich panel data make it possible to estimate economically interpretable parameters including true *ex ante* and *ex post* rates of return to schooling and their distributions in the population.

Appendix

Census Data

- The Census samples used in this chapter are taken from the 1940, 1950, 1960, 1970, 1980 and 1990 Public-Use Census Samples.
- The 1940 sample consists of the self-weighting subsample which represents 1% of the population.
- The 1950 sample consists of sample-line persons (for whom questions regarding earnings were asked) which represent about 0.303% of the population.
- The 1960 sample is a self-weighting 1% sample.

Census Data

- The 1970 sample is taken from two Public-Use A samples: the 1% State sample (5% form) and the 1% State sample (15% form).
- It is a self-weighting sample of 2% of the population.
- The 1980 and 1990 Census samples are both 5% Public Use A samples.
- The 1980 sample is self-weighting but the 1990 sample is not.
- For 1990, we use person weights to re-weight the sample back to random proportions.

Census Data

The following sample restrictions are imposed for each Census year:

age

Sample includes individuals age 16-64. For Census years when a **quarter-of-birth** variable is available, we take into account the quarter of birth in calculating the age of each individual from the *year of birth* variable provided in the data set.

race

Only individuals reported as being black or white are included in the analysis.

Census Data

earnings

The earnings measure used is annual earnings, which includes both wage and salary and business income for the Census years when business income is available. For Census years when earnings are reported in intervals, we use the midpoint of the interval as the individual's earnings.

imputations

Individuals with imputed information on age, race, sex, education, weeks worked or income are excluded. For years when all the imputation flags are not provided, we omit individuals on the basis of the available imputation flags.

Census Data

The following variables are constructed:

experience

Potential experience is measured by Age − Years of Education − 6.

Census Data

years of education

For the 1940-1980 Censuses, years of education are reported as the highest grade completed. For the 1990 Census, years of education are reported differently: by categories for first through fourth grade and for fifth through eighth grade, by year for ninth through 12th grade, and then by degree attained. To maintain comparability with the other Census samples, we impute the number of years of school associated with each category or degree. For those with some college but no degree or for those with an associate degree, we assign 14 years of school. For those with a bachelor's degree, we assign 16 years of school. For professional degrees we assign 17 years and for masters degrees and beyond, including doctoral degrees, we assign 18 years of school.

Current Population Survey (CPS) Data

- The CPS samples used in this chapter are taken from the 1964-2000 CPS March Supplements.

The following sample restrictions are used for each year:

age

Sample includes individuals age 18-65.

race

Sample separated into whites and all non-whites.

earnings

Annual wage and salary income (deflated using the CPI-U) is used as the earnings measure in each year.

Current Population Survey (CPS) Data

The following variables are constructed for our analysis:

experience

Potential experience is measured by Age - Years of Education - 6.

years of education

For 1964-1991, years of education are reported as the highest grade completed. Categories of schooling include 9-11 years, 12 years, and 16 years. From 1992-2000, years of education are reported differently. Those completing 12 years of schooling but who do not receive a high school diploma are assigned 11 years. Only those with 12 years of schooling and a diploma are assigned 12 years of schooling. For those with a bachelor's degree, we assign 16 years of school.

Tuition Time Series

- To estimate the private cost of college, we use the time series Total Revenue from Student Fees and Tuition obtained from Snyder (1993, Table 33).
- Tables 24 and 33 of this publication provide, for all institutions, statistics on total educational revenue, total tuition revenue, and total enrollment.
- We divide total revenue for all institutions by total enrollment.
- Supplementing this data with data from Snyder (2000, Tables 175 and 331), we create a consistent time series of total educational revenue, total tuition revenue, and total enrollment for 1940-1995.

Tax Rate Time Series

- We obtain the average marginal tax rate time series from Barro and Sahasakul (1983) and Mulligan and Marion (2000, Table 1, column 1).
- The tax rates used in our progressive tax analysis are obtained from the federal schedule for a single adult with no dependents.
- All income is assumed to be earned income and standard deductions are assumed.
- To obtain after-tax income for 1960-90, we use the TAXSIM version 4.0 program available at http://www.nber.org/taxsim/taxsim-calc4/index.html.
- For 1940 and 1950, we use the actual federal tax schedules (Form 1040) as reported in the Statistics of Income.

Local Linear Regression

- In estimating the nonparametric regressions, we use local linear regression methods.
- As discussed in Fan and Gijbels (1996), the local linear estimator for the conditional expectation $E\left[y_i \middle| x_i = x_0\right]$ can be computed from the minimization problem

$$\min_{a,b} \sum_{i=1}^{n} \left(y_i - a - b_1 \left(x_i - x_0\right)\right)^2 K\left(\frac{x_i - x_0}{h_n}\right),$$

where $K(\cdot)$ is a kernel function and $h_n > 0$ is a bandwidth which converges to zero as $n \to \infty$.

Local Linear Regression

- The estimator of the conditional mean $E\left[y_i \mid x_i = x_0\right]$ is $\widehat{a}$.
- The local linear estimator can be expressed as a weighted average of the y_i observations, $\sum_{i=1}^{n} y_i W_i(x_0)$, where the weights are

$$W_i(x_0) = \frac{K_i \sum_{j=1}^{n} K_j^2 - K_i \sum_{k=1}^{n} K_k}{\sum_{k=1}^{n} K_k \sum_{j=1}^{n} K_j^2 - \left(\sum_{k=1}^{n} K_k\right)^2}.$$

Local Linear Regression

- Our local regression estimator is given by

$$\hat{m}(x_0) = \sum_{i=1}^{N} y(x_i) W_i(x_0),$$

where $y(x_i)$ represents log earnings at experience level x_i and N represents the number of observations.

Local Linear Regression

- The asymptotic distribution of the estimator $\hat{m}(x_0)$ for $m(x_0) = E(y_i|x_i = x_0)$ is given by

$$\sqrt{nh_n}\,(\hat{m}(x_0) - m(x_0)) \sim N(B_n, V_n) + o_p(1)$$

where the bias and variance expressions are given by

$$\begin{aligned} B_n &= h_n^2 \cdot (0.5m''(x_0)) \cdot \int_{-\infty}^{\infty} u^2 K(u)du \\ V_n &= \frac{\sigma^2(x_0)}{f(x_0)} \int_{-\infty}^{\infty} K^2(u)du, \end{aligned}$$

where $\sigma^2(x_0) = E(\{y_i - E(y_i|x_i = x_0)\}^2|x_i = x_0)$ and $f(x_0)$ is the density of x_i at x_0.

Tests of Parallelism

- In Section 13 of this chapter, we perform nonparametric tests of whether the log-earnings-experience profiles are parallel across schooling levels.
- Let s_1 and s_2 denote two different schooling levels (16 years and 12 years, for example).
- We test whether

$$E(y_i|x_i, s = s_1) - E(y_i|x_i, s = s_2)$$
$$= \text{constant across } x_i \in \{10, 20, 30, 40 \text{ years}\}$$

We select the experience values at which the hypothesis is tested to be at least 2 bandwidths apart from the other experience levels, so that the nonparametric estimates are independent from one another.

Tests of Parallelism

- Let $\hat{m}(x_i, s_1)$ denote the estimator for $E(y_i|x_i, s = s_1)$ for experience level x_i and schooling level $s = s_1$.
- The test statistic for testing parallelism for two different schooling levels s_1 and s_2 and two experience levels x_i and x_k is given by

$$(\hat{m}(x_i, s_1) - \hat{m}(x_i, s_2) - (\hat{m}(x_k, s_1) - \hat{m}(x_k, s_2)) \cdot$$
$$(\hat{V}_1 + \hat{V}_2 + \hat{V}_3 + \hat{V}_4)^{-1} \cdot$$
$$(\hat{m}(x_i, s_1) - \hat{m}(x_i, s_2) - (\hat{m}(x_k, s_1) - \hat{m}(x_k, s_2)),$$

  where $\hat{V}_1$, $\hat{V}_2$, $\hat{V}_3$, and $\hat{V}_4$ are estimators for $V_1 = \mathrm{Var}(\hat{m}(x_i, s_1))$, $V_2 = \mathrm{Var}(\hat{m}(x_i, s_2))$, $V_3 = \mathrm{Var}(\hat{m}(x_k, s_1))$, $V_4 = \mathrm{Var}(\hat{m}(x_k, s_2))$.

Tests of Parallelism

- To estimate the variances, we use

$$\mathrm{Var}(\hat{m}(x_i, s_\ell)) = \sum_{i=1}^{N} \hat{\varepsilon}(x_i, s_\ell)^2 \, W_i^2(x_i), \qquad \ell = 1, 2,$$

  where $\hat{\varepsilon}(x_i, s_\ell) = y(x_i, s_\ell) - \hat{m}(x_i, s_\ell)$, $\ell = 1, 2$, is the fitted residual from the nonparametric regression evaluated at experience level x_i.

- In Table 1, we report test results based on the test statistic that straightforwardly generalizes the test statistic given above to four experience levels.