ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Heckman, James Joseph; Lochner, Lance J.; Todd, Petra E.

Working Paper

Earnings functions, rates of return and treatment effects: the Mincer equation and beyond

IZA Discussion Papers, No. 1700

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Heckman, James Joseph; Lochner, Lance J.; Todd, Petra E. (2005) : Earnings functions, rates of return and treatment effects: the Mincer equation and beyond, IZA Discussion Papers, No. 1700, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/33472

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 1700

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

James J. Heckman
Lance J. Lochner
Petra E. Todd

August 2005

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

James J. Heckman
University of Chicago, University College London, American Bar Foundation and IZA Bonn

Lance J. Lochner
University of Western Ontario

Petra E. Todd
University of Pennsylvania and IZA Bonn

Discussion Paper No. 1700
August 2005

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 1700
August 2005

ABSTRACT

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

Numerous studies regress log earnings on schooling and report estimated coefficients as "Mincer rates of return". A more recent literature uses instrumental variables. This chapter considers the economic interpretation of these analyses and how the availability of repeated cross section and panel data improves the ability of analysts to estimate the rate of return. We consider under what conditions the Mincer model estimates an ex post rate of return. We test and reject the model on six cross sections of U.S. Census data. We present a general nonparametric approach for estimating marginal internal rates of return that takes into account tuition, income taxes and forms of uncertainty. We also contrast estimates based on a single cross-section of data, using the synthetic cohort approach, with estimates based on repeated cross-sections following actual cohorts. Cohort-based models fitted on repeated cross section data provide more reliable estimates of ex post returns. Accounting for uncertainty affects estimates of rates of return. Accounting for sequential revelation of information calls into question the validity of the internal rate of return as a tool for policy analysis. An alternative approach to computing economic rates of return that accounts for sequential revelation of information is proposed and the evidence is summarized. We distinguish ex ante from ex post returns. New panel data methods for estimating the uncertainty and psychic costs facing agents are reviewed. We report recent evidence that demonstrates that there are large psychic costs of schooling. This helps to explain why persons do not attend school even though the financial rewards for doing so are high. We present methods for computing distributions of returns ex ante and ex post. We review the literature on IV estimation. The link of the estimates to the economics is not strong. The traditional instruments are weak, and this literature has not produced decisive empirical estimates. We exposit new methods that interpret the economic content of different instruments within a unified framework.

JEL Classification: C31

Keywords: rate of return to schooling, internal rate of return, uncertainty, psychic costs, panel data, distribution

Corresponding author:

James J. Heckman
University of Chicago
Department of Economics
1126 E. 59th Street
Chicago IL 60637
USA
Email: jheckman@uchicago.edu

* The first part of this chapter was prepared in June 1998. It previously circulated under the title "Fifty Years of Mincer Earnings Regressions". Heckman's research was supported by NIH R01-HD043411, NSF 97-09-873, NSF SES-0099195 and NSF SES-0241858. We thank George Borjas, Pedro Carneiro, Flavio Cunha, Jim Davies, Reuben Gronau, Eric Hanushek, Lawrence Katz, John Knowles, Mario Macis, Derek Neal, Dan Schmierer, Jora Stixrud, Ben Williams, Kenneth Wolpin, and participants at the 2001 AEA Annual Meeting, the Labor Studies Group at the 2001 NBER Summer Institute, and participants at Stanford University and Yale University seminars for helpful comments. Our great regret is that Sherwin Rosen, our departed friend and colleague, who thought long and hard about the issues discussed in this chapter, could not give us the benefit of his wisdom.

1 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.

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. 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.

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. 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.[1]

This model is widely used as a vehicle for estimating "returns" to schooling quality,[2] for measuring the impact of work experience on male-female wage gaps,[3] and as a basis for economic studies

[1] See, *e.g.* Psacharopoulos (1981), Psacharopoulos and Patrinos (2004) and Willis (1986) for extensive surveys of Mincer returns.

[2] See Behrman and Birdsall (1983) and Card and Krueger (1992).

[3] See Mincer and Polachek (1974).

of returns to education in developing countries.[4] 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.[5]

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.

This chapter makes the following points:

(1) We test important predictions underlying the Mincer model using six waves of U.S. Census data, 1940-1990.[6] We find, as does other recent literature, that Mincer's original model fails to capture central features of empirical earnings functions in recent decades. Our 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.

(2) In response to the evidence against the Mincer model, 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. 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.

The estimated marginal rates of return are often implausible, calling into question the empirical

[4] See Glewwe (2002).
[5] See Bils and Klenow (2000).
[6] Mincer's analysis focused on 1960 U.S. census data (earnings for 1959).

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.

(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. 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.

(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.

(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.[7] 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.[8]

[7] Weisbrod (1962) developed the concept of the option value of schooling. For one formalization of his analysis, see Comay, Melnik, and Pollatschek (1973).

[8] Schooling choices are made sequentially. Thus if the function relating the value of completing schooling at each year of schooling is nonconcave, the return to one stage may be low but the return to the next stage may be high, hence creating an option value at the stage with low terminal payoff. The earlier stage must be completed to obtain the higher return arising at the later stage.

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. 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, 1970). 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. 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.

(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. 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. 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.

(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. 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.

(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. 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 schooling do not do so.

In this chapter, 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 not a valid guide for accurately estimating rates of return to schooling, the Mincer vision of using economics to explain earnings data remains valid.

This chapter proceeds in the following way. Section 2 reviews 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. Section 3 presents 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. Section 4 extracts 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.

Section 5 introduces 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. Section 6 considers 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.

Section 7 discusses 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. Section 8 discusses 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.

Section 9 surveys a recent literature that estimates distributions of *ex post* returns. Section 10 decomposes 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. Section 11 concludes.

2 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:

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

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 $E(\varepsilon|s,x) = 0$.[9] This regression model is motivated by two conceptually different frameworks used by Mincer (1958, 1974). While algebraically similar, their economic content is very different. In section 3, we formally test and reject predictions of these models on the type of Census data originally used by Mincer. In section 4, we implement a more general nonparametric approach to estimating internal rates of return that does not require an explicit model specification.

2.1 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. 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 section 10 shows are important determinants of the return to schooling and its distribution.

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

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

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

[9] Psacharopoulos (1981) and Psacharopoulos and Patrinos (2004) provide surveys of an enormous Mincer-based earnings literature.

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.[10]

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. 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.

2.2 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). 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.[11]

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

[10] This term also disappears if the retirement age, T, is allowed to increase one-for-one with s (*i.e.*, $\frac{\partial T(s)}{\partial s} = 1$), so post-school working life is the same for persons of all schooling levels.

[11] Chiswick and Mincer (1972) explicitly analyze income inequality with this model. We discuss earnings distributions and distributions of rates of return in section 10.

age t. Potential earnings at t are

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

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. 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.

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

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

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

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

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

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. 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,

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

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 as

$$\ln Y(s, x) = \bar{\alpha} + \bar{\rho}_s s + \bar{\beta}_0 x + \bar{\beta}_1 x^2 + [(\alpha - \bar{\alpha}) + (\rho_{si} - \bar{\rho}_s)s + (\beta_{0i} - \bar{\beta}_0)x + (\beta_{1i} - \bar{\beta}_1)x^2],$$

where the terms in brackets are part of the error.[12] Mincer originally assumed that $(\alpha - \bar{\alpha}), (\rho_{si} - \bar{\rho}_s), (\beta_{0i} - \bar{\beta}_0), (\beta_{1i} - \bar{\beta}_1)$ 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 in sections 7 and 8.

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. Assuming that post-school investment patterns are identical across persons

[12] In the random coefficients model, the error term of the derived regression equation is heteroskedastic.

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

and

(ii) log-earnings age profiles diverge with age across schooling levels ($\frac{\partial \ln Y(s,x)}{\partial s \partial t} = \frac{\rho_0 \kappa}{T} > 0$ *).*

In section 3, 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.[13] 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.

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.

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 are uncorrelated, the variance of log earnings reaches a minimum when experience is approximately equal to $1/\rho_0$. (See the derivation in Appendix A.) At this experience level, variance in earnings is solely a consequence of differences in schooling levels or ability and is unrelated to

[13] Mincer (1974) provided informal empirical support for the implications using 1960 census data.

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. 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.[14]

3 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.[15]

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 in section 6.

Nonparametric local linear regression is used to generate the estimates.[16] 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. 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

[14] In addition to Mincer (1974), studies by Schultz (1975), Smith and Welch (1979), Hause (1980), and Dooley and Gottschalk (1984) also provide evidence of this pattern for wages and earnings.

[15] Business income is not available in the 1940 Census. Appendix B provides detailed information on the construction of our data subsamples and variables.

[16] Details about the nonparametric estimation procedure are given in Appendix C. The bandwidth parameter is equal to 5 years. Estimates are not very sensitive to changes in the bandwidth parameter in the range of 3-10 years.

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.

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.[17] 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.[18]

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 $< .001$).[19]

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. 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.[20] 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. We discuss this issue in section 10.

[17] Murphy and Welch (1992) also document differences in earnings-experience profiles across education levels using data from the 1964-1990 Current Population Surveys.

[18] The formulae for the test statistics are given in Appendix C.

[19] It is also rejected for nonparametric specifications of the experience term. These results are available on request from the authors.

[20] The U-shaped profile of the variance of earnings argues against the Rutherford (1955) model of earnings as revived by Atkeson and Lucas (1992).

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. 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.

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.

4 Estimating Internal Rates of Return

Given the evidence against the validity of the Mincer earnings specification presented in section 3 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. 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 in section 4.2 below, and also later in section 10.

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 in section 2, such as the restriction that log earnings increase linearly with schooling and the restriction that log earnings-experience profiles are parallel across schooling classes.

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. 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.[21]

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.[22] Individuals are assumed to choose s to maximize the present discounted value of lifetime earnings,[23]

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

The first order condition for a maximum yields

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

[21] To estimate social rates of return, we need to account for the social opportunity costs of funds and full social returns including crime reduction. See Lochner and Moretti (2004).

[22] The standard framework implicitly assumes that individuals know these functional relationships, credit markets are perfect, education does not enter preferences, and there is no uncertainty.

[23] This expression embodies an institutional feature of the U.S. economy where income from all sources is taxed but one cannot write off tuition and non-pecuniary costs of education. However, we assume that agents can write off interest on their loans. This assumption is consistent with the institutional feature that persons can deduct mortgage interest, that 70% of American families own their own homes, and that mortgage loans can be used to finance college education. The expressions based on (2) can easily be modified to account for other tax treatments of tuition.

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

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

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.

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. In light of the substantial estimates of psychic costs presented in Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005a,b,c,d), 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 in section 10. Accounting for these psychic costs lowers the internal rate of return.

Consider the additional common assumption that $T'(s) = 1$ (*i.e.*, no loss of work life from schooling). This simplifies the first order condition to

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

As noted in section 2, 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). 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.

From equation (4) we observe, more generally, that the difference between after-tax interest rates 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). 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.

The evidence for 1980 and 1990 presented in section 3 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.

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.

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.[24] Typically, internal rates of return are based on non-marginal differences in schooling. Incorporating tuition (and psychic costs) and taxes, the

[24]Becker (1964) states this logic and Hanoch (1967) applies it.

internal rate of return for schooling level s_1 versus s_2, $r_I(s_1, s_2)$, solves

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

As with $\tilde{r}$ above, r_I will equal the Mincer coefficient on schooling under the assumptions of parallelism over 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 ($T'(s) = 1$).[25] In the next section, we compare rate of return estimates based on specification (1) to those obtained by directly solving for $r_I(s_1, s_2)$ in equation (5).

4.1 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.[26] 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. Then, we predict earnings under this specification for the first 47 years of experience, and the IRR

[25] When tuition and psychic costs are negligible, proportional taxes on earnings will have no effect on estimated internal rates of return, because they reduce earnings at the same rate regardless of educational choices.

[26] As lower schooling levels are reported only in broader intervals in the 1990 Census, we can only compare 6 years against 10 years and cannot compare 6 years against 8 years or 8 against 10 years as we do for the earlier Census years. We assume the private cost to elementary and high school is zero in all the calculations.

is taken to be the root of equation (5).[27] 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.[28]

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. 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). Sheepskin effects are an important feature of the data.[29] 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.) 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.

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. This procedure does not impose any assumption other than continuity on the earnings-experience relationship. 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 reported in section 3, which show that earnings profiles in recent decades are no longer parallel in experience across schooling categories.

[27] Strictly speaking, we solve for the root of the discrete time analog of equation (5).

[28] They would be identical if our internal rate of return calculations were computed in continuous time. Because we use discrete time to calculate internal rates of return, $r_I = e^{\rho_s} - 1$, which is approximately equal to ρ_s when it is small.

[29] We use the term "sheepskin effects" to refer to exceptionally large rates of return at degree granting years of schooling. We cannot, however, distinguish in some years of the Census data which individuals receive a diploma among individuals reporting 12 or 16 years of completed schooling.

The general estimates in Table 3 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. 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.[30] Sections 7 and 8 consider estimation when schooling choices are endogenous.

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. 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.

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).

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 curve, but do not report any effects of relaxing the quadratic-in-experience assumption on estimated marginal rates of return to schooling. We find that the quadratic-in-experience assumption is fairly innocuous; the assumptions of linearity

[30] Though, it is worth noting that the fraction of white men completing high school is relatively stable after 1970. Among black men, high school graduation rates continued to increase until the early 1980s. Heckman, Lyons, and Todd (2000), Chandra (2003) and Neal (2004) show the importance of selection adjustments in estimating wage functions, but there have been few adjustments of rates of return for selection. This important topic is neglected in the recent literature.

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.

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.[31] Below, in section 10, 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. (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).[32] Row three 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. The final row 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. (See Appendix B for details.)

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

[31] Because we assume that schooling is free (direct schooling costs are zero) through high school and because internal rates of return are independent of flat taxes when direct costs of schooling are zero, internal rates of return to primary and secondary school are identical across the first three specifications in the table. Empirically, taking into account progressive tax rates has little impact on the estimates for these school completion levels. (Tables are available upon request.) For these reasons, we only report in Table 4 the IRR estimates for comparisons of school completion levels 12 and 14, 12 and 16, and 14 and 16.

[32] Average college tuition was computed by dividing the total tuition and fees revenue in the U.S. by total college enrollment that year. Federal and state support are not included in these figures. See Appendix B for further details on the time series we used for both tuition and taxes. We lack data on psychic costs, although the estimates from structural models suggests that they may be sizeable. See Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005d).

blacks. 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. 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. As discussed in section 10, however, accounting for psychic costs has a substantial effect on estimated rates of return.[33]

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 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. 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.

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.*, $T'(s) = 0$). We find virtually identical results for all years and schooling comparisons for both assumptions about the schooling - worklife relationship.[34] 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.

[33] We do not produce any estimates of the time series of psychic costs. We only report a one time snapshot. Thus we do not know whether psychic costs increased or decreased.

[34] Results available from authors upon request.

4.2 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.

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.

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

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

Assume for the sake of argument (but contrary to the evidence in section 3) 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. 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.[35] Individuals take fitted log earnings profiles as predictions for their own future earnings, ignoring any potential person-specific deviations from that profile. Ignoring taxes, for this case, the IRR estimator $\hat{r}_I$ solves

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

[35] Assuming a symmetric distribution for ε, this is equivalent to placing themselves at the median of the earnings distribution.

which is the discrete time analogue to the model of equation (2) for two schooling levels s and $s+j$.[36] If tuition and psychic costs are negligible ($v=0$),

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

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.

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

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

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 plim $\hat{r}_I = e^{\rho_s} - 1$ as before. 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.

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

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

[36] We assume here that $T(s) - s = T$ for all s, or that $T'(s) = 1$.

where

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

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. We discuss how to decompose variability into predictable and unpredictable components in section 10. 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 $M(s,j) < 1$.[37]

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. 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.

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

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

[37] More generally if $v \neq 0$, then $\hat{r_I}$ converges to the root of equation (6). Neglecting this term leads to an upward bias, as previously discussed.

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.

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).

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. 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.

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 for $\varepsilon = 0$ ("unadjusted earnings") and (ii) that they forecast using mean earnings within each education and experience category rather than using predicted earnings placing themselves at $\varepsilon = 0$ ("adjusted earnings"). 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.[38]

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. In other words, predicted

[38] We lack the required panel data on individuals to compute *ex post* rates of return. See the discussion in section 10.

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.

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). 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, 2005d; Heckman and Navarro, 2005) but require panel data and cannot be applied to Census cross-sections. We review the evidence from the panel literature in section 10.

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.

5 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.[39] 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 estimates of option values from the recent dynamic literature in

[39] See Keane and Wolpin (1997), Belzil and Hansen (2002), Cunha, Heckman, and Navarro (2005d) and Heckman and Navarro (2005) for such models.

section 10.[40]

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 which 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. 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. This result runs counter to the intuitions 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.

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. 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 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

[40] See Heckman and Navarro (2005).

will include the expected value of newly revealed information that can be acted on. 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 future information. 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. 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 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).

The empirical evidence presented in section 3 (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. 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.

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 estimates of option values from the literature in section 10.

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.[41] Thus they face a lottery where the chance of being admitted to the next round of schooling does not depend on earnings. 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. In this environment, the total expected value of attaining $s \in \{1, 2, \ldots, \bar{S}\}$ 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 *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

[41] They also consider the probability of failing conditional on attending the next grade. The results from such an analysis are quite similar to those discussed here.

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. 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.[42] 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.

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 option value of high school represents

[42] Proof: $V_{\bar{S}} = Y_{\bar{S}}$ at $\bar{S}$, so

$$E_{\bar{S}-2}\left(V_{\bar{S}-1}\right) = \left(1 - \pi_{\bar{S},\bar{S}-1}\right) Y_{\bar{S}-1} + \pi_{\bar{S},\bar{S}-1} \max\left\{Y_{\bar{S}-1}, \frac{Y_{\bar{S}}}{1+r}\right\},$$

since there is no uncertainty about earnings conditional on each schooling level. For proportional earnings growth at rate r, both versions of the Mincer model imply that $Y_s = \frac{1}{1+r} Y_{s+1}$ for all s. Thus, people may differ in their earnings levels and face different individual specific interest rates as in the accounting identity model. They may also face different $\pi_{s+1,s}$. For any sequence of $\pi_{s+1,s}$ and r, we obtain

$$E_{\bar{S}-2}\left(V_{\bar{S}-1}\right) = Y_{\bar{S}-1} = \frac{Y_{\bar{S}}}{1+r}.$$

Backward induction produces $E_{s-2}\left(V_{s-1}\right) = Y_{s-1} = \frac{Y_s}{1+r}$ for all s, which implies no option value for any schooling level.■

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.[43]

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.

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 (2005). Suppose that there is uncertainty about net earnings conditional on s, so that actual lifetime earnings for someone with s years of school are

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

This form of uncertainty is a one time, schooling-specific shock. The literature discussed in section 10 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 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

[43] An influential book by Dixit and Pindyck (1994) defines option values as arising only in an environment of uncertainty. This definition is too restrictive. Options include any extra choices created by completing one stage of schooling beyond stopping and earning at that stage.

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 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 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

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

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

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

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

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:

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

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 *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 direct costs of schooling C_s, 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.

$\widetilde{R}_{s,s-1}$ 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 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). In section 10, 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.[44]

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. 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, $s+1$.

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.

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.

Keane and Wolpin (1997) develop sequential models of schooling. Although not the focus of their

[44] Our definition of the *ex post* return is a bit ambiguous because at different stages after $s-1$, information about V_s (which is defined over streams of future earnings at different stages of schooling) is revealed. We use V_s in the full information, end of life version when all information is in.

analysis, option values can be derived from the estimated value functions associated with different schooling levels.[45] Heckman and Navarro (2005) present a more general framework for information revelation with serially correlated unobservables and establish semiparametric identification of the model and explicitly estimate option values. We briefly discuss their work in section 10.

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.[46] We assume that $\sigma = 0.1$ in the results presented in the tables. 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. 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 in section 8 predicts an *upward* bias in *OLS* estimates of the return to schooling. In a sequential model, people with a good draw at lower schooling levels drop out, thus producing a downward bias.

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.

[45] In the ordered choice models in Cameron and Heckman (1998) and Heckman and Navarro (2005), there is no option value arising from sequential resolution of uncertainty, because of the assumed one sided nature of the information revelation process.

[46] We also considered models with an AR(1) process for the shocks: $\log(\epsilon_s) = \rho \log(\epsilon_{s-1}) + v_s$ where $v_s \sim N(0, \sigma)$ but for the sake of brevity we do not report them. The case where $\rho = 0$ corresponds to Tables 6a and 6b. For $\rho = 1$, $E(\epsilon_{s+1}|\epsilon_s) = \epsilon_s$ and a good or bad shock affects expected future outcomes in the same proportion as current outcomes. In this model, the outcome of ϵ_s has no effect on schooling decisions. In the linear case corresponding to Table 6a, expected rates of return as measured by $E_{s-1}(R_{s,s-1})$ range between those reported in the table (when ρ is zero) and the linear increase in earnings, $r = 0.10$ (when ρ is near one). Expected returns for the more general non-linear case differ little from those shown in Table 6b, since nearly everyone chooses to attend the highest level of schooling regardless of the value for ρ. This implies that returns always reflect the expected increase in earnings between the current schooling level and the highest possible schooling level, which is, on average, independent of ρ.

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. 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.[47] 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 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. 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.[48]

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

[47] We have assumed that individuals cannot choose to recall the wage streams associated with earlier schooling choices (*i.e.*, someone with s years of school cannot choose to work at a lower schooling level and obtain Y_{s-1} or $Y_{s-2}, \ldots$) if they receive a low realization for Y_s. Allowing for resort to the earnings opportunities created at all earlier schooling opportunities provides a force offsetting the tendency for option values to decline with schooling. These opportunities provide the agent with a form of insurance.

[48] Results available from the authors upon request.

Census years with schooling levels three and five representing high school and college graduation, respectively. 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. 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 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. 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. 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 $\bar{Y}_s E_{s-1}\left(\epsilon_s | \epsilon_s > \frac{E_s(V_{s+1})}{(1+r)\bar{Y}_s}\right)$ rather than $\bar{Y}_s$ (*i.e.*, observed earnings are based on a sample selecting not to continue). 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). 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 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. 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.

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.[49]

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. 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. 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

[49] As information unfolds after s, one could define a sequence of *ex post* value functions depending on what is revealed after stage s.

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.

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 $\beta > 0$.[50] As long as $\alpha_s > \alpha_{s'}$ for all $s > s'$, any two earnings streams will only cross once. Letting $Y_s(a)$ denote the earnings for someone with s years of school at age a, we have

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

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 or value function at age a of someone choosing to attend s_1 years of school with the option of continuing will be $\bar{Y} = (1-p)Y_{s_1} + pY_{s_2}$.

For $\alpha_0 < \alpha_{s_1} < \alpha_{s_2}$, $\bar{Y}$ 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 $s_1 < s < s_2$.[51] This is illustrated in Figure 8. Because there are three crossings between Y_0 and $\bar{Y}$, internal rate of return equations for the value functions produced from this model can generate up to three possible real 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.

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

[50] The example can easily be extended to account for tuition costs and more general lifecycle earnings profiles.

[51] The left hand side of this condition ensures that $\bar{y}$ jumps from zero to some point above $\bar{y}_0$ at age s_1. Then, $\bar{y}$ increases with age at a slower rate than does $\bar{y}_0$. The right hand condition guarantees that at some later age s, $\bar{w}$ will be below $\bar{y}_0$. Finally, we know at age s_2, $\bar{y}$ will jump above w_0, since both α_{s_1} and α_{s_2} are both greater than α_0.

depends on the discount rate. Writing equations out explicitly in terms of interest rate r, we obtain

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

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

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

Take a three period example. In this case, the IRR for the second level of schooling solves

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

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

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.

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.

Both of these cases can arise 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. 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. 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.

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. For an exception, see Cunha, Heckman, and Navarro (2005d) and Heckman and Navarro (2005) 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.

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. Estimates reported in Heckman and Navarro (2005) suggest that these option values are negligible. So our theoretical concerns may be misplaced. Much more work has to be done to see if their work holds up. 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 that ignores the sequential resolution of uncertainty, more general methods of the type presented in section 4 are required to obtain internal rates of return.

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."

6 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. 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. 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.

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.

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. 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. This suggests a preference for cohort-based estimates when changes in cohort quality are taking place.

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). 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.

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. 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. 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. Our estimates also take into account average marginal tax rates and tuition costs using the time series generated from CPS data.[52] 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. 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.[53] Two sets of estimates are generated: (i) regressions are estimated separately for each year of CPS data (to produce a set of cross-section estimates), and (ii) 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 described in the previous section.[54]

Figures 7a and 7b show cohort and cross-section high school and college completion IRR es-

[52] An average marginal tax rate of 25% is assumed for all years after 1994, the final year of tax rates reported in Mulligan and Marion (2000). This corresponds to the average of all rates since 1950, after which rates changed very little from year to year.

[53] In estimating earnings profiles for those with 10 years of education, we combine individuals with 9-11 years, with separate intercept terms for each of the education levels. This is done to increase precision in estimation. See Appendix B for additional details on the coding of the education variables.

[54] In addition to the quadratic specification, we also tried using a cubic and quartic in experience to extrapolate for the missing experience levels. For cohorts with 25 or fewer years of data, extrapolations based on higher order polynomial specifications were unreliable, so we adopted the more parsimonious quadratic specification.

timates for white men, which corresponds to CPS estimates 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.[55] The cohort-based estimates reported in Figure 7a 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. 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. More dramatic differences are observed for the college-going decision of white men as shown in Figure 7b. 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. 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.

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. 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 (2005d) present empirical explorations of alternative expectation-formation

[55] We do not estimate returns for cohorts beyond 1983, since there are too few years of earnings observations for those cohorts to produce stable and reliable estimates.

models. We review methods for estimating agent information sets in section 10.

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. 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.

7 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 also ignores ability bias.[56] 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. 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, although the economic interpretation of the estimates he reports is not at all clear, as is typical of the entire recent instrumental variable literature.

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

[56] See Katz and Autor (1999) for a survey. An exception is Angrist and Krueger (1991). For an analysis of the quality of their instruments see Staiger and Stock (1997).

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).[57] 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

$$\ln y_i = \alpha_i + \rho_i S_i, \tag{11}$$

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. 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 (11) can be written as

$$\ln y_i = \bar{\alpha} + \bar{\rho} S_i + \{\varepsilon_{\alpha_i} + \varepsilon_{\rho_i} S_i\}. \tag{12}$$

Equations (11) and (12) 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.)

Given that the growth rate ρ_i is a random variable, it has a distribution that can be studied using the methods surveyed in sections 9 and 10. 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.

The original Mincer model assumed that the growth rate of earnings with schooling, ρ_i, is

[57] Recall our discussion in section 2.2.

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. 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).

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 explicitly derived from economic theory and is based on the explicit analysis of Becker's model by Rosen (1977).[58] We consider conditions under which it is possible to estimate the effect of schooling on the schooled in his model. The next section considers the more general and recent analysis of Carneiro, Heckman, and Vytlacil (2005).

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

$$U(y, S) = \ln y\,(S) - \varphi(S) \qquad \varphi'(S) > 0 \qquad \text{and} \qquad \varphi''(S) > 0.$$

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. This specification is written in terms of annualized earnings and abstracts from work experience.[59] 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

$$\frac{g'(S)}{g(S)} = \varphi'(S). \tag{13}$$

The term $\frac{g'(s)}{g(s)}$ is the percentage change of earnings with schooling or the "growth rate" at level s.

[58] Random coefficient models with coefficients correlated with the regressors are systematically analyzed in Heckman and Robb (1985, 1986). They originate in labor economics with the work of Lewis (1963). Heckman and Robb analyze training programs but their mathematics applies to estimating the returns to schooling.

[59] Adding work experience in a multiplicatively separable way invokes the Mincer model.

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 $\frac{g'(S)}{g(S)} = r$.

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 (13), 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.[60]

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.

We integrate the first order condition (13) to obtain the following hedonic model of earnings,

$$\ln y_i = \alpha_i + \rho_i S_i - \frac{1}{2} k_1 S_i^2. \tag{14}$$

To achieve the familiar looking Mincer equation, assume $k_1 = 0$.[61] This assumption rules out diminishing "returns" to schooling in terms of years 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

[60] In that model, aggregate allocations of persons to schooling are determined by an arbitrage condition that returns must be equalized across choices.

[61] The Card model (1999) produces a Mincer-like model where ρ_i is the Mincer return for individual i. The mean return in the population is $E(\rho_i)$. It is an *ex post* return derived under the assumption that log earnings are linear in schooling, contrary to the literature, previously discussed, that shows pronounced nonlinearities and sheepskin effects. (See the discussion in sections 3 and 4.)

r_i is constant for everyone but not necessarily the same constant. 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.[62]

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, $COV(\rho_i, \rho_i - r_i) = 0$.

Under linearity of the estimating equations and exogeneity of the regressors, least squares estimates conditional expectations. Our analysis is more general. 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.

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 $COV(\rho_i, \rho_i - r_i) > 0$. If the costs of schooling are sufficiently positively correlated with the growth rate, then schooling is negatively correlated with the growth rate.

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

[62] Recall our discussion of Section 4. From equation (4) if term 1 is zero, and we assume multiplicative separability (or no experience) then $\rho_i (=$ term 2$) = \tilde{r} +$ term 3 which arises from tuition and psychic costs where $\tilde{r}$ is the opportunity cost of funds.

which one can identify $\bar{\rho}$, the mean growth rate in the population. First we consider the case where the marginal cost of funds, r_i, is observed and then other cases in the following subsections.[63]

7.1 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 in section 8.2. The causal effect may also be of interest in its own right if the goal is to estimate earnings pricing equations. We discuss some simple approaches for identifying causal effects before turning to a more systematic analysis in section 8.

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. 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 scale. Recall that $S_i = \dfrac{(\rho_i - r_i)}{k}$, so that $\rho_i = r_i + kS_i$ and $\bar{\rho} = E(\rho_i) = \bar{r} + kE(S_i)$.

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. Form

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

[63] Our discussion is based in part on Heckman and Vytlacil (1998).

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

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

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).[64] 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.[65] 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.

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 $r_i \perp\!\!\!\perp (\alpha_i, \rho_i)$. 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 $L_i \perp\!\!\!\perp (M_i, \rho_i, \alpha_i)$.[66] Thus, we can identify the mean return to schooling.

7.2 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 the analysis of section 7.1 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, $\ln y_i = \alpha_i + \rho_i(\rho_i - L_i\gamma - M_i)/k$. We obtain the vector moment equations:

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

[64] The symbol $\not\perp\!\!\!\perp$ means "not independent of."

[65] See, *e.g.* Greene (2003).

[66] As we have stressed, the independence conditions are overly strong, but can be weakened to a mean independence assumption provided that we only seek to recover conditional means.

so $\overline{\rho}$ is identified from the population moments because the covariances on both sides are available.[67] 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

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

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.

7.3 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,

$$E(\ln y_i \mid S_i) = E(\alpha_i \mid S_i) + E(\rho_i S_i \mid S_i) = E(\alpha_i \mid S_i) + E(\rho_i \mid S_i)S_i.$$

If there is an L 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 as in section 7.2,

$$\begin{aligned} E\left(\ln y_i \mid L_i\right) &= E(\alpha_i \mid L_i) + E(\rho_i S_i \mid L_i) \\ &= \eta + \bar{\rho}E(S_i \mid L_i).^{68} \end{aligned}$$

[67] One can use the GMM formula presented in Hansen (1982) to construct an efficient estimator if there is more than one non-constant element in L_i.

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. 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 in section 10 and use it to motivate recent developments in the literature on identifying information available to agents when they make their schooling decisions.

7.4 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.

8 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 (11) in which there are only two levels of schooling. Our discussion can be generalized (see Heckman and Vytlacil, 2005b and Heckman, Urzua, and Vytlacil, 2004), 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 (14) varies in the population. Recent research on instrumental variables in the correlated coefficient model establishes a close link between IV and the selection model (Heckman, 1976) that Willis

[68] $\eta = \bar{\alpha} + \left(\frac{\sigma_\rho^2}{k}\right) - \frac{E(\rho_i M_i)}{k}$, where $\sigma_\rho^2 = VAR(\rho_i)$.

and Rosen apply to obtain their estimates. As shown in Heckman, Urzua, and Vytlacil (2004), 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.

Since 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.[69] Heckman (1997) and Heckman and Vytlacil (1998) show how models of schooling that capture key features of economic theory are intrinsically nonlinear. See also the discussion in Heckman and Vytlacil (1998).

In the model of section 7, 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.

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.

Following Heckman and Vytlacil (2000, 2005a,b), Heckman, Urzua, and Vytlacil (2004), 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.[70]

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

[69] Card's model becomes nonlinear if one constrains it to rule out negative schooling and schooling above observed magnitudes.

[70] Heckman, Urzua, and Vytlacil (2004) systematically compare these literatures. See also Heckman and Vytlacil (2005a,b).

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. Attention has focused in recent years on IV estimates of the coefficient of schooling in a regression of log earnings on schooling. In the special cases analyzed in section 7, 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.

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. 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.

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. 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 ($E(G|G \leq C)$ *vs.* $E(G|C' < G \leq C)$). Neither is estimated by the average growth rate, and hence by the IV method discussed in section 7. 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.[71]

[71] There is an additional assumption behind this example which we will maintain throughout this chapter: that the

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. 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 economic questions. 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.[72]

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. This section draws from Heckman and Vytlacil (2005a,b), Heckman, Urzua, and Vytlacil (2004) 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. 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.

policy does not have important general equilibrium effects.

[72] Heckman and Vytlacil (2005b) and Heckman, Urzua, and Vytlacil (2004) present an extensive discussion of this issue.

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, 2005b).

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.

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.[73] Heckman and Robb (1985, 1986) make the point that IV does not identify interpretable parameters in a selection model or a generalized Roy model.

8.1 The Generalized Roy Model of Schooling

To focus the discussion, and motivate the empirical literature, we consider a two outcome model. Heckman and Vytlacil (2005a,b) and Heckman, Urzua, and Vytlacil (2004) extend this discussion to ordered choice and general unordered choice models with multiple outcomes.

As noted in section 7, 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.[74] 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. In the generalized Roy framework, the potential outcomes associated with two different schooling levels are generated by

[73] As opposed to what we would expect from a single skill model where the best individuals in one sector would also be the best individuals in the sector they did not choose if they were placed there instead.

[74] See Becker and Chiswick (1966), Chiswick (1974) and Mincer (1974).

two random variables (U_0, U_1):

$$\ln Y_0 = \alpha + U_0 \tag{15a}$$

$$\ln Y_1 = \alpha + \bar{\beta} + U_1 \tag{15b}$$

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. 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.

Observed earnings are written in a "switching regression" form,

$$\ln Y = S \ln Y_1 + (1 - S) \ln Y_0 = \alpha + \beta S + U_0 = \alpha + \bar{\beta} S + \{U_0 + S(U_1 - U_0)\}. \tag{16}$$

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 switching regression model (1958, 1972). It is equivalent to the familiar semilog specification of the earnings-schooling equation popularized by Mincer (1974), given in equation (11), which in the current notation writes log earnings $\ln Y$ as a function of S,

$$\ln Y = \alpha + \bar{\beta} S + U, \tag{17}$$

where $U = U_0 + S(U_1 - U_0)$.[75] In terms of the notation of section 7, $U_0 = \varepsilon_\alpha$, $U_1 - U_0 = \varepsilon_\rho$.

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

[75] For simplicity, throughout this section we suppress explicit notation for dependence of the parameters on the covariates X unless it is clarifying to make this dependence explicit.

form

$$S = \begin{cases} 1 \text{ if } \ln Y_1 - \ln Y_0 \geq C \Longleftrightarrow \beta \geq C \\ 0 \text{ otherwise.} \end{cases} \tag{18}$$

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 (16). This produces the "correlated random coefficient model" that is often applied to general versions of (16). Decision rules similar to (18) characterize many other economic choices.

The conventional approach to estimating selection models postulates normality of (U_0, U_1) in equations (15a) and (15b), 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 (18). 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.[76] However in recent years these assumptions have often been viewed as unacceptably strong by empirical labor economists (See, *e.g.* Angrist and Krueger, 1999).[77]

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, 2005a,b) 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. 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 (2004) and Heckman and Vytlacil (2005b) present extensive discussions of the relationship between the two approaches.

[76] Willis and Rosen (1979) is an example of the application of the generalized Roy model. Textbook treatments of the normal selection model are available in Amemiya (1985) and Ruud (2000). Aakvik, Heckman, and Vytlacil (2005) and Heckman, Tobias, and Vytlacil (2001, 2003) derive all of the treatment parameters and distributions of treatment parameters for several parametric models including the normal. The Heckman, Tobias and Vytlacil papers present one elementary discussion of the normal selection model applied to the generalized Roy framework. Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005d) estimate the distribution of treatment effects under semiparametric assumptions. We review this work in the last two sections of this chapter.

[77] A large literature, starting with Heckman and Sedlacek (1985) and exemplified most recently by Blundell, Reed, and Stoker (2003) shows that correcting for selection and sectoral choices, a log normality assumption for sectoral earnings fits the data well.

Heckman and Vytlacil work with general nonseparable models,

$$\ln Y_1 = \mu_1(X, U_1) \text{ and } \ln Y_0 = \mu_0(X, U_0). \tag{19}$$

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. As demonstrated by Heckman and Vytlacil (2000, 2005a,b) 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.[78]

A latent variable model that captures decision rule (18) in a general way is:

$$\begin{aligned} S^* &= \mu_S(Z) - U_S \\ S &= 1 \text{ if } S^* \geq 0. \end{aligned} \tag{20}$$

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 $S = 0$. 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 separability between Z and U_S in (20) 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 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 (2005a,b) and Heckman, Urzua, and Vytlacil (2004).

The separability that is required to justify (20) and that underlies the entire *LATE*-based liter-

[78] Heckman, Lochner, and Taber (1998) analyze generalized Roy models of schooling in a general equilibrium framework.

ature cannot be justified in many choice-theoretic models of schooling including dynamic discrete models (See Heckman and Navarro, 2005). The method of IV applied to a heterogeneous outcome model is fundamentally asymmetric. 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 "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 (2005a,b) and Heckman, Urzua, and Vytlacil (2004).

Heckman and Vytlacil (2001a, 2005a) assume that (a) Z has some variables that shift $\mu_S(Z)$ given X (the other variables) – an exclusion condition that is standard in the IV literature; (b) The unobservables (U_0, U_1, U_S) are independent of Z given X (a standard instrumental variables condition) and (c) $0 < \Pr(S = 1 \mid X) < 1$, 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. 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.[79]

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 $\mu_S(z) = P(z)$.[80] (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 $U_S = F_{Vs}(V_S)$.) The propensity score $P(z)$ is a monotonic transformation of the mean utility of attending school and we will refer to it

[79] Vytlacil (2002) shows that under the conditions stated in this paragraph, separability (20) implies and is implied by the monotonicity and independence conditions of Imbens and Angrist (1994) and Angrist and Imbens (1995). Heckman and Vytlacil (2005a,b) present three alternative definitions of the MTE for a general nonlinear model which are equivalent in a linear model.

[80] We say a random variable is uniformly distributed over [0,1] if its density is identically equal to 1 over this interval.

as the mean utility.

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.

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

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

the return to the population average person given characteristics $X = x$. This quantity is sometimes called "the" causal effect of S.[81] Others report the "return" for those who attend school:

$$\begin{aligned} E\left(\beta \mid S=1, X=x\right) &= E(\ln Y_1 - \ln Y_0 \mid S=1, X=x) \\ &= \bar{\beta}(x) + E(U_1 - U_0 \mid S=1, X=x).^{82} \end{aligned}$$

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.

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

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

Angrist and Krueger (1991) and Meghir and Palme (2001) estimate this parameter. 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 $E(\ln Y_1 - \ln Y_0 \mid \ln Y_1 - \ln Y_0 = 0) = 0$.

Depending on the conditioning sets and the summary statistics desired, a variety of "causal

[81] It is the Average Treatment Effect (ATE) parameter. Card (1999, 2001) defines it as the "true causal effect" of education. See also Angrist and Krueger (1999, 2001). Our chapter demonstrates that there is no "true causal effect."

[82] It is the Treatment on the Treated parameter as discussed by Heckman and Robb (1985).

effects" can be defined. Different causal effects answer different economic questions. As noted by Heckman and Robb (1986), Heckman (1997) and Heckman and Vytlacil (2005a,b), under one of two conditions

I: $U_1 = U_0$ (common effect model)

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). 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.[83]

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, it is far clearer to state an economic question and find the answer to it. This obvious and traditional approach is not pursued in the literature. Heckman and Vytlacil (2001c, 2005a,b) develop this approach using a standard welfare framework. 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:

$$E(V(Y) \mid \text{Alternative Policy}) - E(V(Y) \mid \text{Baseline Policy}). \tag{21}$$

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

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

[83] For example, Heckman and Robb (1985) note that in his survey of the union effects on wages, Lewis (1986) confuses these different "effects." This is especially important in his comparison of cross section and longitudinal estimates where he inappropriately compares conceptually different parameters.

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. 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 (21) or answers the stated economic question. Heckman and Vytlacil (2001c, 2005a,b) show how (21) can be represented as a weighted average of the *MTE*. The weights are given in Table 9b. See Heckman, Urzua, and Vytlacil (2004) for further examples.

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. 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 (1994) is often invoked by empirical analysts to justify an instrumental variable estimate, without providing any precise definition of the economic question it addresses.

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.

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.[84]

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[85] 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. 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.[86] 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.[87]

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

The index model (18) and (20) 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 $U_S = u_S$.[88] 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 $P(Z) = u_S$.

Under their assumptions, Heckman and Vytlacil (1999, 2001b, 2005a,b) establish that all of the

[84] See Carneiro, Heckman, and Vytlacil (2005).

[85] An estimand is the probability limit of an estimator.

[86] See Heckman, Urzua, and Vytlacil (2004) for an analysis of this case.

[87] Heckman and Vytlacil (2005b) show that for every policy it is possible in principle to define an instrumental variable that generates the correct policy relevant treatment effect. However, such an instrument may not be feasible in any given data set because of support problems. (Support is the range of a random variable where it has positive density.) Different policies define different policy relevant instrumental variables.

[88] Björklund and Moffitt (1987) introduced this parameter in the context of the parametric normal Roy model. See Heckman and Vytlacil (2005a,b) for a discussion of this literature.

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 or *ATE*, $E(Y_1 - Y_0 \mid S = 1, x)$, and treatment on the untreated or *TUT*, $E(Y_1 - Y_0 \mid S = 0, x)$. See Table 9a (from Heckman and Vytlacil, 2000, 2005a,b) for the treatment parameters expressed in terms of *MTE* and Table 9b for the weights. The analysis of Heckman and Vytlacil (2001b, 2005a,b) unites the selection literature and the modern *IV* literature using a common analytical framework. Heckman, Urzua, and Vytlacil (2004) discuss how to construct the weights.[89]

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.

Figure 10a plots the marginal treatment effect (*MTE*) derived from a generalized normal Roy model using the parameterization of (17) and (18) shown at the base of the figure. 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.[90] 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.

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).[91] *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. The decline in *MTE* reveals that the "gross return" (β) declines with U_S. Those more likely to attend school

[89] Their paper provides software for doing so.

[90] Recall that $S = \mathbf{1}(S^* > 0) = \mathbf{1}(\mu_S(Z) > U_S)$ so that the higher U_S, the less likely is a person to attend college or have $S = 1$.

[91] The form of the Roy model we use assumes additive separability and generates U_0,U_1 and U_S from a common unobservable ε. Thus, in this example, the distribution of $U_1 - U_0$ given U_S is degenerate.

(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 ($E(U_0 \mid X = x, S = 1) - E(U_0 \mid X = x, S = 0) < 0$). 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, 2004, for further analysis and examples). 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.

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 of section 4 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. (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.) College educated persons retire at age T_1. High school educated persons retire at age T_0. 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.

As discussed in section 3, 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.[92] Mincer further assumes that $T_1 - T_0 = \tau$ so working lives are the same for both schooling classes. 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

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

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

$$\gamma = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{Y}_0} \approx \ln \bar{Y}_1 - \ln \bar{Y}_0.$$

This is the "Mincer return" to school. An alternative expression for the return is

$$R = \frac{\gamma - \frac{CA(\tau)}{\bar{Y}_0 e}}{1 + \frac{CA(\tau)}{\bar{Y}_0 e}}.$$

This shows that the Mincer return γ, is greater than the true return, R, whenever costs are positive. When costs are zero ($C = 0$), R equals the Mincer return, γ. 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

[92] Mincer assumes more general period-specific growth rates. The argument in the text can be modified to account for this at the cost of more notational complexity.

to college; otherwise, it does not. When $C = 0$, an alternative way to state this criterion is that it pays to go to college if

$$1 + \gamma > (1 + r)^{\tau}.$$

When $\tau = 1$, this simplifies to the conventional criterion that $\gamma > r$.

The evidence presented in sections 3 and 4 of this chapter 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. Additionally, the evidence presented below in section 10 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 γ is an ingredient of returns, it is not in general a return, as the expression for R reveals.[93]

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.

8.3 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 $\Pr(S = 1 \mid X, \beta) \neq \Pr(S = 1 \mid X)$. In this framework, β is an *ex post* "causal effect," which may not be known to agents *ex ante*. In

[93] In addition, in this two choice example, there are no option values. Accounting for them is a factor toward raising the rate of return above the measured growth rate.

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.

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).

In the notation of equation (17), the usual assumption in the literature is that $COV(S, U_0) > 0$. 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:

$$\text{plim } \hat{\beta}_{OLS} = \beta + \frac{COV(S, U_0)}{V(S)} > \beta.$$

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 $COV(Z, S) \neq 0$ and $COV(Z, U_0) = 0$, then:

$$\text{plim } \hat{\beta}_{IV} = \beta + \frac{COV(Z, U_0)}{COV(Z, S)} = \beta.$$

Therefore we expect that $\hat{\beta}_{IV} < \hat{\beta}_{OLS}$.

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.

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.[94] Card (1999, 2001)

[94] Nonclassical measurement error is measurement error of a true variable that is stochastically dependent on the true value of the variable. Thus the mean, the variance and other moments of the measurement error may depend on the true value.

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.

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. 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;[95] 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.[96] In our notation, the probability limits of the least squares and *IV* estimators are

$$\text{plim } \hat{\beta}_{OLS} = \bar{\beta} + \frac{COV(S, U_0)}{V(S)} + \frac{COV[S, S(U_1 - U_0)]}{V(S)} \tag{22}$$

$$\begin{aligned} \text{plim } \hat{\beta}_{IV} &= \frac{COV[Z, S(U_1 - U_0)]}{COV(Z, S)} + \bar{\beta} + \frac{COV(Z, U_0)}{COV(Z, S)} \\ &= \bar{\beta} + \frac{COV[Z, S(U_1 - U_0)]}{COV(Z, S)}. \end{aligned} \tag{23}$$

In general, plim $\hat{\beta}_{OLS}$ can be larger than, smaller than or equal to plim $\hat{\beta}_{IV}$. We can rewrite (22)

[95] This argument is based on *LATE* (Imbens and Angrist, 1994). Card does not provide a precise definition of the concept. Carneiro, Heckman, and Vytlacil (2005) precisely define and estimate the return for the average marginal person. See also the discussion in Heckman and Vytlacil (2005b).

[96] Card's model was described in section 7. It allows for multiple levels of schooling, but assumes a common rate of return across all schooling levels or else reports an average return to schooling across schooling levels. Heckman, Urzua, and Vytlacil (2004) and Heckman and Vytlacil (2005b) develop methods for identifying marginal returns to different schooling levels. For simplicity, we assume a two outcome schooling model.

and (23) as:

$$\begin{aligned}
\text{plim } \hat{\beta}_{OLS} &= \bar{\beta} + E(U_0|S=1) - E(U_0|S=0) + E(U_1 - U_0|S=1) \\
&= E(\beta|S=1) + E(U_0|S=1) - E(U_0|S=0) \\
\text{plim } \hat{\beta}_{IV} &= \bar{\beta} + E(U_1 - U_0|S=1) \\
&\quad + \frac{COV[Z, (U_1 - U_0)|S=1] \Pr(S=1)}{COV(Z,S)} \\
&= E(\beta|S=1) + \frac{COV[Z, (U_1 - U_0)|S=1] \Pr(S=1)}{COV(Z,S)}.
\end{aligned}$$

Therefore, plim $\hat{\beta}_{IV}$ >plim $\hat{\beta}_{OLS}$ if $\frac{COV[Z,(U_1-U_0)|S=1]\Pr(S=1)}{COV(Z,S)} > E(U_0|S=1) - E(U_0|S=0)$.[97]

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 $\frac{COV[Z,(U_1-U_0)|S=1]\Pr(S=1)}{COV(Z,S)} > 0$.

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.[98] As a consequence of these two conditions,

$$COV(Z,S) < 0 \text{ and } COV(Z,U_1) = COV(Z,U_0) = 0. \tag{24}$$

In the simple two outcome model of schooling, individuals enroll in school if benefits are higher than costs as is clear from equation (18) ($S=1$ 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 $COV(Z, U_1 - U_0) = 0$, $COV(Z, U_1 - U_0|S=1) > 0$ (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 $U_1 - U_0$). But in that case, because $COV(Z,S) < 0$, plim $\hat{\beta}_{IV}$ < plim $\hat{\beta}_{OLS}$. Implicit in Card's analysis is the assumption that it is not possible for the average

[97] This argument appears in Carneiro and Heckman (2002).

[98] Z may be a vector, but in this example we assume it is a scalar.

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 $E(\beta|S=1) < E(\beta|\beta = C(Z))$. 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 ($k_1 < 0$ in section 7), 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.

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, $\hat{\beta}_{IV} > \hat{\beta}_{OLS}$. In their setup, $\frac{COV[Z,(U_1-U_0)|S=1]\Pr(S=1)}{COV(Z,S)} < 0$, $E(U_0|S=1) - E(U_0|S=0) < 0$ and $\frac{COV[Z,(U_1-U_0)|S=1]\Pr(S=1)}{COV(Z,S)} > E(U_0|S=1) - E(U_0|S=0)$.[99] *OLS* estimates are downward biased for $E(\beta|S=1)$ because $E(U_0|S=1) - E(U_0|S=0) < 0$. 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.[100] This possibility is featured in Willis and Rosen (1979), who speculate that, contrary to conventional wisdom, $COV(U_1, U_0) < 0$, although, with their model, they cannot identify this correlation from the data. Carneiro, Heckman, and Vytlacil (2005) and Cunha, Heckman, and Navarro (2005d) 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 (22)), and since $E(Y_0|S=0) > E(Y_0|S=1)$, the *OLS* estimate is an underestimate of $E(Y_1 - Y_0|S=1)$.

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

[99] Carneiro (2002) shows that their conclusions hold for white females and across different data sets.

[100] In such a model we need to have more than one dimension of ability.

in school.[101] 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). 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.[102]

8.4 Estimating the *MTE*

We now show how the local *IV* methods of Heckman and Vytlacil (1999, 2001b, 2005a,b) can be used to estimate average returns to school for any population of interest. Heckman, Urzua, and Vytlacil (2004) show how to estimate the *MTE* and generate all of the weights shown in Table 9b. They also provide software for doing so. Using equation (16) the conditional expectation of $\log Y$ $(= \ln Y_0(1-S) + \ln Y_1 S)$ is

$$E(\ln Y \mid Z = z) = E(\ln Y_0 \mid Z = z) + E(\ln Y_1 - \ln Y_0 \mid Z = z, S = 1) \Pr(S = 1 \mid Z = z)$$

where we keep the conditioning on X implicit. From the index structure generated by decision rules (18) and (20), we may write this expectation as

$$E(\ln Y \mid Z = z) = E(\ln Y_0) + E(\beta \mid P(z) \geq U_S, P(Z) = P(z))P(z).$$

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

[101] However, even in the one ability model, Griliches (1977) shows that it is possible that the most able do not enroll in school because their opportunity costs of doing so are too high.

[102] For example, returns to job training or unionism vary across individuals and individuals make choices based on them. The productivity of different inputs varies across firms and they choose different quantities of inputs according to the productivity patterns they face (this is relevant for the estimation of production functions). Different consumers have different demand elasticities for a good and their choice of quantities depends on their elasticity.

values of Z, z and z', assuming $P(z) < P(z')$, we obtain the IV formula:

$$\begin{aligned}
&\frac{E(\ln Y \mid P(Z) = P(z)) - E(\ln Y \mid P(Z) = P(z'))}{P(z) - P(z')} \\
= \;& \bar{\beta} + \frac{E(U_1 - U_0 \mid P(z) \geq U_S)P(z) - E(U_1 - U_0 \mid P(z') \geq U_S)P(z')}{P(z) - P(z')} \\
= \;& E\left(\beta \mid P\left(z\right) < U_S \leq P\left(z'\right)\right) \\
= \;& \Delta^{LATE}(P(z), P(z')),
\end{aligned}$$

where Δ^{LATE} is the $LATE$ parameter. This is the average return to schooling for individuals who have U_S between $P\left(z\right)$ and $P\left(z'\right)$ $(P\left(z\right) < U_S \leq P\left(z'\right))$. 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. Otherwise IV estimates an economically difficult-to-interpret combination of MTE parameters with weights given in Table 9b.

Another representation of $E(\ln Y \mid P(Z) = P(z))$ reveals the index structure underlying this model more explicitly and writes

$$E(\ln Y \mid P(Z) = P(z)) = \alpha + \bar{\beta} P(z) + \int_{-\infty}^{\infty} \int_{0}^{P(z)} (U_1 - U_0) f(U_1 - U_0 \mid U_S = u_S)\, du_S\, d(U_1 - U_0). \quad (25)$$

Differentiating with respect to $P(z)$, we obtain MTE:

$$\begin{aligned}
\frac{\partial E(\ln Y \mid P(Z) = P(z))}{\partial P(z)} &= \bar{\beta} + \int_{-\infty}^{\infty} (U_1 - U_0) f(U_1 - U_0 \mid U_S = P(z))\, d(U_1 - U_0) \\
&= \Delta^{MTE}(P(z)).
\end{aligned}$$

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 (2004) elaborate on this point.

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 $E(\ln Y|P(z))$ is linear in $P(z)$, standard instrumental variables methods identify "the" effect of S on $\ln Y$. 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 (2004) implement it and find evidence in support of nonlinearity in the data they analyze.

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, 2005a,b).[103]

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, 2005a,b; Heckman, Urzua, and Vytlacil, 2004).

8.5 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. As noted by Heckman and Vytlacil (2005a) and Heckman, Urzua, and Vytlacil (2004), since the question being addressed by the recent literature is not clearly stated, it is not obvious that IV is better than OLS. The

[103] Software is available in Heckman, Urzua, and Vytlacil (2004).

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.

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. (See Cunha, Heckman, Lochner, and Masterov, 2005). Unless one controls for ability, the quality of the instruments is in question. Many data sets lack direct measures of ability.

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). 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). 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.

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).

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 (2004). 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.

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 (2004). Both papers use a logit model for schooling choice. 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. "Distance to College at 14" and "Local Unemployment Rate at 17" have weak effects. The only strong exclusion is tuition, conditioning on ability.

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,[104] 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. (They trim 5% of the observations in the sample.[105]) 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).

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

[104] Formally, for nonparametric analysis, we need to investigate the support of $P(Z)$ conditional on X. However, the partially linear structure that we will impose below implies that we only need to investigate the marginal support of $P(Z)$.

[105] The importance of trimming in a semiparametric model similar to the one we use is illustrated in Heckman, Ichimura, and Todd (1997).

cannot do. It can produce a number. It cannot produce even a conventional treatment effect, much less a rate of return to schooling in these data. The pattern of support of $P(Z)$ is similar in the Heckman, Urzua, and Vytlacil (2004) study. See Figure 12 taken from their analysis.

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.[106] Carneiro, Heckman, and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2004) 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.

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. Heckman, Urzua, and Vytlacil (2004) report similar findings.

Carneiro, Heckman, and Vytlacil (2005) partition the estimated MTE into two components, one depending on X and the other on u_S,

$$\begin{aligned} MTE(x,u_S) &= E(\ln Y_1 - \ln Y_0|X=x, U_S=u_S) \\ &= \mu_1(X) - \mu_0(X) + E(U_1 - U_0|U_S=u_S). \end{aligned}$$

[106] See Blackburn and Neumark (1993); Bishop (1991); Grogger and Eide (1995); Heckman and Vytlacil (2001a); Murnane, Willett, and Levy (1995); Meghir and Palme (2001).

Figure 13 plots the component of the MTE that depends on U_S but not on X where the confidence interval bands are bootstrapped. $E(U_1 - U_0 \mid U_S = u_S)$ is declining in u_S for values of u_S below 0.7, and then it is flat and if anything it slightly rises.[107] 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).[108] The magnitude of the heterogeneity in returns is substantial: returns can vary from 13% to 40% per year of college.[109]

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.

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).

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)

[107] Notice that the decision rule is $S = 1$ if $P(Z) - U_S \geq 0$ so, for a given Z, individuals with a higher U_S are less likely to go to college.

[108] U_S may be interpreted as the unobservable cost of college.

[109] The bootstrapped confidence intervals are very wide. However, the estimates of each point of the curve are highly correlated which will reduce the imprecision of the implied treatment parameters. For example, Carneiro, Heckman and Vytlacil can reject (at the 10% level) the hypothesis that $MTE(\bar{x}, v = 0.05) = MTE(\bar{x}, v = 0.5)$ (although, they cannot reject that $MTE(\bar{x}, v = 0.5) = MTE(\bar{x}, v = 0.95)$).

support of $P(Z)$, which is close to the full unit interval.[110]

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 $P(Z)$. 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.

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. 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%.[111] 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.

Carneiro, Heckman, and Vytlacil (2005) compare all of these estimated summary measures of

[110] Alternative ways to deal with the problem of limited support are to construct bounds for the parameters or to use a parametric extrapolation outside of the observed support. Bounds on the treatment effects are generally wide even though the support is almost full. Parametric extrapolation outside of the support is potentially sensitive to the choice of extrapolation model. Estimates based on locally adapted extrapolations show much less sensitivity than do estimates based on global approximation schemes. See Carneiro, Heckman, and Vytlacil (2005) for further discussion.

[111] The policy consists of a subsidy of \$500 for individuals who have higher than median ability and for whom both parents have less than a high school degree. The purpose of this simulation is to approximate a means tested tuition subsidy for high ability individuals. The standard error of this estimate is quite high, even though it overweights the MTE where it is more precisely estimated. The reason is that the flexible form we use for the selection equation, although useful for the estimation of the MTE, creates some imprecision in the policy simulation because not all the coefficients on the terms involving tuition are precisely estimated, at least for this policy. It is easier to simulate policies in models where tuition enters the choice equation in a simpler way, although in some cases the standard errors of the MTE become larger because it is important (for the standard errors) to be flexible in the way the instruments affect choices. See Carneiro, Heckman, and Vytlacil (2005) for further details.

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,[112] so that they can compare these estimates with the policy relevant treatment effect. The *OLS* estimate of the return to a year of college is 7.88% while the *IV* estimate is 16.49%.[113] 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.

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 explained in section 8.2 and confirmed in the empirical work of Carneiro, Heckman, and Vytlacil (2005) reported here and in Cunha, Heckman, and Navarro (2005d), 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. 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 (2005d) shows that the single skill or efficiency units representation of the labor market which is implicit in most of the literature is invalid.

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

[112] This is obtained by integrating X with respect to $f_X(x|PRT) = f_X(x|\mu_S(Z) - U_S < 0, \mu_S(Z') - U_S \geq 0)$.

[113] When they compute the *IV* parameter by weighting up the *MTE* using the *IV* weights, they get an estimate of 12.12%, which is close to the *IV* estimate we obtain using the direct method.

should be divided by 3.5 to get the annual returns to college reported in Carneiro, Heckman, and Vytlacil (2005).) The final three columns show the IV based on an estimated MTE using (a) a parametric normal model (third column); (b) a semiparametric polynomial estimation method and (c) 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.

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.

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.

8.6 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 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. 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. 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 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.[114] 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.

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 correlation is reported in the third column of Table 14b.

The instrument used by Carneiro, Heckman, and Vytlacil (2005) is $P(Z)$. Regressing schooling

[114] In a recent paper Stock and Yogo (2002) propose a different test. However they still find that the rule of thumb first proposed in Staiger and Stock (1997) works well in general.

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. 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.

8.7 Summary of the Modern Literature on Instrumental Variables

Heckman and Vytlacil (2001a, 2005a,b) 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. 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. The approach based on the MTE unites the selection and IV literatures. As noted by Heckman, Urzua, and Vytlacil (2004), both methods use $P(Z)$ but one conditions on it (the selection model) while the other (the IV literature) does not.

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, 2005d). 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.

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 (2004) they produce wildly different "effects" of schooling on earnings. 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.

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, 2005a; Heckman, Urzua, and Vytlacil, 2004). The monotonicity assumptions (or index structure assumptions embodied in (18) or (20) 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. 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, 2005). 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.

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, 2005b; Heckman, Urzua, and Vytlacil, 2004, for further discussion of this point).

9 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. In section 8, we presented a variety of mean treatment effects which are defined by the conditioning variables used. Different means answer different policy questions.

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, 2005a,c,d) 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. 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.[115] 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. From knowledge of the mean treatment effects presented in section 8, 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. The methods can be used to explain

[115] It is also possible to generate all mean, median or other quantile gains to schooling, to identify all pairwise treatment effects in a multi-outcome setting, and to determine how much of the variability in returns across persons comes from variability in the distributions of the outcome selected and how much comes from variability in opportunity distributions.

effects of schooling (and other interventions) on earnings, employment and health. In this chapter, we focus on earnings measures.

Under the assumptions of section 8, 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.

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.[116] The standard treatment effect model analyzed in section 8 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.[117]

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".

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.

If the selection problem can be solved and the marginal distributions of Y_1 and Y_0 are identified and obtained, 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.[118]

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 $F(Y_1 \mid S = 1, X, Q) = F(Y_1 \mid X, Q)$.

[116] Thus $(Y_1, Y_0) \perp\!\!\!\perp Z \mid X$ and $\Pr(S = 1 \mid Z, X)$ depends on Z for all X.
[117] See Heckman, Lochner, and Taber (1998) for a treatment of general equilibrium policy evaluation.
[118] See Heckman and Smith (1998) and Heckman, Smith, and Clements (1997).

Matching is a form of nonparametric least squares that 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

$$F(Y_0, Y_1 \mid X) = \int F(Y_0 \mid X, Q) F(Y_1 \mid X, Q) \, d\mu(Q \mid X).$$

$\mu(Q \mid X)$ 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.[119]

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,

$$Y_1 = Y_0 + \Delta \tag{26}$$

where Δ is the difference in means between Y_1 and Y_0 for the selection corrected distribution.[120] This assumes that schooling has the same effect on everyone (with the same X) and that effect is Δ.

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 $F_0 = F_0(y_0 \mid X)$. This assumption generates a deterministic relationship which turns out to be the tight upper bound of the Fréchet bounds.[121] An alternative assumption is that people are perfectly inversely ranked so the best in one distribution is the worst in the other:

[119] See Heckman and Vytlacil (2005a,b)

[120] Δ may be a function of X.

[121] An upper bound is "tight" if it is the smallest possible upper bound. A lower bound is tight if it is the largest lower bound.

$Y_1 = F_1^{-1}\left(1 - F_0(Y_0)\right)$. This is the tight Fréchet lower bound.[122]

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.[123] However, it imposes a strong and arbitrary dependence across distributions. When the perfect ranking assumption is relaxed and tested, it is rejected.[124]

A more general framework attacks this problem in a different way than does matching or 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) and Cunha, Heckman, and Navarro (2005a,c,d). In this chapter we draw on the analysis of Carneiro, Hansen, and Heckman (2003).[125] 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.

The basic idea is to restrict the dependence among the (U_0, U_1, U_S) by factor models or other restrictions (Urzua, 2005). 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).

[122] More generally, one can associate quantiles across distributions more freely. Heckman, Smith, and Clements (1997) use Markov transition kernels that stochastically map quantiles of one distribution into quantiles of another. They define a pair of Markov kernels $M(y_1, y_0 \mid X)$ and $\tilde{M}(y_0, y_1 \mid X)$ such that

$$F_1(y_1 \mid X) = \int M(y_1, y_0 \mid X) dF_0(y_0 \mid X)$$

$$F_0(y_0 \mid X) = \int \tilde{M}(y_0, y_1 \mid X) dF_1(y_1 \mid X).$$

Allowing these kernels to be degenerate produces a variety of deterministic transformations, including the two previously presented, as special cases of a general mapping. Different $(M, \tilde{M})$ pairs produce different joint distributions. These stochastic or deterministic transformations supply the missing information needed to construct the joint distributions.

[123] See *e.g,* Heckman, Smith, and Clements (1997).

[124] However, testing it requires invoking other assumptions. See Cunha, Heckman, and Navarro (2005a,c,d).

[125] Using the model of equations (19) and (20), under the assumption that $(Z, X) \perp\!\!\!\perp (U_0, U_1, U_S)$ and including the condition that (i) $\mu_S(Z)$ is a nontrivial function of Z conditional on X and that (ii) the support assumptions on $\mu_1(X), \mu_0(X)$ and $\mu_S(Z)$ matches that of U_1, U_0, U_S, they establish nonparametric identification of $F(U_0, U_S), F(U_1, U_S)$ up to a scale for U_S, and full identification of $\mu_1(X)$, $\mu_0(X)$ over their supports and $\mu_S(Z)$ suitably scaled.

Assume separability between unobservables and observables and that Y_1 and Y_0 are scalars:

$$\begin{aligned} Y_1 &= \mu_1(X) + U_1 \\ Y_0 &= \mu_0(X) + U_0. \end{aligned}$$

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. 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).

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 $\sigma_S^2 = Var(U_S)$). See Heckman (1976) or Cunha, Heckman, and Navarro (2005d). 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 $(U_0, U_1, U_S/\sigma_S)$.

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 (2005a). 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.[126] Formal proofs and general conditions are given in Carneiro, Hansen,

[126] We identify $F(Y_0 \mid X)$ in the limit sets where $P(Z) = 0$ and $F(Y_1 \mid X)$ in the limit sets where $P(Z) = 1$. Heckman, Urzua, and Vytlacil (2004) present evidence on the validity of such limit arguments in their empirical analysis.

and Heckman (2003). Normality plays no central role in the analysis of this section. We use it because it has a familiarity in the economics of education due to the application of the Generalized Roy model by Willis and Rosen (1979).[127]

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). One 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. 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 $E(\varepsilon_S) = 0$. To set the scale of the unobserved factor, we normalize one "loading" (coefficient on θ) to 1. 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.

Under normality or from the general semiparametric identification analysis of Carneiro, Hansen, and Heckman (2003), we can identify

$$\begin{aligned} COV(U_0, \frac{U_S}{\sigma_S}) &= \frac{\alpha_0\alpha_S}{\sigma_S}\sigma_\theta^2 \\ COV(U_1, \frac{U_S}{\sigma_S}) &= \frac{\alpha_1\alpha_S}{\sigma_S}\sigma_\theta^2 \end{aligned}$$

where $\sigma_S^2 = Var(\varepsilon_S)$. From the ratio of the second covariance to the first we obtain $\frac{\alpha_1}{\alpha_0}$. Thus we obtain the sign of the dependence between U_0, U_1 because

$$COV(U_0, U_1) = \alpha_0\alpha_1\sigma_\theta^2.$$

[127] Separability can also be relaxed using the methods of Matzkin (2003).

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.

Knowledge of the sign of $\frac{\alpha_1}{\alpha_0}$ 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 $COV(U_0, U_1) < 0$. The evidence reported in Carneiro, Hansen, and Heckman (2001, 2003) and Cunha, Heckman, and Navarro (2005a,c,d) supports their conjecture. Those with high levels of U_1 have lower levels of U_0.[128]

With additional information, we can identify the full joint distribution. We now present some examples. Cunha, Heckman, and Navarro (2005d) present a more comprehensive analysis.

Example 1 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). 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 :

$$\begin{aligned} COV(Y_1, T) &= \alpha_1 \alpha_T \sigma_\theta^2, \\ COV(Y_0, T) &= \alpha_0 \alpha_T \sigma_\theta^2, \\ COV(S^*, T) &= \frac{\alpha_S}{\sigma_S} \alpha_T \sigma_\theta^2.^{129} \end{aligned}$$

[128] Our terminology is different from that of Willis and Rosen (1979). What they call a "one factor model" is an efficiency units model where $U_1 = U_0$ (or, more generally, that U_1 and U_0 are perfectly dependent). As our analysis reveals, it is possible to have (U_1, U_0) not perfectly dependent and have a one factor representation. U_1 and U_0 would be perfectly dependent in a one factor model only when the uniquenesses are identically zero: $\varepsilon_0 = \varepsilon_1 = 0$.

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 ($\alpha_T = 1$). 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

$$\frac{COV\left(Y_1, S^*\right)}{COV\left(S^*, T\right)} = \frac{\alpha_1 \alpha_S \sigma_\theta^2}{\alpha_S \alpha_T \sigma_\theta^2} = \alpha_1,$$

because $\alpha_T = 1$ (normalization). From the preceding argument without the test score, we obtain α_0 since

$$\frac{COV\left(Y_1, S^*\right)}{COV\left(Y_0, S^*\right)} = \frac{\alpha_1 \alpha_S \sigma_\theta^2}{\alpha_0 \alpha_S \sigma_\theta^2} = \frac{\alpha_1}{\alpha_0}.$$

From knowledge of α_1 and α_0 and the normalization for α_T, we obtain σ_θ^2 from $COV\left(Y_1, T\right)$ or $COV\left(Y_0, T\right)$. We obtain α_S (up to scale σ_S) from $COV\left(S^*, T\right) = \alpha_S \alpha_T \sigma_\theta^2$ 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.

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,

$$\begin{aligned} \alpha_S &= \alpha_1 - \alpha_0 - \alpha_C \\ \varepsilon_S &= \varepsilon_1 - \varepsilon_0 - \varepsilon_C \\ Var\left(\varepsilon_S\right) &= Var\left(\varepsilon_1\right) + Var\left(\varepsilon_0\right) + Var\left(\varepsilon_C\right). \end{aligned}$$

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.

[129] Conditioning on X, Z, we can remove the dependence of Y_1, Y_0, T and S^* on these variables and effectively work with the residuals $Y_0 - \mu_0\left(X\right) = U_0$, $Y_1 - \mu_1\left(X\right) = U_1$, $T - \mu_T\left(X\right) = U_T$, $S^* - \mu_S\left(Z\right) = U_S$.

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).[130] From the variance of T given X, we obtain $Var(\varepsilon_T)$ since we know $Var(T)$ (conditional on X) and we know $\alpha_T^2\sigma_\theta^2$:

$$Var(T) - \alpha_T^2\sigma_\theta^2 = \sigma_{\varepsilon_T}^2.$$

(Recall that we keep the conditioning on X implicit.) By similar reasoning, it is possible to identify $Var(\varepsilon_0)$, $Var(\varepsilon_1)$ and the fraction of $Var(U_S)$ 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.

We have assumed normality because it is convenient. Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005a,c,d) 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.

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, that leads into our analysis of *ex ante vs. ex post* returns is discussed next.

Example 2 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

$$\begin{aligned} Y_{1t} &= \mu_{1t}(X) + \alpha_{1t}\theta + \varepsilon_{1t} \qquad t = 1, 2 \\ Y_{0t} &= \mu_{0t}(X) + \alpha_{0t}\theta + \varepsilon_{0t} \qquad t = 1, 2. \end{aligned}$$

[130] The easiest case to understand writes

$$\mu_C(Z) = Z\gamma \qquad \mu_1(X) = X\beta_1 \qquad \mu_0(X) = X\beta_0$$
$$\mu_S(Z, X) = X(\beta_1 - \beta_0) - Z\gamma.$$

We identify the coefficients of the index $\mu_S(Z, X)$ up to scale σ_{U_S}, but we know $\beta_1 - \beta_0$ from the earnings functions. Thus if one X is not in Z and its associated coefficient is not zero, we can identify σ_{U_S}. See, *e.g.* Heckman (1976).

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. 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.

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 where

$$U_S = \varepsilon_{11} + \varepsilon_{12} - \varepsilon_{01} - \varepsilon_{02} - \varepsilon_C.$$

Following our discussion of Example 1, 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.[131]

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.

From the available data, we can identify the following covariances:

$$\begin{aligned}
COV(U_S, U_{12}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{12}\sigma_\theta^2 \\
COV(U_S, U_{11}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{11}\sigma_\theta^2 \\
COV(U_S, U_{01}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{01}\sigma_\theta^2 \\
COV(U_S, U_{02}) &= (\alpha_{12} + \alpha_{11} - \alpha_{02} - \alpha_{01} - \alpha_C)\alpha_{02}\sigma_\theta^2 \\
COV(U_{11}, U_{12}) &= \alpha_{11}\alpha_{12}\sigma_\theta^2 \\
COV(U_{01}, U_{02}) &= \alpha_{01}\alpha_{02}\sigma_\theta^2.
\end{aligned}$$

[131] If not, then $\mu_C(Z)$, $\sigma^2_{U_S}$ and ε_C are identified up to normalizations.

If we normalize $\alpha_{01} = 1$ (recall that one normalization is needed to set the scale of θ), we can form the ratios

$$\frac{COV(U_S, U_{12})}{COV(U_S, U_{01})} = \alpha_{12} \qquad \frac{COV(U_S, U_{11})}{COV(U_S, U_{01})} = \alpha_{11}$$

$$\frac{COV(U_S, U_{02})}{COV(U_S, U_{01})} = \alpha_{02}.$$

From these coefficients and the remaining covariances, we identify σ_θ^2 using $COV(U_{11}, U_{12})$ and/or $COV(U_{01}, U_{02})$. Thus if the factor loadings are nonzero,

$$\frac{COV(U_{11}, U_{12})}{\alpha_{11}\alpha_{12}} = \sigma_\theta^2$$

and

$$\frac{COV(U_{01}, U_{02})}{\alpha_{01}\alpha_{02}} = \sigma_\theta^2.$$

We can recover σ_θ^2 (since we know $\alpha_{11}\alpha_{12}$ and $\alpha_{01}\alpha_{02}$) from $COV(U_{11}, U_{12})$ and $COV(U_{01}, U_{02})$. 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) $COV(Y_{11}, Y_{01}) = \alpha_{11}\alpha_{01}\sigma_\theta^2$; $COV(Y_{12}, Y_{01}) = \alpha_{12}\alpha_{01}\sigma_\theta^2$; $COV(Y_{11}, Y_{02}) = \alpha_{11}\alpha_{02}\sigma_\theta^2$ and $COV(Y_{12}, Y_{02}) = \alpha_{12}\alpha_{02}\sigma_\theta^2$. 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.

As in Example 1, 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 (2005a,c,d) and Heckman and Navarro (2005) use multiple factors. We offer an example in the next section.

The key idea to constructing joint distributions of counterfactuals using the analysis of Cunha, Heckman, and Navarro (2005a,c,d) 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. 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.

10 *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 we can measure realized outcomes and assess what components of those outcomes were known at the time schooling choices were made.[132]

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:

$$Y_{i,t} = X_{i,t}\beta + S_i\tau + U_{i,t}, \tag{27}$$

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.

The error term $U_{i,t}$ is usually decomposed into two or more components. For example, it is

[132] An alternative approach summarized by Manski (2004) is to use survey methods to elicit expectations. We do not survey that literature in this chapter.

common to specify that

$$U_{i,t} = \phi_i + \delta_{i,t}. \tag{28}$$

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. 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).

An alternative specification of the error process postulates a factor structure for earnings, that uses the representation introduced in section 9:

$$U_{i,t} = \theta_i\alpha_t + \varepsilon_{i,t}, \tag{29}$$

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. Hause (1980) and Heckman and Scheinkman (1987) analyze such earnings models. Any process in the form of equation (28) can be written in terms of (29). The latter specification is more directly interpretable as a pricing equation than (28).

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 (27), (28), and (29) or versions of them describe *ex post* variability but tell us nothing about which components of (27) or (29) 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 (29), 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.

The methodology developed in Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005b,c,d) 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. 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.

For certain 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) and Cunha, Heckman, and Navarro (2005b,c,d) use choice information to extract *ex ante* or forecast components of earnings and to distinguish them from realized earnings under different market environments. 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.

10.1 A Generalized Roy Model

To state these issues more precisely, consider a version of the generalized Roy (1951) economy with two sectors.[133] This builds on the second example of section 9. 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. 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.

$Y_{1,i}$ is the *ex post* present value of earnings in the college sector, discounted over horizon T for

[133] See Heckman (1990) and Heckman and Smith (1998) for discussions of the generalized Roy model. In this chapter we assume only two schooling levels for expositional simplicity, although our methods apply more generally.

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. $Y_{1,i}$ 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 in section 9 and the references cited.

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. 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

$$S_i = \begin{cases} 1, \text{ if } E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right) \geq 0 \\ 0, \text{ otherwise.}^{134} \end{cases} \tag{30}$$

Under perfect foresight, the postulated information set would include $Y_{1,i}, Y_{0,i}$, and C_i. 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.

The decision rule is more complicated in the absence of full risk diversifiability and depends

[134] If there are aggregate sources of risk, full insurance would require a linear utility function.

on the curvature of utility functions, the availability of markets to spread risk, and possibilities for storage. (See Cunha, Heckman, and Navarro, 2005c, and Navarro, 2004, 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, Heckman, and Navarro, 2005b).

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 their analysis, we consider how this question might be addressed in the linear-in-the-parameters Card model.

10.2 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 discounted lifetime earnings of person i as

$$Y_i = \alpha + \rho_i S_i + U_i, \tag{31}$$

where ρ_i is the person-specific *ex post* return, S_i is years of schooling, and U_i is a mean zero unobservable. 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, $S_i = \lambda(\eta_i, Z_i, \tau_i)$, 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.

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. 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.

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 in section 8 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. 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?

Suppose that the model for schooling can be written in linear in parameters form, as in the Card model exposited in section 7:

$$S_i = \lambda_0 + \lambda_1 \eta_i + \lambda_2 \nu_i + \lambda_3 Z_i + \tau_i, \tag{32}$$

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 also be correlated with U_i and η_i and ν_i in (31). 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.

As a simple example, consider the model of section 7.1. We observe the cost of funds, r_i, and assume $r_i \perp\!\!\!\perp (\rho_i, \alpha_i)$. This assumes that the costs of schooling are independent of the "return" ρ_i and the payment to raw ability, α_i. 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_i).

Suppose that agents do not know ρ_i at the time they make their schooling decisions but instead know $E(\rho_i) = \bar{\rho}$.[135] If agents act on this expected return to schooling, decisions are given by

$$S_i = \frac{\bar{\rho} - r_i}{k}$$

and *ex post* earnings observed after schooling are

$$Y_i = \bar{\alpha} + \bar{\rho} S_i + \{(\alpha_i - \bar{\alpha}) + (\rho_i - \bar{\rho}) S_i\}.$$

In the notation introduced in the Card model, $\eta_i = \bar{\rho}$ and $\nu_i = \rho_i - \bar{\rho}$.

In this case,

$$COV(Y, S) = \bar{\rho} Var(S)$$

because $(\rho_i - \bar{\rho})$ is independent of S_i. Note further that $(\bar{\alpha}, \bar{\rho})$ can be identified by least squares because $S_i \perp\!\!\!\perp [(\alpha_i - \bar{\alpha}), (\rho_i - \bar{\rho}) S_i]$.

If, on the other hand, agents know ρ_i at the time they make their schooling decisions, *OLS* breaks down for identifying $\bar{\rho}$ because ρ_i is correlated with S_i. We can identify $\bar{\rho}$ and the distribution of ρ_i using the method of instrumental variables presented in section 7.1. Under our assumptions, r_i is a valid instrument for S_i.

In this case

$$COV(Y_i, S) = \bar{\rho} Var(S) + COV(S, (\rho - \bar{\rho}) S).$$

Since we observe S, 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 $COV(\ln y, S)$. 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_i is not known at the

[135] This is a rational expectations assumption. Under rational expectations with the currently specified information set, the mean *ex ante* return is the same as the mean *ex post* return, but the distributions of these returns may be very different.

time schooling decisions are made and that agents base their choice of schooling on $E(\rho_i) = \bar{\rho}$. If the economist does not observe r_i, but instead observes determinants L_i satisfying the conditions in section 7.2, then we can still conduct the Durbin-Wu-Hausman test to discriminate between the two hypotheses, but we cannot form $COV(\rho, S)$ directly.

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 in section 7.3 but OLS is no longer consistent even if, in making their schooling decisions, agents forecast ρ_i 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.

Even ignoring selection bias, if we misspecify the information set, in the case where r_i 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 $(\rho_i - \bar{\rho})$ and are correlated with S_i, we do not identify $\bar{\rho}$. The Durbin-Wu-Hausman test is not informative on the stated question. For example, if local labor market variables proxy the opportunity cost of school (the r_i), and also predict the evolution of *ex post* earnings $(\rho_i - \bar{\rho})$, they are invalid. The question of determining the appropriate information set is front and center and cannot in general be inferred using IV methods.

The method developed by Cunha, Heckman, and Navarro (2005a,c,d) 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 discussed in section 8.

10.3 Identifying Information Sets

Cunha, Heckman, and Navarro (2005a,c,d) 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 (2005) extend their analysis to multiple schooling levels.

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 $E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right)$. Under the correct model of expectations, we could form the residual

$$V_{\mathcal{I}_{i,0}} = \left(Y_{1,i} - Y_{0,i} - C_i\right) - E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right),$$

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 (31) and (32), 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 $V_{\mathcal{I}_{i,0}} = \nu_i$.

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 $V_{\widetilde{\mathcal{I}}_{i,0}} = \left(Y_{1,i} - Y_{0,i} - C_i\right) - E\left(Y_{1,i} - Y_{0,i} - C_i \mid \widetilde{\mathcal{I}}_{i,0}\right)$. 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 (31) and (32), 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.

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.[136] There may be several information sets that satisfy this property.[137] 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.

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 $\mathcal{I}_{i,0}$

[136] This procedure is a Sims (1972) version of a Wiener-Granger causality test.

[137] Thus different combinations of variables may contain the same information. The issue of the existence of a smallest information set is a technical one concerning a minimum $\sigma-$algebra that satisfies the condition on $\mathcal{I}_{i,0}$.

or their distribution. Thus, in the example of equations (31) and (32) 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. This strategy is pursued in Cunha, Heckman, and Navarro (2005a,d) for a two-choice model of schooling, and generalized by Cunha, Heckman, and Navarro (2005c).

10.4 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 Example 2 presented in section 9. 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. 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 $COV(U_S, U_{12}) = 0$, $COV(U_S, U_{01}) = 0, COV(U_S, U_{02}) = 0$. However realized earnings are correlated with each other through the realized θ.

Under the assumptions of Example 2, 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.

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.

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. 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:

$$C_i = Z_i\gamma + \theta_i\alpha_C + \varepsilon_{i,C}.$$

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 in section 9 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.

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:

$$\begin{aligned} S_i^* &= E\left(\sum_{t=0}^{T} \frac{\left(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}\right) - \left(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t}\right)}{(1+r)^t} - (Z_i\gamma + \theta_i\alpha_C + \varepsilon_{iC}) \;\middle|\; \mathcal{I}_{i,0}\right) \\ S_i &= 1 \text{ if } S^* \geq 0;\ S_i = 0 \text{ otherwise.} \end{aligned} \tag{33}$$

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.

Constructing (33) 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. 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.

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.[138]

To formally characterize the CHN empirical procedure, it is useful to introduce some additional notation. Let $\odot$ denote the Hadamard product ($a \odot b = (\mathrm{a}_1 b_1, \ldots, \mathrm{a}_L b_L)$) 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. For a proposed information set $\widetilde{\mathcal{I}}_{i,0}$ which may or may not be the true information set on which agents

[138] The term 'heterogeneity' is somewhat unfortunate. Under this term, CHN include trends common across all people (*e.g.*, macrotrends). The real distinction they are making is between components of realized earnings forecastable by agents at the time they make their schooling choices *vs.* components that are not forecastable.

act, CHN define the proposed choice index $\tilde{S}_i^*$ in the following way:

$$\begin{aligned}\tilde{S}_i^* &= \sum_{t=0}^{T} \frac{E\left(X_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t}\left(\beta_{1,t}-\beta_{0,t}\right) + \sum_{t=0}^{T} \frac{\left[X_{i,t} - E\left(X_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t}\left(\beta_{1,t}-\beta_{0,t}\right) \odot \Delta_{X_t} \qquad (34)\\ &+E(\theta_i \mid \widetilde{\mathcal{I}}_{i,0})\left[\sum_{t=0}^{T}\frac{(\alpha_{1,t}-\alpha_{0,t})}{(1+r)^t} - \alpha_C\right] + \left[\theta_i - E\left(\theta_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]\left\{\left[\sum_{t=0}^{T}\frac{(\alpha_{1,t}-\alpha_{0,t})}{(1+r)^t} - \alpha_C\right] \odot \Delta_\theta\right\}\\ &+\sum_{t=0}^{T}\frac{E\left(\varepsilon_{1,i,t}-\varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} + \sum_{t=0}^{T}\frac{\left[(\varepsilon_{1,i,t}-\varepsilon_{0,i,t}) - E\left(\varepsilon_{1,i,t}-\varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t}\Delta_{\varepsilon_t}\\ &-E\left(Z_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\gamma - \left[Z_i - E\left(Z_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]\gamma \odot \Delta_Z - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right) - \left[\varepsilon_{iC} - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]\Delta_{\varepsilon_C}.\end{aligned}$$

To conduct their test, CHN fit a schooling choice model based on the proposed model (34). They estimate the parameters of the model including the Δ parameters. This decomposition for $\tilde{S}_i^*$ assumes that agents know the β, the γ, and the α.[139] If it 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 $\widetilde{S}_i^* = S_i^*$.

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 the example coefficients based on Example 2 of section 9, previously discussed, 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. 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 $\mathcal{I}_{i,0}$.

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.

[139] Cunha, Heckman, and Navarro (2005b,c,d) relax this assumption.

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.

One plausible scenario 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 (34). 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.[140,141]

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.[142] 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.

Statistical decompositions do not tell us which components of (29) 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

[140] This test has been extended to a nonlinear setting, allowing for credit constraints, preferences for risk, and the like. See Cunha, Heckman, and Navarro (2005b) and Navarro (2004).

[141] A similar but distinct idea motivates the Flavin (1981) test of the permanent income hypothesis and her measurement of unforecastable income innovations. She picks a particular information set $\widetilde{\mathcal{I}}_{i,0}$ (permanent income constructed from an assumed ARMA (p, q) time series process for income, where she estimates the coefficients given a specified order of the AR and MA components) and tests if $V_{\widetilde{\mathcal{I}}_{i,0}}$ (our notation) predicts consumption. Her test of 'excess sensitivity' can be interpreted as a test of the correct specification of the ARMA process that she assumes generates $\widetilde{\mathcal{I}}_{i,0}$ which is unobserved (by the economist), although she does not state it that way. Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2004) extend her analysis but, like her, maintain an *a priori* specification of the stochastic process generating $\mathcal{I}_{i,0}$. Blundell, Pistaferri, and Preston (2004) claim to test for 'partial insurance.' In fact their procedure can be viewed as a test of their specification of the stochastic process generating the agent's information set. More closely related to our work is the analysis of Pistaferri (2001), who uses the distinction between expected starting wages (to measure expected returns) and realized wages (to measure innovations) in a consumption analysis.

[142] Keane and Wolpin assume one factor where the θ is a discrete variable and they assume all factor loadings are identical across periods. However, their specification of the uniquenesses or innovations is more general than that used in factor analysis. See our discussion of their model in Section 10.8. The analysis of Hartog and Vijverberg (2002) is another example and uses variances of *ex post* income to proxy *ex ante* variability, removing "fixed effects" (person specific θ).

enter (34), 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.

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. 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.[143] A central issue, discussed next, is how far one can go in identifying income information processes without specifying preferences, insurance, and market environments.

10.5 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 that has been used thus far breaks down.

If we postulate information processes *a priori*, and assume that preferences are known up to

[143] In a model with complete autarky with preferences Ψ, ignoring costs,

$$I_i = \sum_{t=0}^{T} E\left[\frac{\Psi\left(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}\right) - \Psi\left(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \;\middle|\; \widetilde{\mathcal{I}}_{i,0} \right],$$

where ρ is the time rate of discount, we can make a similar decomposition but it is more complicated given the nonlinearity in Ψ. For this model we could do a Sims noncausality test where

$$\begin{aligned} V_{\widetilde{\mathcal{I}}_{i,0}} \;=\; & \sum_{t=0}^{T} \frac{\Psi\left(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}\right) - \Psi\left(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} - \\ & \sum_{t=0}^{T} E\left[\frac{\Psi\left(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}\right) - \Psi\left(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \;\middle|\; \widetilde{\mathcal{I}}_{i,0} \right]. \end{aligned}$$

This requires some specification of Ψ. See Carneiro, Hansen, and Heckman (2003), who assume $\Psi(Y) = \ln Y$ and that the equation for $\ln Y$ is linear in parameters. Cunha, Heckman, and Navarro (2005b) and Navarro (2004) generalize that framework to a model with imperfect capital markets where some lending and borrowing is possible.

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, Heckman, and Navarro (2005b) 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.

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. Cunha, Heckman, and Navarro (2005b) add 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. 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, Heckman, and Navarro (2005c).

10.6 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. See CHN who combine NLSY and PSID data sets. We summarize the analysis of Cunha, Heckman, and Navarro (2005c) in this subsection. See their paper for their exclusions and identification conditions.

Following the preceding theoretical analysis, they consider only two schooling choices: high school and college graduation.[144] For simplicity and familiarity, we focus on their results based on complete contingent claims markets. Because they assume that all shocks are idiosyncratic and the operation of complete markets, schooling choices are made on the basis of expected present value income maximization. Carneiro, Hansen, and Heckman (2003) assume the absence of any credit

[144] Heckman and Navarro (2005) present a model with multiple schooling levels.

markets or insurance. Cunha, Heckman, and Navarro (2005b) check whether their empirical findings about components of income inequality are robust to different assumptions about the availability of credit markets and insurance markets. They estimate an Aiyagari-Laitner economy with a single asset and borrowing constraints and discuss risk aversion and the relative importance of uncertainty. We summarize the evidence from alternative assumptions about market structures below.

10.6.1 Identifying Joint Distributions of Counterfactuals and the Role of Costs and Ability as Determinants of Schooling

Suppose that the error term for $Y_{s,t}$ is generated by a two factor model,

$$Y_{s,t} = X\beta_{s,t} + \theta_1 \alpha_{s,t,1} + \theta_2 \alpha_{s,t,2} + \varepsilon_{s,t}. \tag{35}$$

We omit the "i" subscripts to eliminate notational burden. Cunha, Heckman, and Navarro (2005c) report that two factors are all that is required to fit the data.

They use a test score system of K ability tests:

$$T_{jT} = X_T \omega_{jT} + \theta_1 \alpha_{jT} + \varepsilon_{jT}. \qquad j = 1, \ldots, K \tag{36}$$

Thus factor 1 is identified as an ability component. The cost function C is specified by:

$$C = Z\gamma + \theta_1 \alpha_{C,1} + \theta_2 \alpha_{C,2} + \varepsilon_C. \tag{37}$$

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:

$$S^* = E\left(Y_{1,0} + \frac{Y_{1,1}}{1+r} - Y_{0,0} - \frac{Y_{0,1}}{1+r} \mid X, \bar{\theta}\right) - E\left(C \mid Z, X, \bar{\theta}, \varepsilon_C\right) \tag{38}$$

$$S = \mathbf{1}\left(S^* \geq 0\right). \tag{39}$$

Cunha, Heckman, and Navarro (2005c) report evidence that the estimated factors are highly non-normal.[145]

Table 15 presents the conditional distribution of *ex post* potential college earnings given *ex post* potential high school earnings, decile by decile, as reported by Cunha, Heckman, and Navarro (2005c). The table displays a strong positive dependence between the relative positions of individuals in the two distributions. In particular, for all high school deciles more than 50% of the individuals located at any decile in the high school earnings distribution will be within one decile of their original position in the college earnings distribution. However, the dependence is far from perfect. For example, almost 10% of those who are at the sixth decile of the high school distribution would be in the eighth decile of the college distribution. Note that this comparison is not being made in terms of positions in the overall distribution of earnings. CHN 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 perfect dependence across components of counterfactual distributions that is maintained in much of the recent literature (*e.g.* Juhn, Murphy, and Pierce, 1993) is too strong.

Figure 16 presents the marginal densities of predicted (actual) earnings for college students and counterfactual college earnings for actual high school students. When we compare the densities of present value of earnings in the college sector for persons who choose college against the counterfactual densities of college earnings for high school 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 college earnings for high school graduates. Despite the shift, many high school graduates would earn more as college graduates.

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 $703.78 thousand dollars over

[145] They assume that each factor k, $k \in \{1, 2\}$, is generated by a mixture of J_k normal distributions,

$$\theta_k \backsim \sum_{j=1}^{J_k} p_{k,j} \phi \left(f_k \mid \mu_{k,j}, \tau_{k,j} \right),$$

where the $p_{k,j}$ are the weights on the normal components.

the life cycle. She would earn \$1,021.97 thousand if she had chosen to be a college graduate.[146] This implies a return of 46% to a college education over the whole life cycle (*i.e.,* a monetary gain of 318.19 thousand dollars). In table 17, CHN note that the typical college graduate earns \$1,122.69 thousand dollars (above the counterfactual earnings of what a typical high school student would earn in college), and would make only \$756.13 thousand dollars over her lifetime if she 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 50% above that of the return for a high school graduate. In monetary terms a college graduate has a gain of going to college almost 50,000 dollars higher over her lifetime than does the typical high school graduate.

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, CHN follow convention in the literature and actually present growth rates in terms of present values, and not true rates of return (ignoring option values).[147] 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. 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. An economic interpretation of figure 17 is that agents who choose a college education are the ones who tend to gain more from it.

With their methodology, CHN can also determine returns to the marginal student. This could also be estimated by the *MTE* method discussed in section 8. Table 18 reveals that the average individual who is just indifferent between a college education and a high school diploma earns \$743.40 thousand dollars as a high school graduate or \$1,089.97 thousand dollars as a college graduate. This implies a return of 48%. 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

[146] These numbers may appear to be large but are a consequence of using a 3% discount rate.
[147] Recall our discussion in section 8.2

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.

A major question that emerges from the CHN analysis is, why, if high school graduates have such positive returns to attending college, don't all attend? People do not pick schooling levels based only on monetary returns. Recall that their choice criterion (equation 38) includes also the pecuniary and non-pecuniary costs of actually attending college. Figure 18 shows the estimated density of the monetary value of this cost both overall and by schooling level. While almost no high school graduate perceives a negative cost (*i.e.*, a benefit) of attending college; around one third of college graduates actually perceive it as a benefit. 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. While in both cases the cost of attending college is positive for both the average college graduate and the average high school graduate, costs on average are smaller for college graduates. We know that 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. This is the answer to the 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) discussed in section 4 are so high. 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.[148] Explanations based on psychic costs are not intrinsically satisfactory. One can rationalize any economic choice data by an appeal to psychic costs. Thus the high monetary returns to schooling and the relatively low take up rates in schooling can be explained in this way. "Psychic costs" may be a stand in for credit constraints and risk aversion.

[148] "Psychic costs" can stand in for expectational errors and attitudes towards risk. We do not distinguish among these explanations in this chapter. The estimated costs are too large to be due to tuition alone. As noted below, given that returns are strongly forecastable, an important role for expectational errors seems unlikely. See the discussion in Cunha, Heckman, and Navarro (2005d).

However, the evidence on psychic costs is more sturdy than this. Carneiro, Hansen, and Heckman (2003) obtain similar conclusions from a model where people are not allowed to borrow or lend. In Cunha, Heckman, and Navarro (2005d), on the other hand, there are no constraints on borrowing or lending. Cunha, Heckman, and Navarro (2005d) present additional evidence on the importance of psychic costs. Navarro (2004) reports high estimations of psychic costs in an Aiyagari-Laitner economy where trades are restricted.

10.6.2 *Ex ante* and *Ex post* Returns: Heterogeneity versus Uncertainty

Figures 19 through 21, from Cunha, Heckman, and Navarro (2005c) separate the effect of heterogeneity from uncertainty in earnings. The figures plot the distribution of *ex ante* and *ex post* outcomes. 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 their papers, the various information sets consist of different components of θ. We 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,[149] so that, $\Theta = \{\theta_1\}$, the reduction in the forecast variance is very small. 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. 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.

Table 20 presents the variance of potential earnings in each state, and returns under different information sets available to the agent. Cunha, Heckman, and Navarro (2005c) conduct this exercise for the forecast of period 1, period 2, and lifetime earnings. They report baseline variances and covariances without conditioning and state the remaining uncertainty as a fraction of the baseline

[149] As opposed to the econometrician who never gets to observe either θ_1 or θ_2.

no-information state variance when different components of θ are known to the agents. Note that knowledge of factor 2 is fundamentally important in reducing forecast variance for period 2 earnings. CHN show that both θ_1 and θ_2 are known to the agents at the time they make their schooling decisions.

This discussion sheds light on the issue of distinguishing predictable heterogeneity from uncertainty. CHN demonstrate that there is a large dispersion in the distribution of the present value of earnings. This dispersion is largely due to heterogeneity, which is forecastable by the agents at the time they are making their schooling choices. CHN provide tests that determine that agents know θ_1 and θ_2. The remaining dispersion is due to luck, or uncertainty or unforecastable factors as of age 17. Its contribution is smaller.

It is interesting to note that knowledge of the factors enables agents to make better forecasts. Figure 21 presents an exercise for returns to college $(Y_1 - Y_0)$ similar to that presented in figures 19 and 20 regarding information sets available to the agent. Knowledge of factor 2 also greatly improves the forecastability of returns. 80% of the variability in returns is forecastable at age 17. The levels are even more predictable (94% for high school; 97% for college). Most variability across people is due to heterogeneity and not uncertainty. However there is still a lot of variability in agent earnings after accounting for what is in the agent's information set. This is intrinsic uncertainty at the time agents make their schooling choices.

10.6.3 *Ex Ante* versus *Ex Post*

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, CHN 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." 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 V since *ex post* all terms on the right hand side of the top equation are known with certainty by the agent.

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. They analyze for example, 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) and Cunha, Heckman, and Navarro (2005a,c,d).

10.7 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. Cunha, Heckman, and Navarro (2005b) estimate a version of this model with restriction on intertemporal trade as in the Aiyagari-Laitner economy. 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% (Cunha, Heckman, and Navarro, 2005b,d) for complete markets, to 6% (Cunha, Heckman, and Navarro, 2005c) for another complete market economy.[150]

This line of work has just begun. It shows what is possible with rich panel data. The empirical evidence on the importance of uncertainty is not yet settled. Yet most of the papers suggest a substantial role for uncertainty in producing returns. Accounting for uncertainty and psychic costs may help to explain the high *ex post* rates of return presented in section 4. 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.

[150] The 6% is a lower bound because in Cunha, Heckman, and Navarro (2005c) some of the data are smoothed.

10.8 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. We have discussed simple models of dynamic sequential decision making in section 5. 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.

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).[151] 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 $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{s,t})$ 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. They make parametric normality assumptions in estimating their models. They impose and do not test the particular information structure that they use.[152] In their model, about 90% of the variance in lifetime returns is predictable at age 16. Their estimate is at the extreme boundary of the estimates produced from the CHN analysis.

Heckman and Navarro (2005) formulate, estimate and identify semiparametric sequential schooling models based on the factor structures exposited in sections 9 and 10 of this chapter. Like Keane and Wolpin, they assume that a complete contingent claims market governs the data. They report substantial effects on empirical estimates from relaxing normality assumptions. Instead of assuming a particular information structure, they test among alternative information structures about the

[151] Thus instead of assuming that θ is continuous, as do CHN, they impose that θ is a discrete-valued random variable that assumes a finite, known number of values.

[152] Keane and Wolpin (1997) impose their discrete factor in the schooling choice and outcome equations rather than testing for whether or not the factor appears in both sets of equations in the fashion of CHN as previously described.

arrival of information on the components of vector θ at different stages of the lifecycle. Their work supports the analysis of CHN 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. Their analysis shows that estimated option values for attending college are relatively small (at most 2% of the total return). This is consistent with a lot of information known about future returns at the time schooling decisions are being made. If their results hold up in subsequent analyses and replications, they imply that the theoretical possibilities that arise from accounting for option values as discussed in section 5 may be empirically unimportant in computing rates of return.[153]

11 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.

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.

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, section 4 considers alternative

[153] Assuming that initial conditions are known by agents, the estimates of Keane and Wolpin (1997) are consistent with small option values although they do not report the option values implicit in their estimates.

approaches to estimating the marginal internal rate of return to schooling across different schooling levels.

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. 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.

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. 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.

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.

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. Results from Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005a,b,c,d) discussed in section 10 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, 2005).

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. 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.

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.[154]

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 facilitates identification of the distribution of rates of return.

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 (2005a,b,c,d), 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. To the extent that individuals are risk averse, the

[154] Manski (2004) presents a comprehensive survey of recent research on measuring expectations.

evidence on uncertainty helps to explain some of the high estimated returns to schooling reported in section 4 and in the entire literature (see Navarro, 2004; Cunha, Heckman, and Navarro, 2005b). At the same time, a substantial amount of observed variability in earnings is predictable at the date schooling decisions are made.

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. 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.

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. The recent literature finds a high 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.

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, 2005b; 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. 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.

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.

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.

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. 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.[155] The new literature is beginning to sort out these competing explanations. Recent developments

[155]Recent work by Heckman and LaFontaine (2005) suggests that in recent cohorts dropouts are relatively more educated than in earlier cohorts so the basic facts work against the hypothesis suggested in the text.

in the literature employ new methods to take advantage of rich longitudinal microdata in order to begin distinguishing among the many possibilities.

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 A 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

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

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

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

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

$$\begin{aligned}\rho_0 \sum_{j=0}^{x-1} (1-j/T) &= 1 - x/T, \qquad \text{or} \\ \rho_0 \left(x - \frac{x(x-1)}{2T} \right) &= 1 - x/T\end{aligned}$$

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. More generally, re-arranging terms and solving for the root of this equation[156] yields the variance minimizing

[156] There is a second root which is greater than T (the maximum working age), so it is ignored.

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.

Appendix B Data description

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. 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.

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.

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 individuals 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.

The following variables are constructed:

experience: Potential experience is measured by Age − Years of Education − 6.

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 is 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.

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 is reported as the highest grade completed. Categories of schooling include 9-11 years, 12 years, and 16 years. From 1992-2000, years of education is 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.

Appendix C Local Linear Regression

In estimating the nonparametric matching regressions, we use local linear regression methods. As discussed in Fan and Gijbels (1996), the local linear estimator for the conditional expectation $E[y_i|\, x_i = x_0]$ can be computed from the minimization problem

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - b_1 (x_i - x_0))^2 K\left(\frac{x_i - x_0}{h_n}\right),$$

where $K(\cdot)$ is a kernel function and $h_n > 0$ is a bandwidth which converges to zero as $n \to \infty$.[157] The estimator of the conditional mean $E[y_i|\, x_i = x_0]$ 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

$$W_i(x_0) = \frac{K_i \sum_{j=1}^{n} K_j^2 - K_i \sum_{k=1}^{n} K_k}{\sum_{k=1}^{n} K_k \sum_{j=1}^{n} K_j^2 - (\sum_{k=1}^{n} K_k)^2}.$$

Taking advantage of the fact that we have many observations with repeated x_i values, our local regression estimator is given by

$$\hat{m}(x_0) = \frac{\sum_{i=1}^{N_x} n_{x_i} y(x_i) W_i(x_0)}{\sum_{i}^{N_X} n_{x_i} W_i(x_0)},$$

where $y(x_i)$ represents average log earnings at experience level x_i, n_{x_i} represents the number of observations at experience level x_i, and N_X represents the number of distinct values of potential experience.[158]

The asymptotic distribution of the estimator $\hat{m}(x_0)$ for $m(x_0) = E(y_i|x_i = x_0)$ is given by

$$(nh_n)^{-1/2}(\hat{m}(x_0) - m(x_0)) \sim N(B_n, V_n) + o_p(1)$$

[157] The kernel function we use in the empirical work is the quartic kernel, given by

$$K(s) = \begin{cases} (15/16)(s^2-1)^2 & \text{if } |s| < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The bandwidth used is equal to 5.

[158] For some of the Census years, there is a problem of non-random sampling with sampling weights provided in the data. The sampling weights are taken into account when calculating the mean log earnings at each experience level.

where the bias and variance expressions are given by

$$\begin{aligned} B_n &= h_n^2 \cdot (0.5m''(x_0)) \cdot \int_{-\infty}^{\infty} u^2 K(u)du \\ V_n &= \sigma^2(x_0) \int_{-\infty}^{\infty} K^2(u)du, \end{aligned}$$

and where $\sigma^2(x_0) = E(\{y_i - E(y_i|x_i = x_0)\}^2|x_i = x_0)$.[159]

Tests of Parallelism

In section 3 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. 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

$$\begin{aligned} &(\hat{m}(x_i, s_1) - \hat{m}(x_i, s_2) - (\hat{m}(x_k, s_1) - \hat{m}(x_k, s_2)) \cdot \\ &(\hat{V}_1 + \hat{V}_2 + \hat{V}_3 + \hat{V}_4)^{-1} \cdot \\ &(\hat{m}(x_i, s_1) - \hat{m}(x_i, s_2) - (\hat{m}(x_k, s_1) - \hat{m}(x_k, s_2)), \end{aligned}$$

where $\hat{V}_1$, $\hat{V}_2$, $\hat{V}_3$, and $\hat{V}_4$ are estimators for $V_1 = Var(\hat{m}(x_i, s_1))$, $V_1 = Var(\hat{m}(x_i, s_2))$, $V_3 = Var(\hat{m}(x_k, s_1))$, $V_3 = Var(\hat{m}(x_k, s_2))$.

Under the null hypothesis of parallelism, the bias terms cancel out, so that it is not necessary

[159] See, *e.g.* Fan and Gijbels (1996), for derivation of these formulae.

to estimate the bias expressions in performing the test.[160] To estimate the variances, we use

$$Var(\hat{m}(x_i, s_1)) = \frac{\sum_{i=1}^{N_X} n_{x_i} \hat{\varepsilon}(x_i, s_1)^2 W_i(x_i)}{\sum_i^{N_X} n_{x_i} W_i(x_i)},$$

where $\hat{\varepsilon}(x_i, s_1) = y(x_i, s_1) - \hat{m}(x_i, s_1)$ is the fitted residual from the nonparametric regression evaluated at experience level x_i.[161] In Table 1, we report test results based on the test statistic that straightforwardly generalizes the test statistic given above to four experience levels.

[160] This cancelling only occurs with the local linear estimator and would not occur if the standard kernel estimator were used instead to generate the nonparametric estimates.

[161] Heckman, Ichimura, Smith, and Todd (1998) show that this estimator has better finite sample performance than a "plug-in" estimator based on the asymptotic variance formulae.

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics 125*(1-2), 15–51.

Aiyagari, S. R. (1994, August). Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics 109*(3), 659–684.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Angrist, J. D. and G. W. Imbens (1995, June). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Angrist, J. D. and A. B. Krueger (1991, November). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics 106*(4), 979–1014.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, pp. 1277–1366. New York: North-Holland.

Angrist, J. D. and A. B. Krueger (2001, Fall). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives 15*(4), 69–85.

Atkeson, A. and R. E. Lucas (1992, July). On efficient distribution with private information. *The Review of Economic Studies 59*(3), 427–453.

Barro, R. J. and C. Sahasakul (1983, October). Measuring the average marginal tax rate from the individual income tax. *Journal of Business 56*(4), 419–452.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research, distributed by Columbia University Press.

Becker, G. S. (1967). Human capital and the personal distribution of income: An analytical approach. Woytinsky Lecture no. 1. Ann Arbor: University of Michigan, Institute of Public Administration.

Becker, G. S. and B. R. Chiswick (1966, March). Education and the distribution of earnings. *The American Economic Review 56*(1/2), 358–369.

Behrman, J. R. and N. Birdsall (1983, December). The quality of schooling: Quantity alone is misleading. *American Economic Review 73*(5), 928–946.

Belzil, C. and J. Hansen (2002, September). Unobserved ability and the return to schooling. *Econometrica 70*(5), 2075–2091.

Belzil, C. and J. Hansen (2005, May). A structural analysis of the correlated random coefficient wage regression model with an application to the OLS-IV puzzle. Discussion paper 1585, IZA.

Ben-Porath, Y. (1967, August). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4, Part 1), 352–365.

Bils, M. and P. J. Klenow (2000, December). Does schooling cause growth? *American Economic Review 90*(5), 1160–1183.

Bishop, J. H. (1991). Achievement, test scores, and relative wages. In M. H. Kosters (Ed.), *Workers and their wages: Changing patterns in the United States*, Number 520 in AEI Studies, pp. 146–186. Washington, D.C.: AEI Press.

Björklund, A. and R. Moffitt (1987, February). The estimation of wage gains and welfare gains in self-selection. *Review of Economics and Statistics 69*(1), 42–49.

Blackburn, M. L. and D. Neumark (1993, July). Omitted-ability bias and the increase in the return to schooling. *Journal of Labor Economics 11*(3), 521–544.

Blundell, R., L. Pistaferri, and I. Preston (2004, October). Consumption inequality and partial insurance. Technical Report WP04/28, Institute for Fiscal Studies.

Blundell, R. and I. Preston (1998, May). Consumption inequality and income uncertainty. *Quarterly Journal of Economics 113*(2), 603–640.

Blundell, R., H. Reed, and T. Stoker (2003, September). Interpreting aggregate wage growth: The role of labor market participation. *American Economic Review 93*(4), 1114–1131.

Bound, J., D. A. Jaeger, and R. M. Baker (1995, June). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Buchinsky, M. and P. Leslie (2000, January). Educational attainment and the changing U.S. wage structure: Dynamic implications without rational expectations. Unpublished manuscript, UCLA Department of Economics.

Cameron, S. V. and J. J. Heckman (1998, April). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy 106*(2), 262–333.

Cameron, S. V. and C. Taber (2004, February). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy 112*(1), 132–182.

Card, D. (1993, October). Using geographic variation in college proximity to estimate the return to schooling. Technical Report 4483, National Bureau of Economic Research.

Card, D. (1995). Earnings, schooling, and ability revisited. In S. Polachek (Ed.), *Research in Labor Economics*, pp. 23–48. Greenwich, CT: JAI Press.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 5, pp. 1801–1863. New York: North-Holland.

Card, D. (2001, September). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Card, D. and A. B. Krueger (1992, February). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy 100*(1), 1–40.

Card, D. and T. Lemieux (2001, May). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *Quarterly Journal of Economics 116*(2), 705–746.

Carneiro, P. (2002). *Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation*. Ph. D. thesis, University of Chicago.

Carneiro, P., K. Hansen, and J. J. Heckman (2001, Fall). Removing the veil of ignorance in assessing the distributional impacts of social policies. *Swedish Economic Policy Review 8*(2), 273–301.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422. 2001 Lawrence R. Klein Lecture.

Carneiro, P. and J. J. Heckman (2002, October). The evidence on credit constraints in post-secondary schooling. *Economic Journal 112*(482), 705–734.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2005). Understanding what instrumental variables estimate: Estimating marginal and average returns to education. Paper presented as *Economic Journal* Lecture at the Royal Economic Society meeting, Durham, England. Also presented as the *Review of Economics and Statistics* Lecture, Harvard University, April 2001. Under review.

Chandra, A. (2003). Is the convergence of the racial wage gap illusory? Working Paper 9476, National Bureau of Economic Research.

Chiswick, B. R. (1974). *Income Inequality: Regional Analyses Within a Human Capital Framework*. New York: National Bureau of Economic Research.

Chiswick, B. R. and J. Mincer (1972, May/June). Time-series changes in personal income inequality in the United States from 1939, with projections to 1985. *Journal of Political Economy 80*(3 (Part II)), S34–S66.

Comay, Y., A. Melnik, and M. A. Pollatschek (1973, June). The option value of education and the optimal path for investment in human capital. *International Economic Review 14*(2), 421–435.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2005). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 7. North-Holland. forthcoming.

Cunha, F., J. J. Heckman, and S. Navarro (2005a). Counterfactual analysis of inequality and social mobility. In S. L. Morgan, D. B. Grusky, and G. S. Fields (Eds.), *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, Chapter 4. Palo Alto: Stanford University Press. forthcoming.

Cunha, F., J. J. Heckman, and S. Navarro (2005b, August). The evolution of uncertainty in the US economy. Presented at the 9th World Congress of the Econometric Society, London. Previously "Separating Heterogeneity from Uncertainty in an Aiyagari-Laitner Economy," presented a the Goldwater Conference on Labor Markets, Arizona State University, March 2004.

Cunha, F., J. J. Heckman, and S. Navarro (2005c). A framework for the analysis of inequality. *Journal of Macroeconomics*. forthcoming.

Cunha, F., J. J. Heckman, and S. Navarro (2005d, April). Separating uncertainty from heterogeneity in life cycle earnings, the 2004 Hicks Lecture. *Oxford Economic Papers 57*(191-261), 1–72.

Dixit, A. K. and R. S. Pindyck (1994). *Investment Under Uncertainty*. Princeton: Princeton University Press.

Dooley, M. D. and P. Gottschalk (1984, February). Earnings inequality among males in the United States: Trends and the effect of labor force growth. *Journal of Political Economy 92*(1), 59–89.

Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute 22*, 23–32.

Eckstein, Z. and K. I. Wolpin (1999, November). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica 67*(6), 1295–1339.

Ellwood, D. T. and T. J. Kane (2000). Who is getting a college education? Family background and the growing gaps in enrollment. In S. Danziger and J. Waldfogel (Eds.), *Securing the Future: Investing in Children from Birth to College*, pp. 283–324. New York: Russell Sage Foundation.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*. New York: Chapman and Hall.

Flavin, M. A. (1981, October). The adjustment of consumption to changing expectations about future income. *Journal of Political Economy 89*(5), 974–1009.

Florens, J.-P., J. J. Heckman, C. Meghir, and E. J. Vytlacil (2002). Instrumental variables, local instrumental variables and control functions. Technical Report CWP15/02, CEMMAP.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon A*(Series 3, 14), 53–77.

Friedman, M. and S. S. Kuznets (1945). *Income from Independent Professional Practice*. New York: National Bureau of Economic Research.

Glewwe, P. (2002, June). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature 40*(2), 436–482.

Greene, W. H. (2003). *Econometric Analysis* (5 ed.). Upper Saddle River, N.J.: Prentice Hall.

Griliches, Z. (1977, January). Estimating the returns to schooling: Some econometric problems. *Econometrica 45*(1), 1–22.

Grogger, J. and E. Eide (1995, Spring). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources 30*(2), 280–310.

Hanoch, G. (1967, Summer). An economic analysis of earnings and schooling. *Journal of Human Resources 2*(3), 310–329.

Hansen, L. P. (1982, July). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

Hartog, J. and W. Vijverberg (2002, February). Do wages really compensate for risk aversion and skewness affection? Technical Report IZA DP No. 426, IZA, Bonn, Germany.

Hause, J. C. (1980, May). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica 48*(4), 1013–1029.

Hausman, J. A. (1978, November). Specification tests in econometrics. *Econometrica 46*, 1251–1272.

Heckman, J. J. (1976, August). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy 84*(4, Part 2), S11–S44. Journal Special Issue: Essays in Labor Economics in Honor of H. Gregg Lewis.

Heckman, J. J. (1979, January). Sample selection bias as a specification error. *Econometrica 47*(1), 153–162.

Heckman, J. J. (1990, May). Varieties of selection bias. *American Economic Review 80*(2), 313–318.

Heckman, J. J. (1997, Summer). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources 32*(3), 441–462. Addendum published vol. 33 no. 1 (Winter 1998).

Heckman, J. J. and B. E. Honoré (1990, September). The empirical content of the Roy model. *Econometrica 58*(5), 1121–1149.

Heckman, J. J., H. Ichimura, J. Smith, and P. E. Todd (1998, September). Characterizing selection bias using experimental data. *Econometrica 66*(5), 1017–1098.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997, October). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies 64*(4), 605–654.

Heckman, J. J., H. Ichimura, and P. E. Todd (1998, April). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*(223), 261–294.

Heckman, J. J. and P. LaFontaine (2005). Bias corrected estimates of GED returns. *Journal of Labor Economics*. forthcoming.

Heckman, J. J., A. Layne-Farrar, and P. E. Todd (1996, November). Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *Review of Economics and Statistics 78*(4), 562–610.

Heckman, J. J., L. J. Lochner, and C. Taber (1998, January). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics 1*(1), 1–58.

Heckman, J. J., T. M. Lyons, and P. E. Todd (2000, May). Understanding black-white wage differentials: 1960-1990. *American Economic Review 90*(2), 344–349.

Heckman, J. J. and S. Navarro (2005). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*. forthcoming.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions. In J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Volume 10, pp. 156–245. New York: Cambridge University Press.

Heckman, J. J. and R. Robb (1986). Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples*, pp. 63–107. New York: Springer-Verlag. Reprinted in 2000, Mahwah, NJ: Lawrence Erlbaum Associates.

Heckman, J. J. and J. Scheinkman (1987, April). The importance of bundling in a Gorman-Lancaster model of earnings. *Review of Economic Studies 54*(2), 243–355.

Heckman, J. J. and G. L. Sedlacek (1985, December). Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy 93*(6), 1077–1125.

Heckman, J. J. and J. A. Smith (1998). Evaluating the welfare state. In S. Strom (Ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pp. 241–318. New York: Cambridge University Press.

Heckman, J. J., J. A. Smith, and N. Clements (1997, October). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies 64*(221), 487–536.

Heckman, J. J., J. L. Tobias, and E. J. Vytlacil (2001, October). Four parameters of interest in the evaluation of social programs. *Southern Economic Journal 68*(2), 210–223.

Heckman, J. J., J. L. Tobias, and E. J. Vytlacil (2003, August). Simple estimators for treatment parameters in a latent variable framework. *Review of Economics and Statistics 85*(3), 748–754.

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2004). Understanding instrumental variables in models with essential heterogeneity. Unpublished manuscript, University of Chicago, Department of Economics. *Review of Economics and Statistics* Lecture, 2002.

Heckman, J. J. and E. J. Vytlacil (1998, Fall). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources 33*(4), 974–987.

Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*, 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2000, January). The relationship between treatment parameters within a latent variable framework. *Economics Letters 66*(1), 33–39.

Heckman, J. J. and E. J. Vytlacil (2001a, February). Identifying the role of cognitive ability in explaining the level of change in the return to schooling. *Review of Economics and Statistics 83*(1), 1–12.

Heckman, J. J. and E. J. Vytlacil (2001b). Local instrumental variables. In C. Hsiao, K. Morimue, and J. L. Powell (Eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, New York, pp. 1–46. Cambridge University Press.

Heckman, J. J. and E. J. Vytlacil (2001c, May). Policy-relevant treatment effects. *American Economic Review 91*(2), 107–111.

Heckman, J. J. and E. J. Vytlacil (2005a). Econometric evaluation of social programs. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume 6*. Amsterdam: Elsevier. forthcoming.

Heckman, J. J. and E. J. Vytlacil (2005b, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Hirshleifer, J. (1970). *Investment, Interest, and Capital*. Englewood Cliffs, NJ: Prentice-Hall.

Hoeffding, W. (1940). Masstabinvariante korrelationtheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik und Universität Berlin 5*, 197–233.

Hungerford, T. and G. Solon (1987, February). Sheepskin effects in the returns to education. *Review of Economics and Statistics 69*(1), 175–177.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Jaeger, D. A. and M. E. Page (1996, November). Degrees matter: New evidence on sheepskin effects in the returns to education. *Review of Economics and Statistics 78*(4), 733–740.

Jöreskog, K. G. and A. S. Goldberger (1975, September). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association 70*(351), 631–639.

Juhn, C., K. M. Murphy, and B. Pierce (1993, June). Wage inequality and the rise in returns to skill. *Journal of Political Economy 101*(3), 410–442.

Kane, T. J., C. Rouse, and D. Staiger (1999). Estimating returns to schooling when schooling is misreported. Technical Report 7235, NBER.

Kane, T. J. and C. E. Rouse (1995, June). Labor-market returns to two- and four-year college. *American Economic Review 85*(3), 600–614.

Katz, L. F. and D. H. Autor (1999). Changes in the wage structure and earnings inequality. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, Chapter 25, pp. 1463–1555. New York: North-Holland.

Katz, L. F. and K. M. Murphy (1992, February). Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics 107*(1), 35–78.

Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Kling, J. R. (2001, July). Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business and Economic Statistics 19*(3), 358–364.

Laitner, J. (1992, December). Random earnings differences, lifetime liquidity constraints, and altruistic intergenerational transfers. *Journal of Economic Theory 58*(2), 135–170.

Lewis, H. G. (1963). *Unionism and Relative Wages in the United States: An Empirical Inquiry*. Chicago: University of Chicago Press.

Lewis, H. G. (1986). *Union Relative Wage Effects: A Survey*. Chicago: University of Chicago Press.

Lillard, L. A. and R. J. Willis (1978, September). Dynamic aspects of earning mobility. *Econometrica 46*(5), 985–1012.

Lochner, L. J. and E. Moretti (2004, March). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review 94*(1), 155–189.

MaCurdy, T. and T. Mroz (1995). Estimating macro effects from repeated cross-sections. Unpublished working paper, Stanford University.

MaCurdy, T. E. (1982, January). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*(1), 83–114.

Manski, C. F. (2004, September). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Matzkin, R. L. (2003, September). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Meghir, C. and M. Palme (2001). The effect of a social experiment in education. Technical Report W01/11, Institute for Fiscal Studies.

Mincer, J. (1958, August). Investment in human capital and personal income distribution. *Journal of Political Economy 66*(4), 281–302.

Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

Mincer, J. (1997). Changes in wage inequality, 1970-1990. In S. W. Polachek and J. Robst (Eds.), *Research in Labor Economics*, Volume 16, pp. 1–18. Greenwich, CT: JAI Press.

Mincer, J. and S. Polachek (1974, Part II, March-April). Family investment in human capital: Earnings of women. *Journal of Political Economy 82*(2), S76–S108.

Mulligan, C. B. and J. G. Marion (2000). Average marginal tax rates revisited: A comment. Working Paper 01.4, Harris School of Public Policy.

Murnane, R. J., J. B. Willett, and F. Levy (1995, May). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*(2), 251–266.

Murphy, K. M. and F. Welch (1990, April). Empirical age-earnings profiles. *Journal of Labor Economics 8*(2), 202–229.

Murphy, K. M. and F. Welch (1992, February). The structure of wages. *Quarterly Journal of Economics 107*(1), 285–326.

Navarro, S. (2004). Understanding schooling: Using observed choices to infer agent's information in a dynamic model of schooling choice when consumption allocation is subject to borrowing constraints. Unpublished manuscript, University of Chicago, Department of Economics.

Neal, D. (2004, February). The measured black-white wage gap among women is too small. *Journal of Political Economy 112*(1, Part 2 Supplement), S1–28.

Noyes, C. R. (1945). Director's comment. In M. Friedman and S. S. Kuznets (Eds.), *Income from Independent Professional Practice*, pp. 405–410. New York: National Bureau of Economic Research.

Pistaferri, L. (2001, August). Superior information, income shocks, and the permanent income hypothesis. *Review of Economics and Statistics 83*(3), 465–476.

Psacharopoulos, G. (1981, October). Returns to education: An updated international comparison. *Comparative Education 17*(3), 321–341.

Psacharopoulos, G. and H. A. Patrinos (2004, August). Returns to investment in education: A further update. *Education Economics 12*(2), 111–134.

Quandt, R. E. (1958, December). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association 53*(284), 873–880.

Quandt, R. E. (1972, June). A new approach to estimating switching regressions. *Journal of the American Statistical Association 67*(338), 306–310.

Robinson, P. M. (1988, July). Root-N-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Rosen, S. (1977). Human capital: A survey of empirical research. In R. Ehrenberg (Ed.), *Research in Labor Economics*, Volume 1, pp. 3–40. Greenwich, CT: JAI Press.

Roy, A. (1951, June). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rutherford, R. S. G. (1955, July). Income distributions: A new model. *Econometrica 23*(3), 277–294.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. New York: Oxford University Press.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without much a priori economic theory. In *New methods in business cycle research: Proceedings from a Conference*. Minneapolis: Federal Reserve Bank of Minneapolis.

Schultz, T. P. (1975). Long-term change in personal income distribution: Theoretical approaches, evidence, and explanations. In D. M. Levine and M. J. Bane (Eds.), *The "Inequality" Controversy: Schooling and Distributive Justice*. New York: Basic Books.

Sims, C. A. (1972, September). Money, income, and causality. *American Economic Review 62*(4), 540–552.

Smith, J. P. and F. R. Welch (1979, June). Inequality: Race differences in the distribution of earnings. *International Economic Review 20*(2), 515–526.

Snyder, T. D. (1993). *120 Years of American Education: A Statistical Portrait*. Washington, D.C.: U.S. Dept. of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Snyder, T. D. (2000). *Digest of Education Statistics 1999*. Washington, DC.: U.S. Dept. of Education, Office of Educational Research and Improvement, National Center for Educational Statistics.

Staiger, D. and J. H. Stock (1997, May). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and M. Yogo (2002). Testing for weak instruments in linear IV regression. Technical Working Paper 284, NBER.

Taber, C. R. (2001, July). The rising college premium in the eighties: Return to college or return to unobserved ability? *Review of Economic Studies 68*(3), 665–691.

Urzua, S. (2005). Schooling choice and the anticipation of labor market conditions: A dynamic choice model with heterogeneous agents and learning. Unpublished manuscript, University of Chicago, Department of Economics.

Vytlacil, E. J. (2002, January). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

Weisbrod, B. A. (1962, October). Education and investment in human capital. *Journal of Political Economy 70*(5, Part 2: Investment in Human Beings), 106–123.

Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics, Volume*, pp. 525–602. New York: North-Holland.

Willis, R. J. and S. Rosen (1979, October). Education and self-selection. *Journal of Political Economy 87*(5, Part 2), S7–S36.

Wu, D. (1973, July). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica 41*, 733–750.

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.

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.

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.

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.

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.

Table 6a: Simulated Returns under Uncertainty with Option Values
(Log Wages Linear in Schooling: $\bar{Y}_{s+1} = (1+r)\bar{Y}_s$)

Education Level (s)	Transition Probability ($p_{s,s-1}$)	Proportional Increase in $\bar{Y}$	Proportional Increase in Observed Earnings	Option/Total Value ($O_{s,s-1}/E_{s-1}(V_s)$)	Average 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.

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

Table 6b: Simulated Returns under Uncertainty with Option Values
(Sheepskin Effects: $\bar{Y}_{s+1} = (1+\rho_{s+1})\bar{Y}_s$ with $\rho_2 = 0.1$, $\rho_3 = 0.3$, $\rho_4 = 0.1$, $\rho_5 = 0.2$)

Education Level (s)	Transition Probability ($p_{s,s-1}$)	Proportional Increase in $\bar{Y}$	Proportional Increase in Observed Earnings	Option/Total Value ($O_{s,s-1}/E_{s-1}(V_s)$)	Average 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.

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

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.

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

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	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.

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; 2005a,b)

Table 9b: Weights*

$h_{TT}(x, u_S) = \left[\int_{u_S}^1 f(p \mid X = x)dp\right] \frac{1}{E(P\mid X=x)}$
$h_{TUT}(x, u_S) = \left[\int_0^{u_S} f(p \mid X = x)dp\right] \cdot \frac{1}{E((1-P)\mid X=x)}$
$h_{PRT}(x, u_S) = \left[\frac{F_{P^*,X}(u_S) - F_{P,X}(u_S)}{\Delta P}\right]$
$h_{IV}(x, u_S) = \left[\int_{u_S}^1 (p - E(P \mid X = x))f(p \mid X = x)dp\right] \frac{1}{Var(P\mid X=x)}$
$h_{OLS} = \frac{E(U_1\mid X=x, U_S=u_S)h_1(x,u_D) - E(U_0\mid X=x, U_S=u_S)h_0(x,u_S)}{MTE(x,u_S)}$
$h_1(x, u_S) = \left[\int_{u_S}^1 f(p \mid X = x)dp\right] \frac{1}{E(P\mid X=x)}$
$h_0(x, u_S) = \left[\int_0^{u_S} f(p \mid X = x)dp\right] \frac{1}{E((1-P)\mid X=x)}$

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

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.

Table 11: Estimates of Various Returns to One Year of College

	Family Background is Exclusion $0.07 < P < 0.98$	Family Background is not Exclusion $0.07 < P < 0.98$
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).

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 (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)

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 (2004)

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).

Table 15

Ex-post Conditional Distribution (College Earnings Conditional on High School Earnings)

$\Pr(d_i < Y_c < d_i + 1 \mid d_j < Y_h < d_j + 1)$*

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.6980	0.2534	0.0444	0.0032	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.2270	0.4150	0.2470	0.0890	0.0180	0.0040	0.0000	0.0000	0.0000	0.0000
3	0.0450	0.2160	0.3420	0.2610	0.1070	0.0260	0.0030	0.0000	0.0000	0.0000
4	0.0140	0.0950	0.2120	0.2930	0.2390	0.1090	0.0370	0.0010	0.0000	0.0000
5	0.0000	0.0300	0.1130	0.2190	0.2940	0.2170	0.1100	0.0170	0.0000	0.0000
6	0.0000	0.0040	0.0340	0.0980	0.2030	0.3080	0.2470	0.0990	0.0070	0.0000
7	0.0000	0.0000	0.0100	0.0340	0.1130	0.2390	0.3190	0.2350	0.0500	0.0000
8	0.0000	0.0000	0.0000	0.0030	0.0240	0.0910	0.2360	0.4010	0.2320	0.0130
9	0.0000	0.0000	0.0000	0.0000	0.0010	0.0060	0.0470	0.2360	0.5400	0.1700
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0110	0.1710	0.8170

*d_i is the ith decile of the College Lifetime Earnings Distribution and d_j is the jth decile of the High School Lifetime Earnings Distribution.

Source: Cunha, Heckman and Navarro (2005c)

Table 16
Average present value of earnings[1] for high school graduates
Fitted and Counterfactual
White males from NLSY79

	High School (fitted)	College (counterfactual)
High School	703.780	1021.970
Std. Err.	14.626	78.214
Random[2]	726.590	1065.900
Std. Err.	20.513	43.054

Average returns[3] for high school graduates	
	High School vs Some College
High School	0.4600
Std. Err.	0.1401

[1]Thousands of dollars. Discounted using a 3% interest rate.

[2]It defines 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.*, (PVEarnings(Col)-PVEarnings(HS))/PVEarnings(HS).

Source: Cunha, Heckman and Navarro (2005c)

Table 17
Average present value of earnings[1] for college graduates
Fitted and Counterfactual
White males from NLSY79

	High School (counterfactual)	College (fitted)
College	756.13	1122.69
Std. Err.	40.571	25.891
Random[2]	726.59	1065.90
Std. Err.	20.513	43.054

Average returns[3] for college graduates	
	High School vs Some College
College	0.50
Std. Err.	0.0805

[1]Thousands of dollars. Discounted using a 3% interest rate.

[2]It defines 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.*, (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Source: Cunha, Heckman and Navarro (2005c)

Table 18

Average present value of earnings[1] for people indifferent between high school and college

Conditional on education level

White males from NLSY79

	High School	College
Average	743.400	1089.970
Std. Err.	24.152	33.255

Average returns[2] for people indifferent between high school and college

	High School vs Some College
Average	0.4800
Std. Err.	0.0853

[1]Thousands of dollars. Discounted using a 3% interest rate.

[2]As a fraction of the base state*, i.e.*, (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Source: Cunha, Heckman and Navarro (2005c)

Table 19

Mean monetary value of total cost of attending college

High School	College	Overall
488.24	232.56	375.27

Mean monetary value of cost of attending college due to ability

High School	College	Overall
40.97	-51.27	0.0

Values in thousands of dollars (2000). Let C denote the monetary value of psychic costs. Then C is given by:

$$C = Z\gamma + \theta_1\alpha_{C1} + \theta_2\alpha_{C2} + \varepsilon_C$$

The contribution of ability to the costs of attending college in monetary value is $\theta_1\alpha_{C1}$. Recall that, on average, the ability is different between those who attend college and those who attend high school.

Source: Cunha, Heckman and Navarro (2005c)

Table 20

Agent's Forecast Variance of Present Value of Earnings

Under Different Information Sets: $I = \{ X, Z, X_T, \varepsilon_C, \Theta \}$

(as a fraction of the variance when no information is available)

	Var(Y_c)	Var(Y_h)	Var(Y_c - Y_h)	Cov(Y_c , Y_h)
For time period 1:[+]				
Variance when $\Theta = \emptyset$	7167.20	5090.46	3073.94	4591.86
Percentage of variance remaining after controlling for the indicated factor:				
$\Theta = \{\theta_1\}$	97.50%	98.34%	99.43%	97.33%
$\Theta = \{\theta_2\}$	18.50%	32.83%	89.52%	2.67%
$\Theta = \{\theta_1, \theta_2\}$	16.01%	31.17%	88.94%	0.00%
For time period 2:[++]				
Variance when $\Theta = \emptyset$	49690.64	167786.87	41137.80	88169.85
Percentage of variance remaining after controlling for the indicated factor:				
$\Theta = \{\theta_1\}$	97.18%	97.54%	98.25%	97.28%
$\Theta = \{\theta_2\}$	7.39%	4.73%	16.55%	2.72%
$\Theta = \{\theta_1, \theta_2\}$	4.57%	2.27%	14.80%	0.00%
For lifetime:[+++]				
Variance when $\Theta = \emptyset$	56857.84	172877.33	44211.74	92761.72
Percentage of variance remaining after controlling for the indicated factor:				
$\Theta = \{\theta_1\}$	97.22%	97.57%	98.33%	97.28%
$\Theta = \{\theta_2\}$	8.79%	5.56%	21.62%	2.72%
$\Theta = \{\theta_1, \theta_2\}$	6.01%	3.13%	19.95%	0.00%

We use an interest rate of 3% to calculate the present value of earnings. In all cases, the information set of the agent is $I = \{ X, Z, X_T, \varepsilon_C, \Theta \}$ and we change the contents of Θ.

[+]Variance of the unpredictable component of earnings between age 17 and 28 as predicted at age 17.

[++]Variance of the unpredictable component of earnings between age 29 and 65 as predicted at age 17.

[+++]Variance of the unpredictable component of earnings between age 17 and 65 as predicted at age 17.

So we would say that the variance of the unpredictable component of period 1 college earnings when using factor 1 in the prediction is 97.5% of the variance when no information is available (*i.e.*, 0.975*7167.2).

Source: Cunha, Heckman and Navarro (2005c)

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

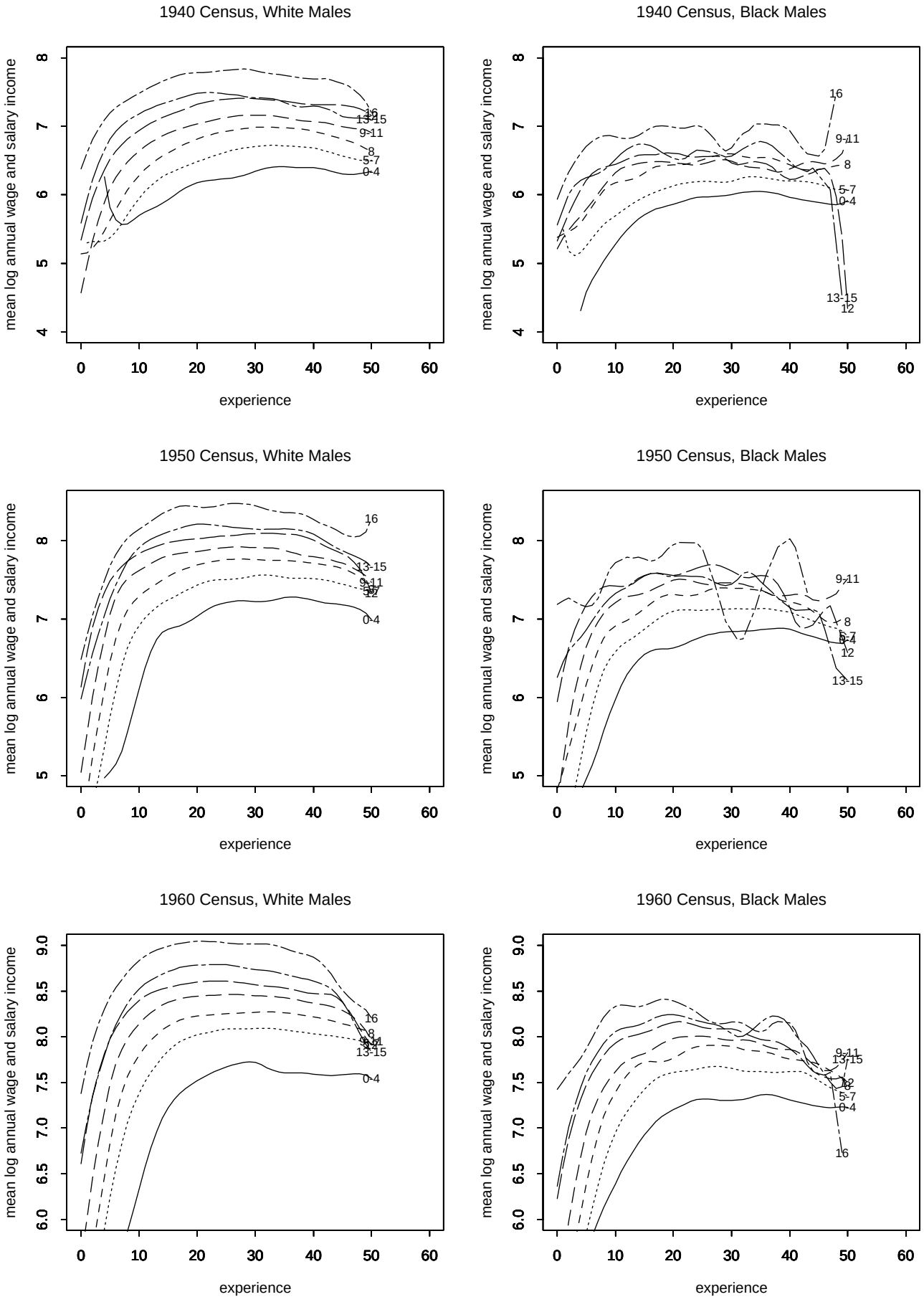


Figure 1b: Experience-Earnings Profiles, 1970-1990

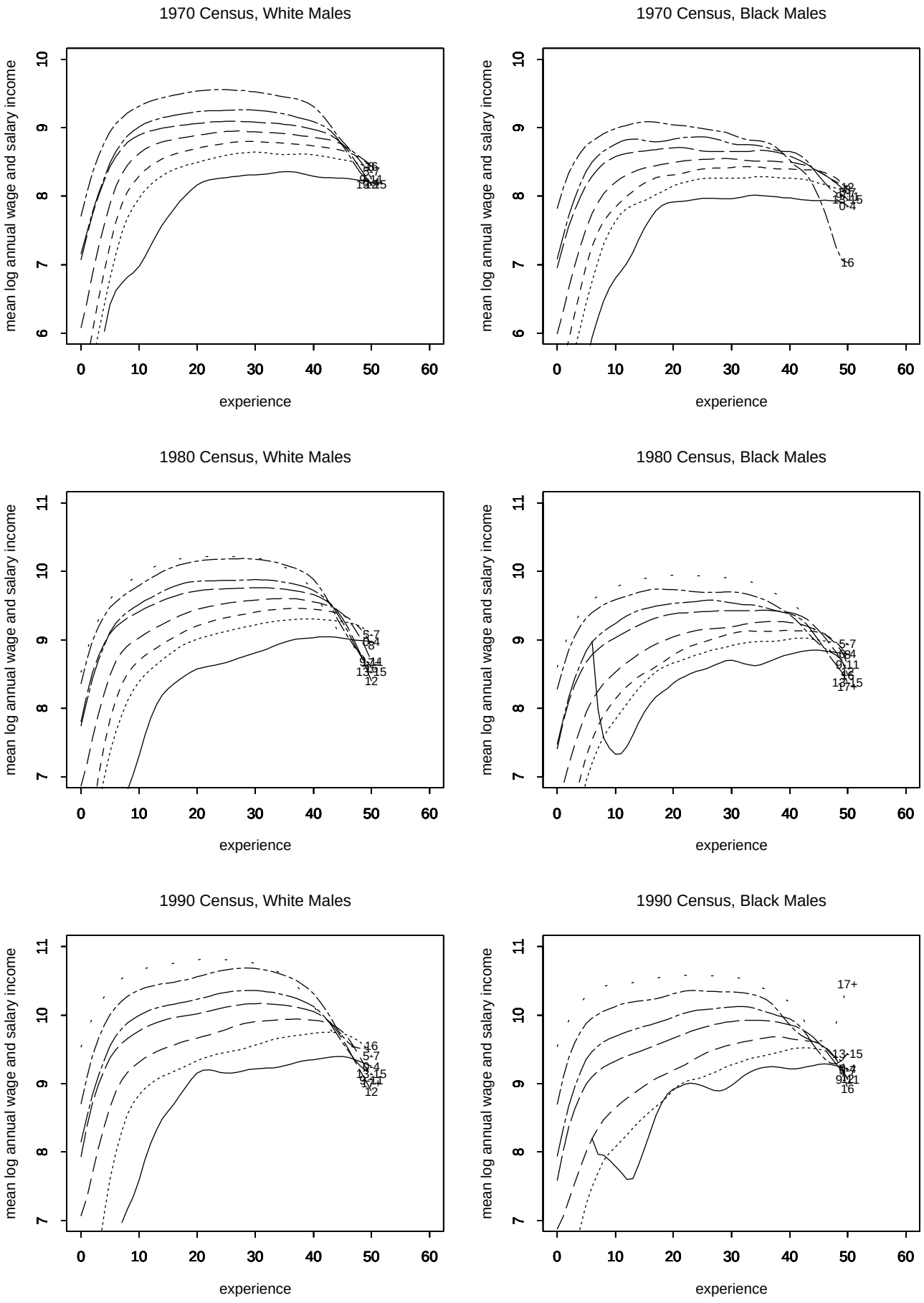


Figure 2: Age-Earnings Profiles, 1940,1960,1980

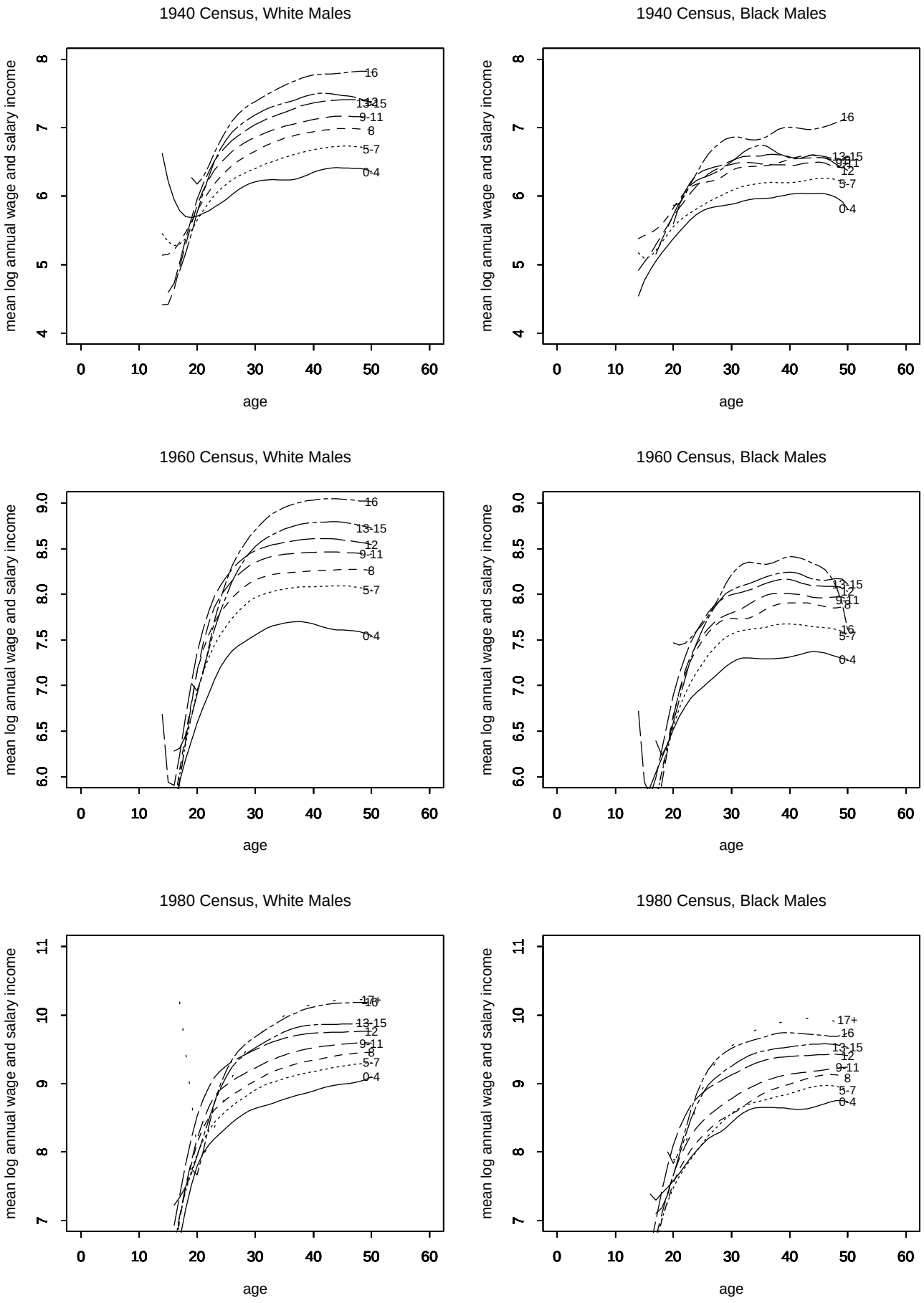


Figure 3: Experience-Variance Log Earnings

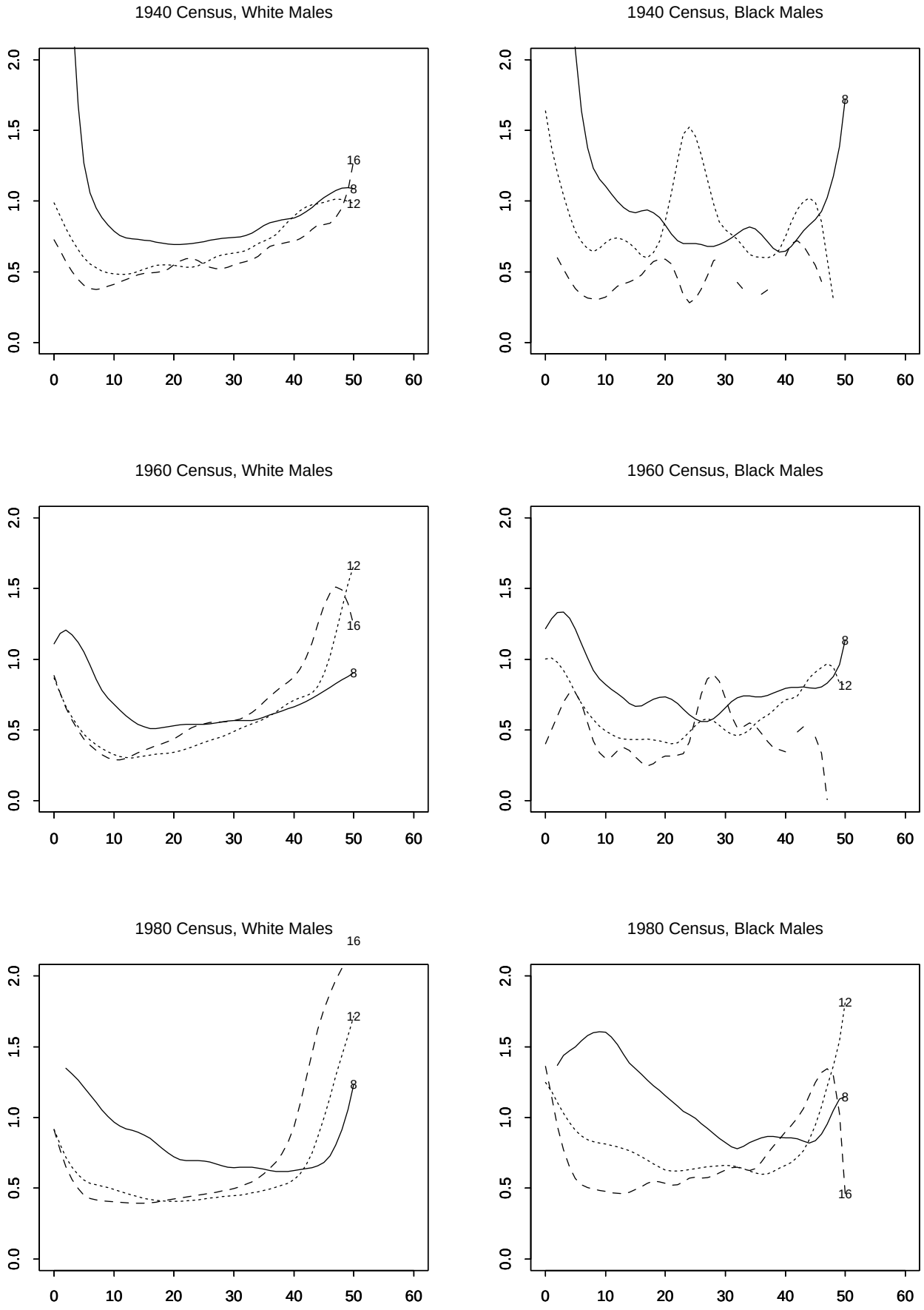


Figure 4a: Average College Tuition Paid (in 2000 dollars)

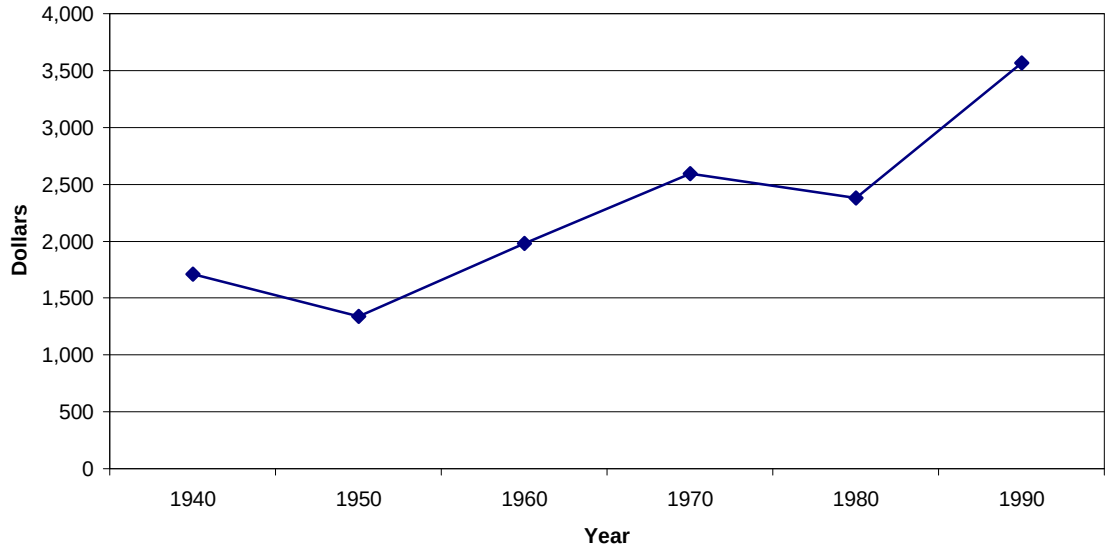


Figure 4b: Marginal Tax Rates
(from Barro & Sahasakul, 1983, Mulligan & Marion, 2000)

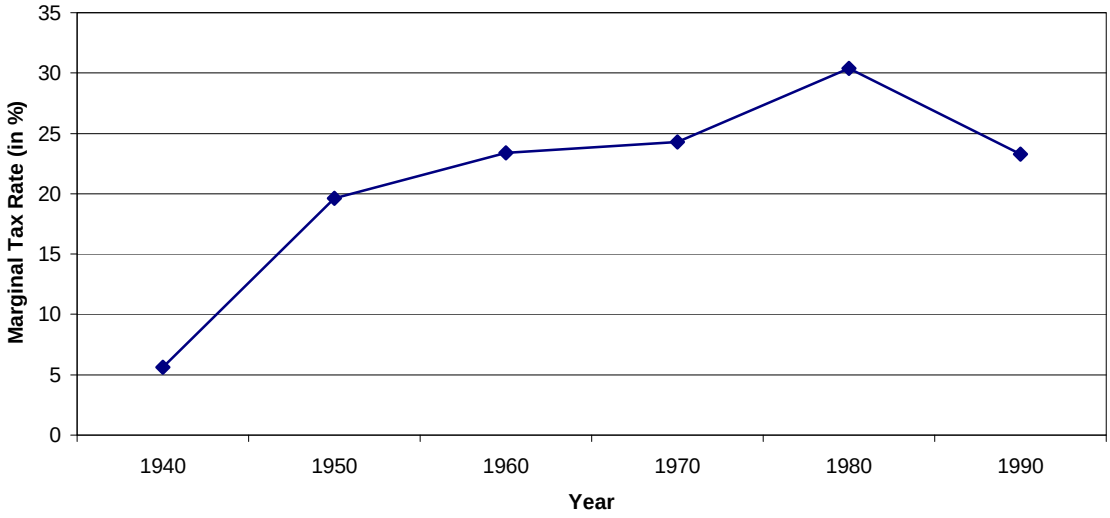


Figure 5: IRR for High School Completion (White and Black Men)

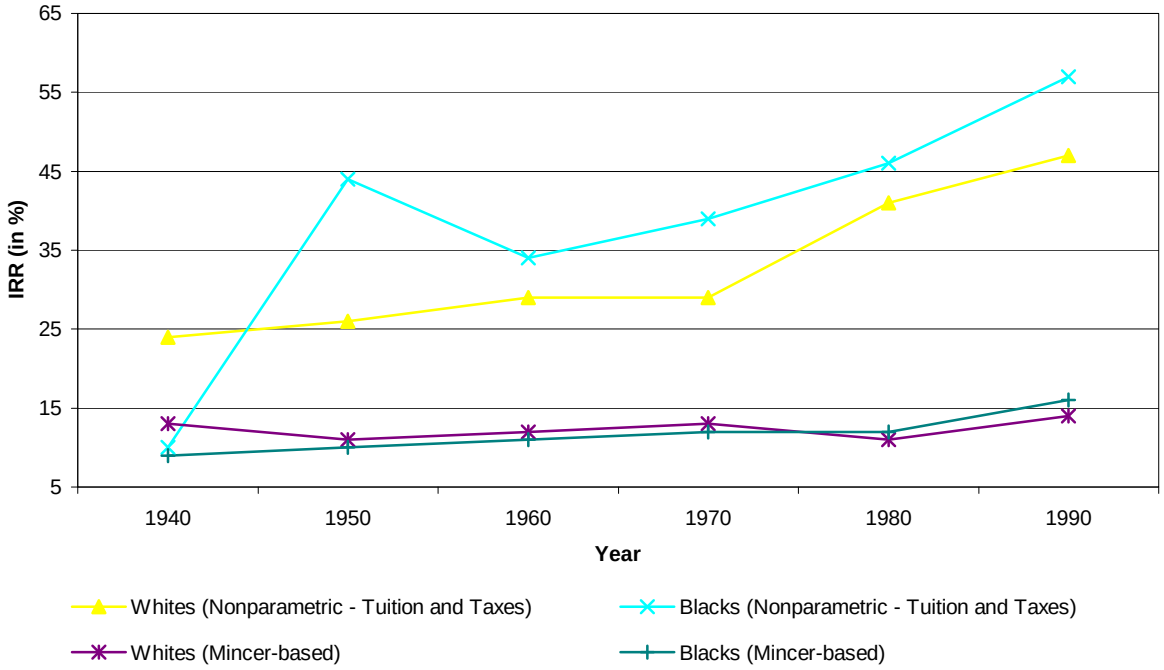


Figure 6: IRR for College Completion (White and Black Men)

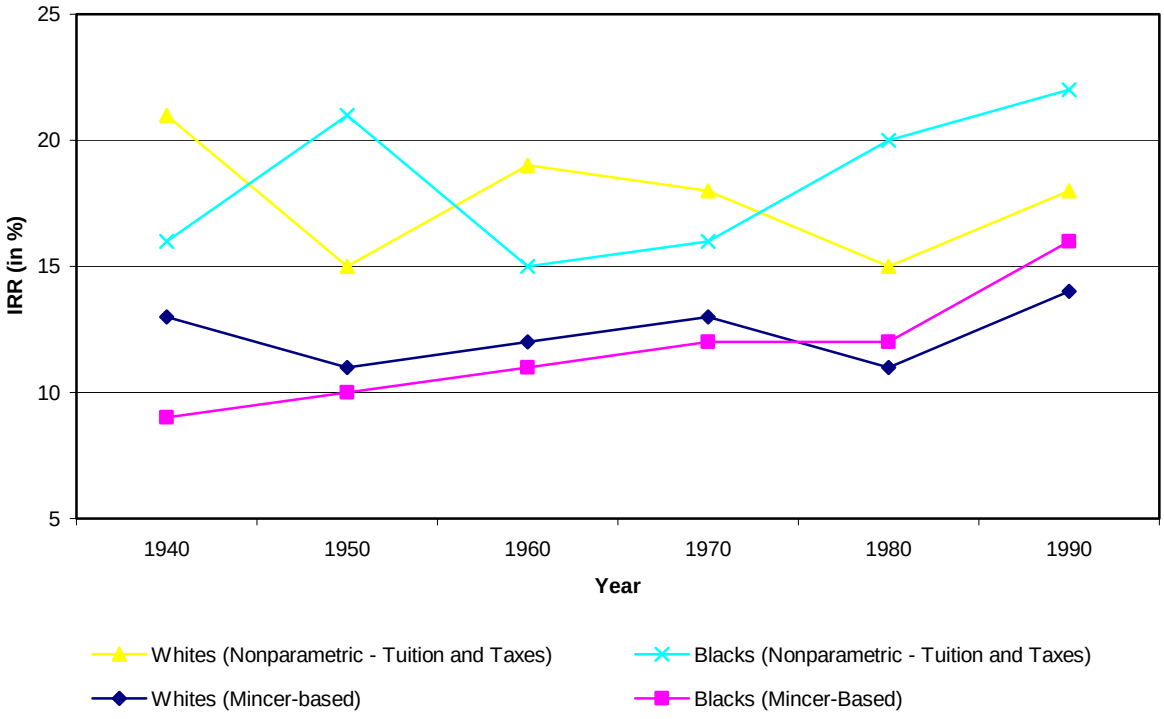


Figure 7a: IRR for 10 vs. 12 Years of Education for White Men (1964−2000 CPS)

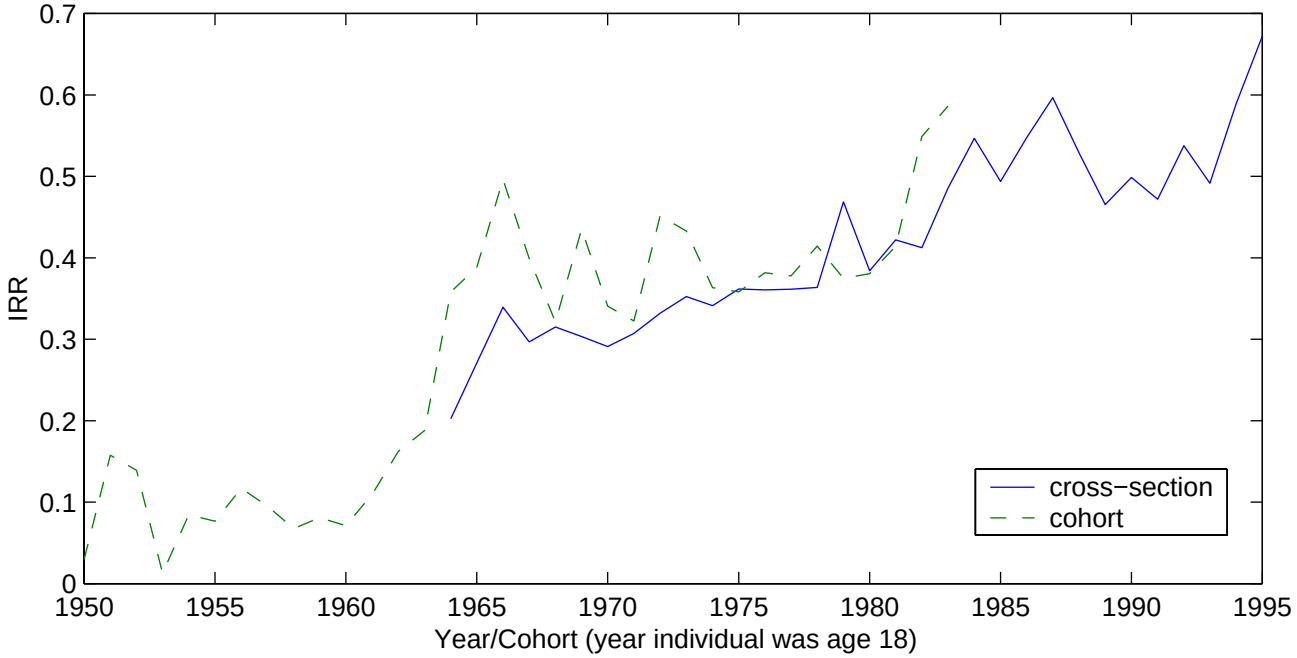


Figure 7b: IRR for 12 vs. 16 Years of Education for White Men (1964−2000 CPS)

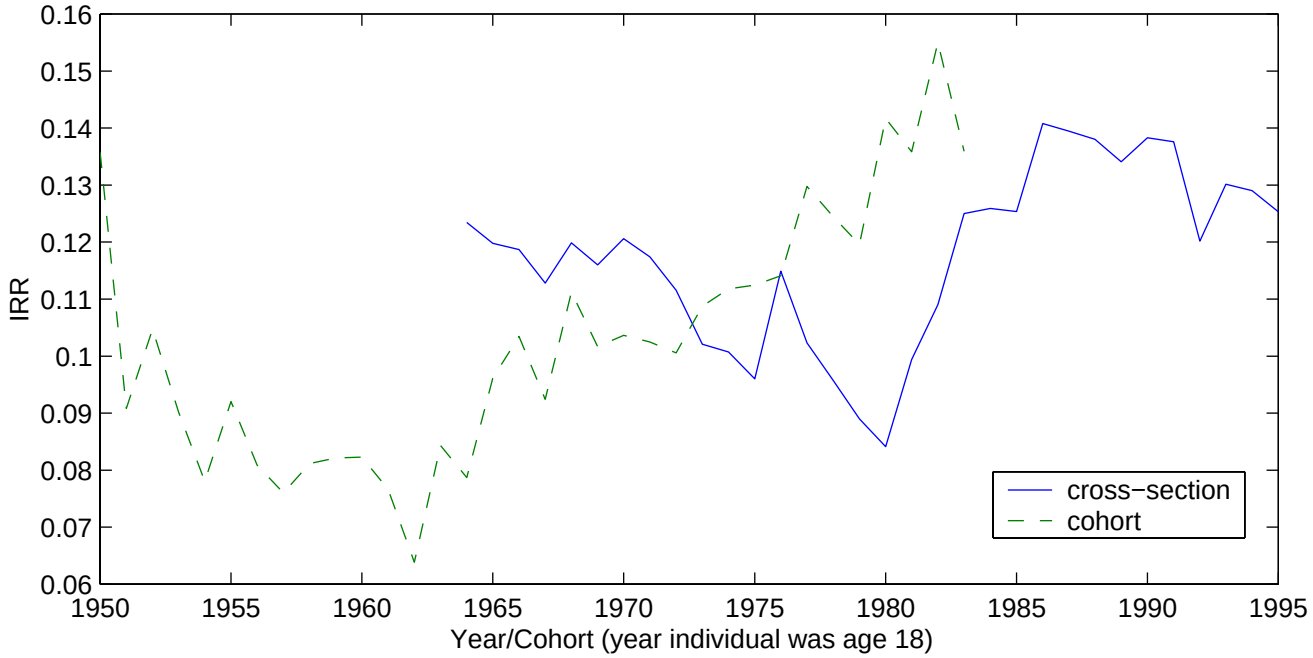


Figure 8: Three Crossings with Option Values

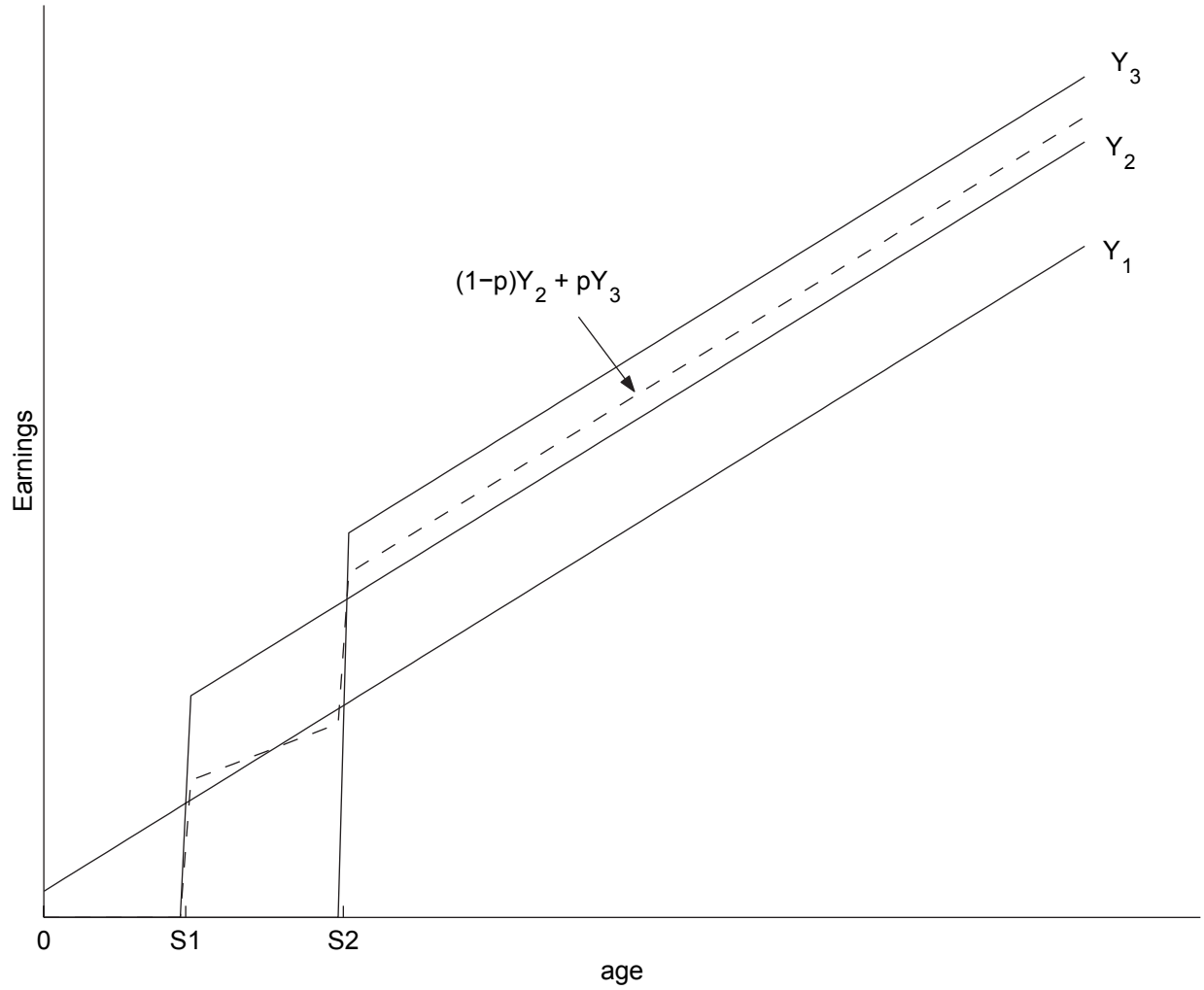


Figure 9
Density of Absolute Returns

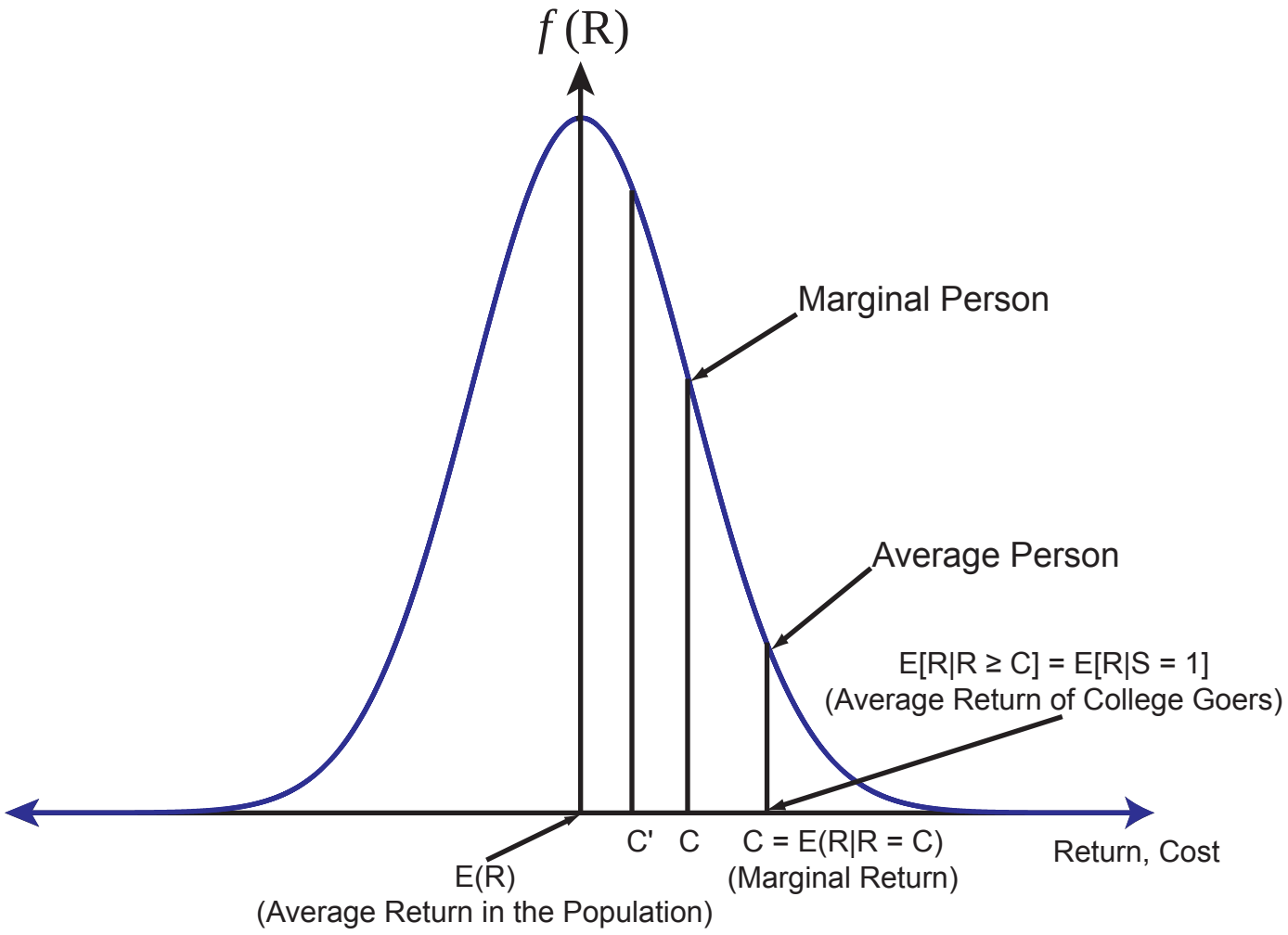


Figure 10a
Weights for the Marginal Treatment Effect for Different Parameters

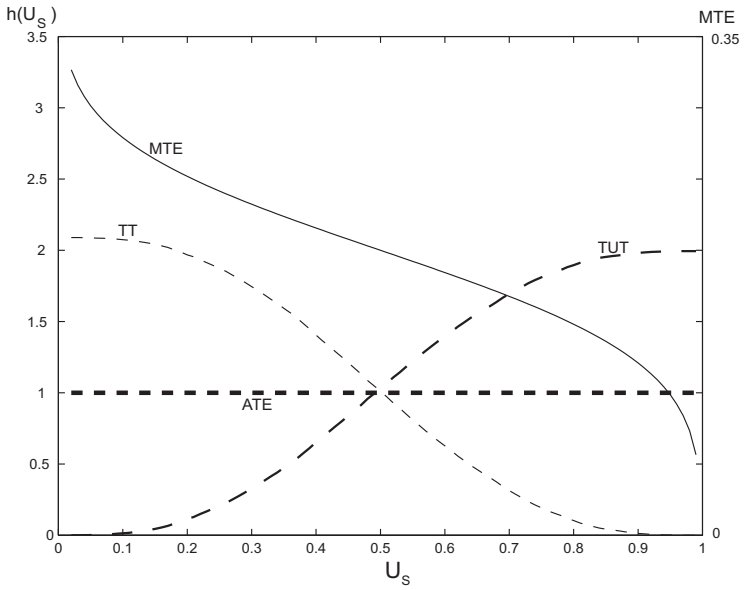
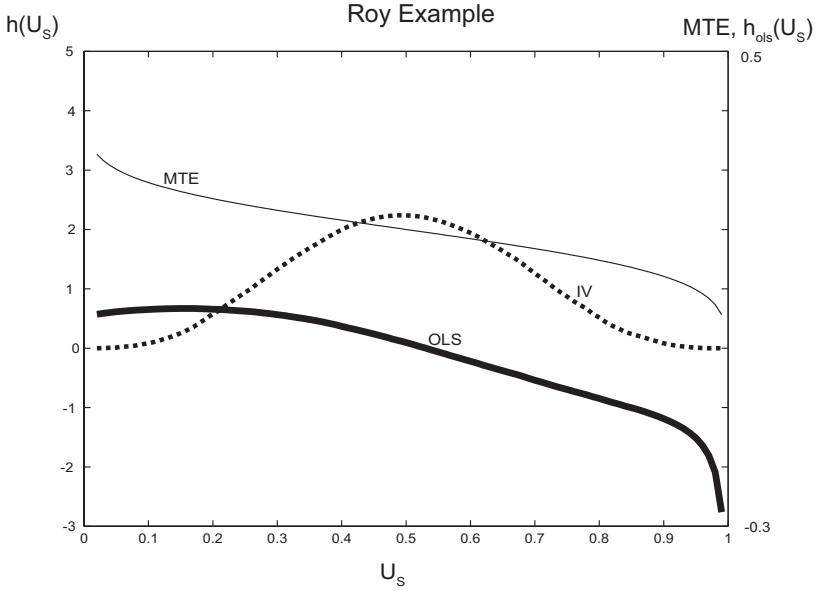


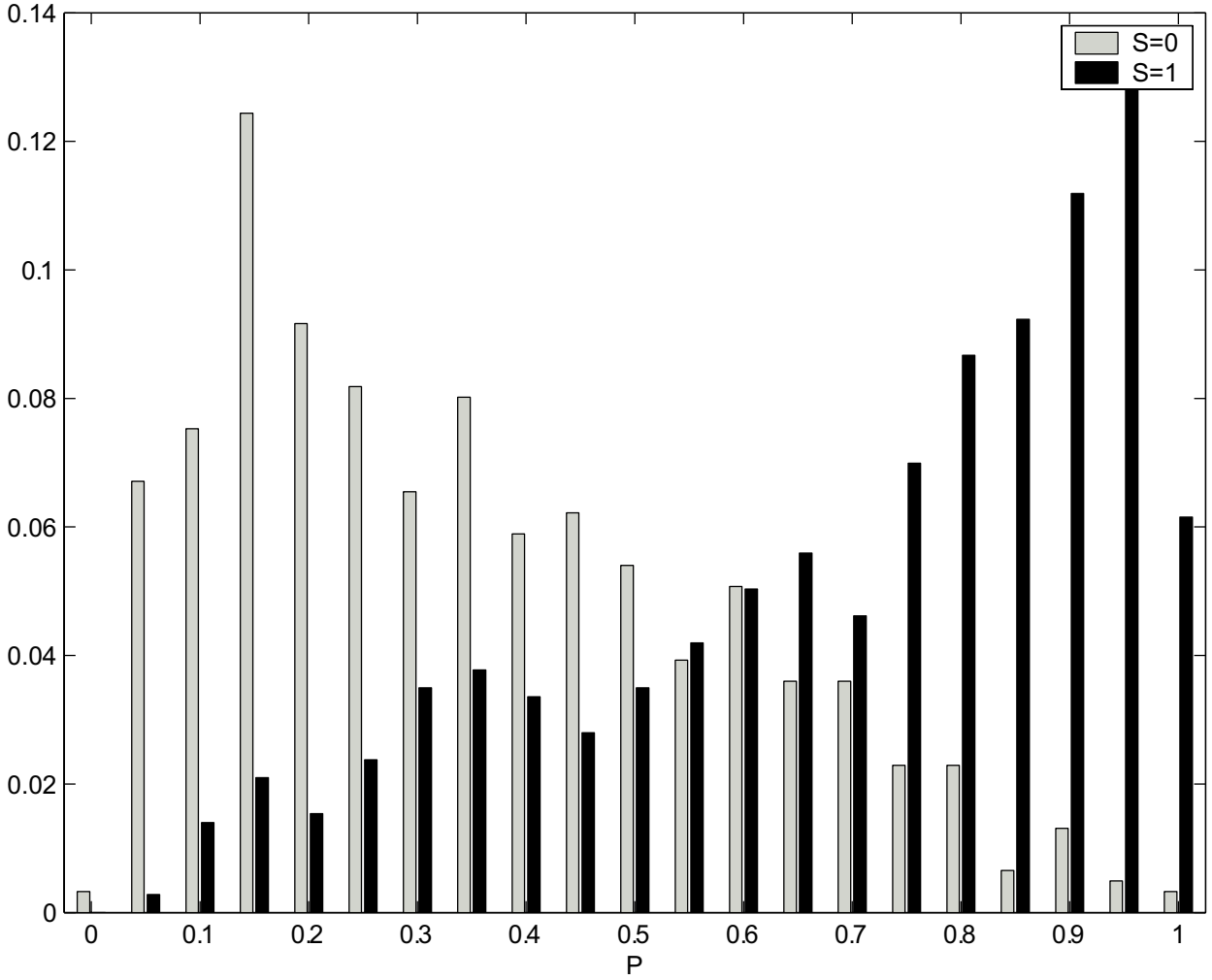
Figure 10b
Marginal Treatment Effect vs. Linear Instrumental Variables
and Ordinary Least Squares Weights
Roy Example



Source: Heckman and Vytlacil (2005b).

$$
\begin{aligned}
lnY_1 &= \alpha + \bar{\beta} + U_1 \\
lnY_0 &= \alpha + U_0 \\
S &= 1 \text{ if } Z - U_S > 0 \\
\alpha &= 0.67 \\
\bar{\beta} &= 0.2
\end{aligned}
\qquad
\begin{aligned}
U_1 &= \sigma_1 \varepsilon \\
U_0 &= \sigma_0 \varepsilon \\
U_S &= \sigma_S \varepsilon
\end{aligned}
\qquad
\begin{aligned}
\sigma_1 &= 0.012 \\
\sigma_0 &= -0.050 \\
\sigma_S &= -1 \\
\varepsilon &\sim N(0,1) \\
Z &\sim N(-0.0026, 0.2700)
\end{aligned}
$$

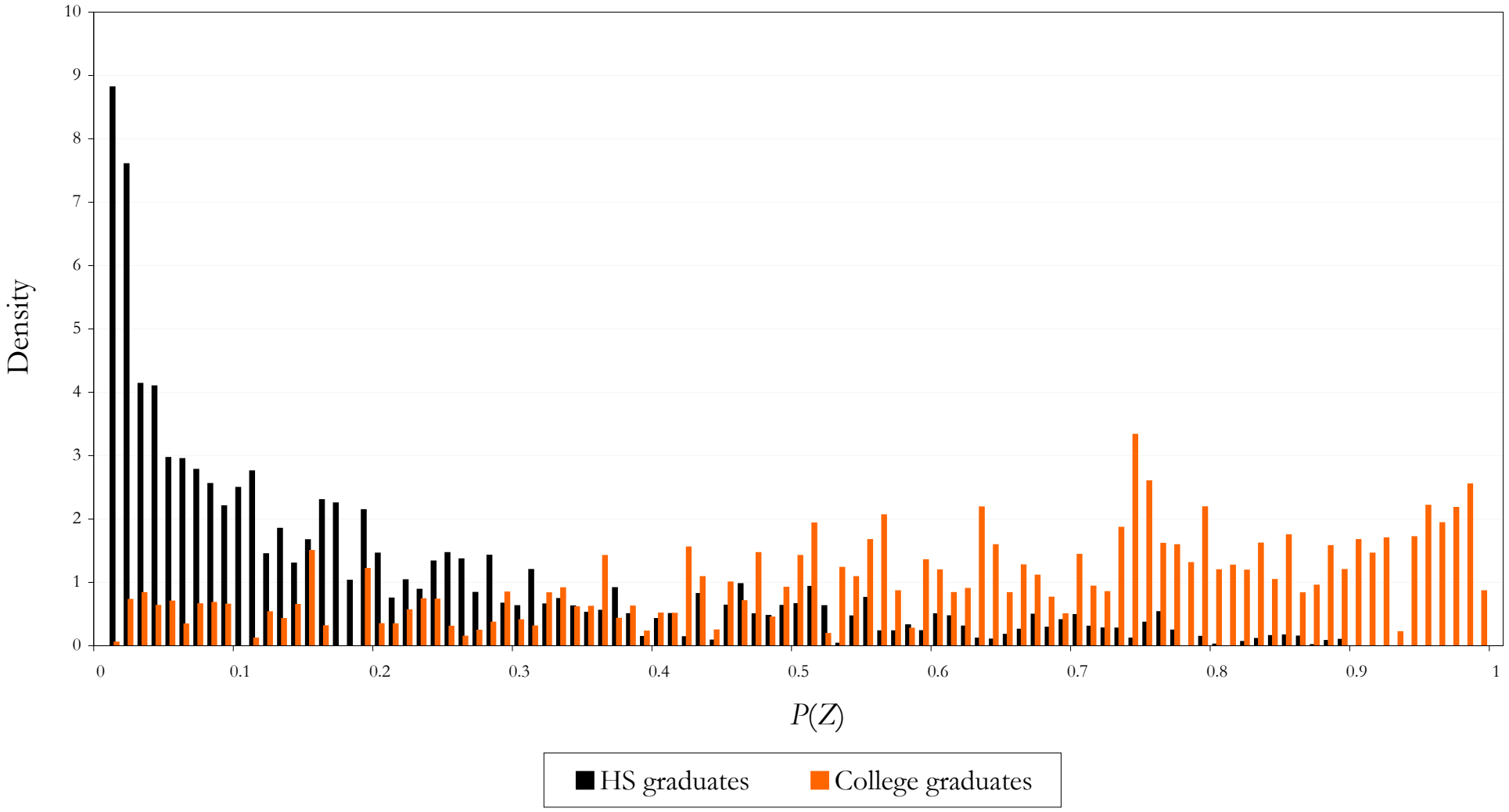
Figure 11 − Density of P given S=0 and S=1 (Estimated Probability of Enrolling in College)



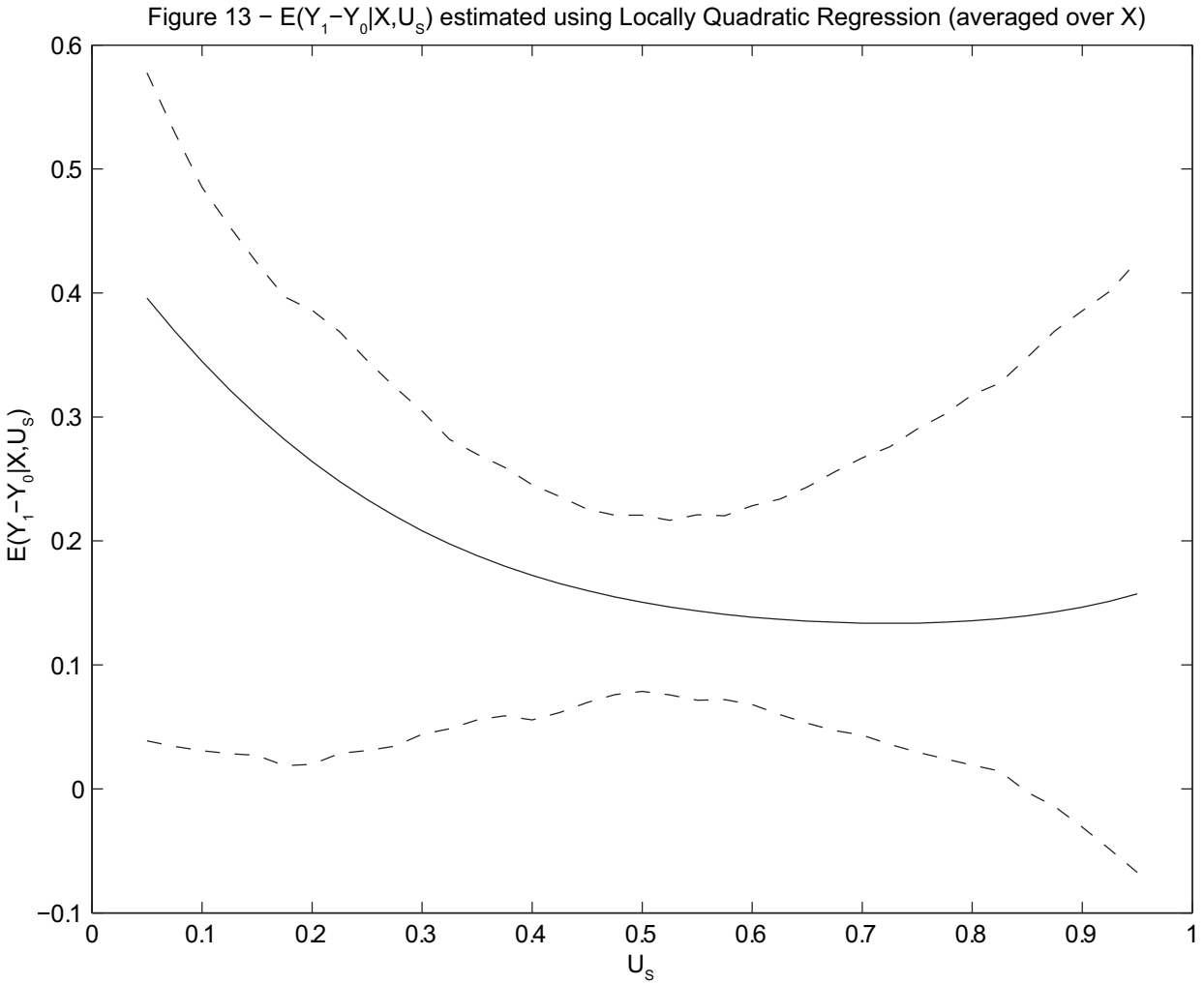
Note: P is the estimated probability of going to college. It is estimated from a logit regression of college attendance on corrected AFQT, father's education, mother's education, number of siblings, tuition, distance to college and local unemployment.
Source: Carneiro, Heckman and Vytlacil (2005)

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



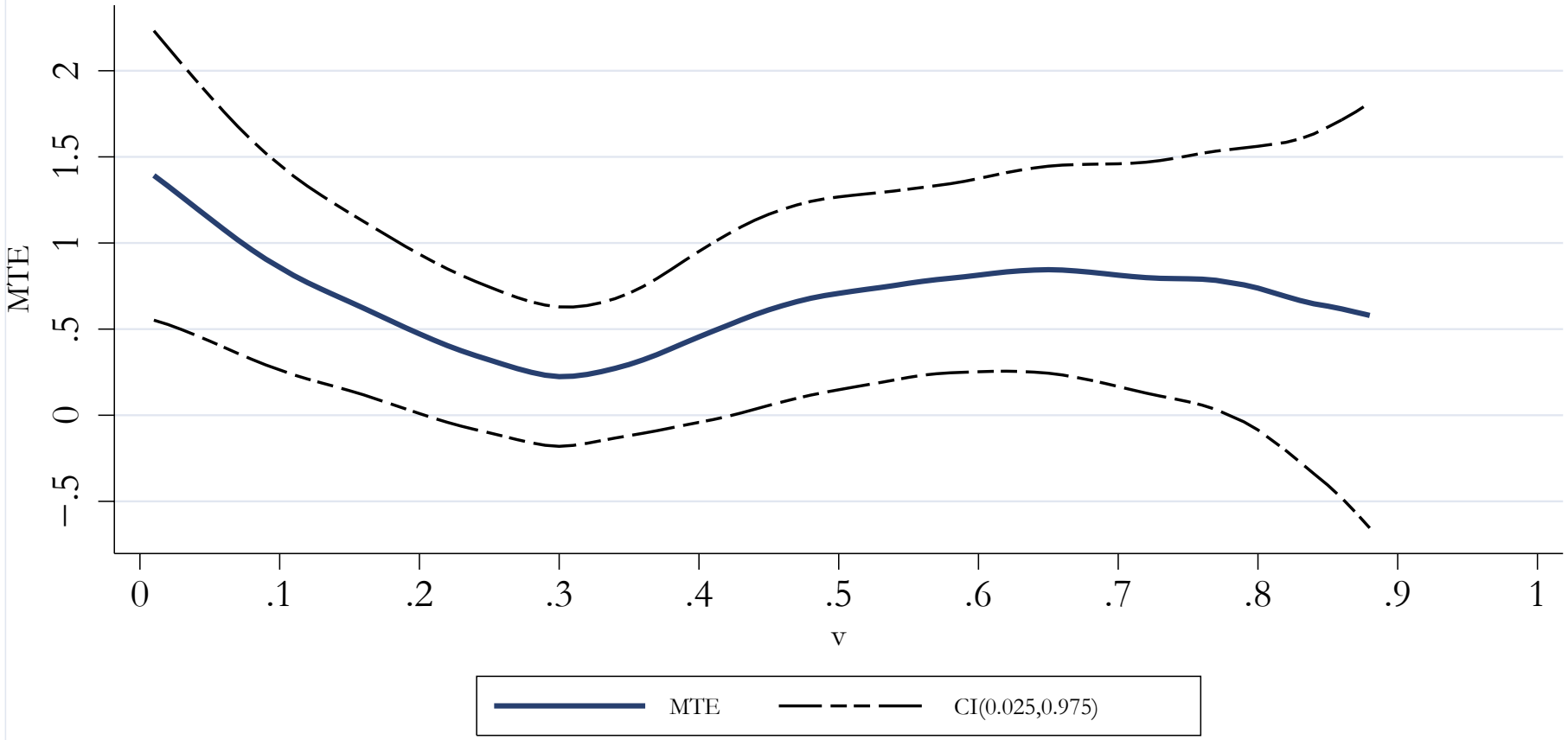
Source: Heckman, Urzua and Vytlacil (2004)



Note: To estimate the function in this figure ($E\left(Y_1 - Y_0 | X, U_S\right)$) we use a two step procedure. We first estimate $\mu_0\left(X\right)$ and $\mu_1\left(X\right)$ 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 $\mu_0\left(X\right) + P * \left[\mu_1\left(X\right) - \mu_0\left(X\right)\right]$ 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 $\mu_1\left(X\right) - \mu_0\left(X\right)$. $E(Y_1 - Y_0 | X, U_S)$ 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).
Source: Carneiro, Heckman and Vytlacil (2005)

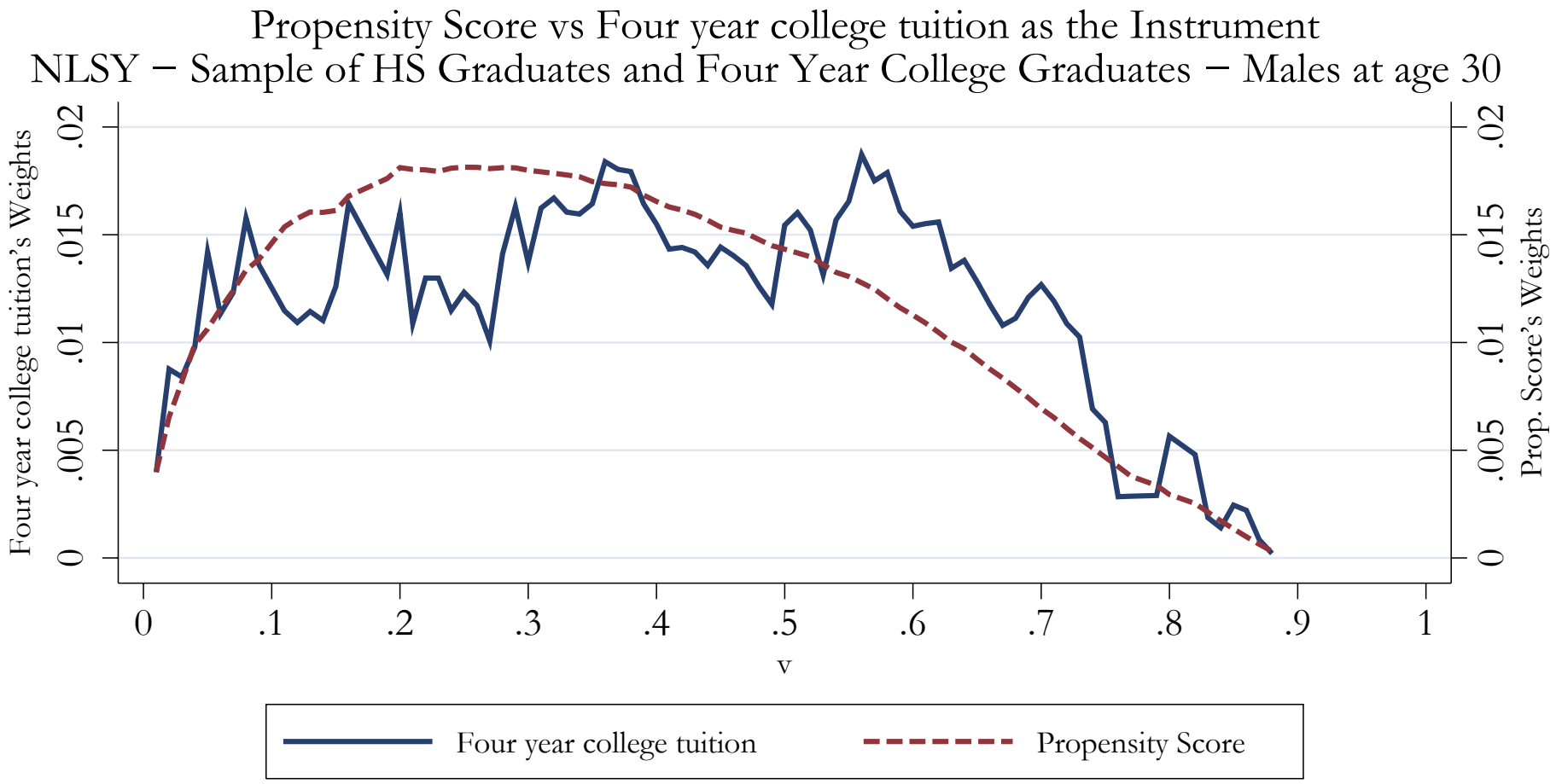
Figure 14. MTE with Confidence Interval

Sample of HS Graduates and Four Year College Graduates − Males at age 30 − Nonparametric



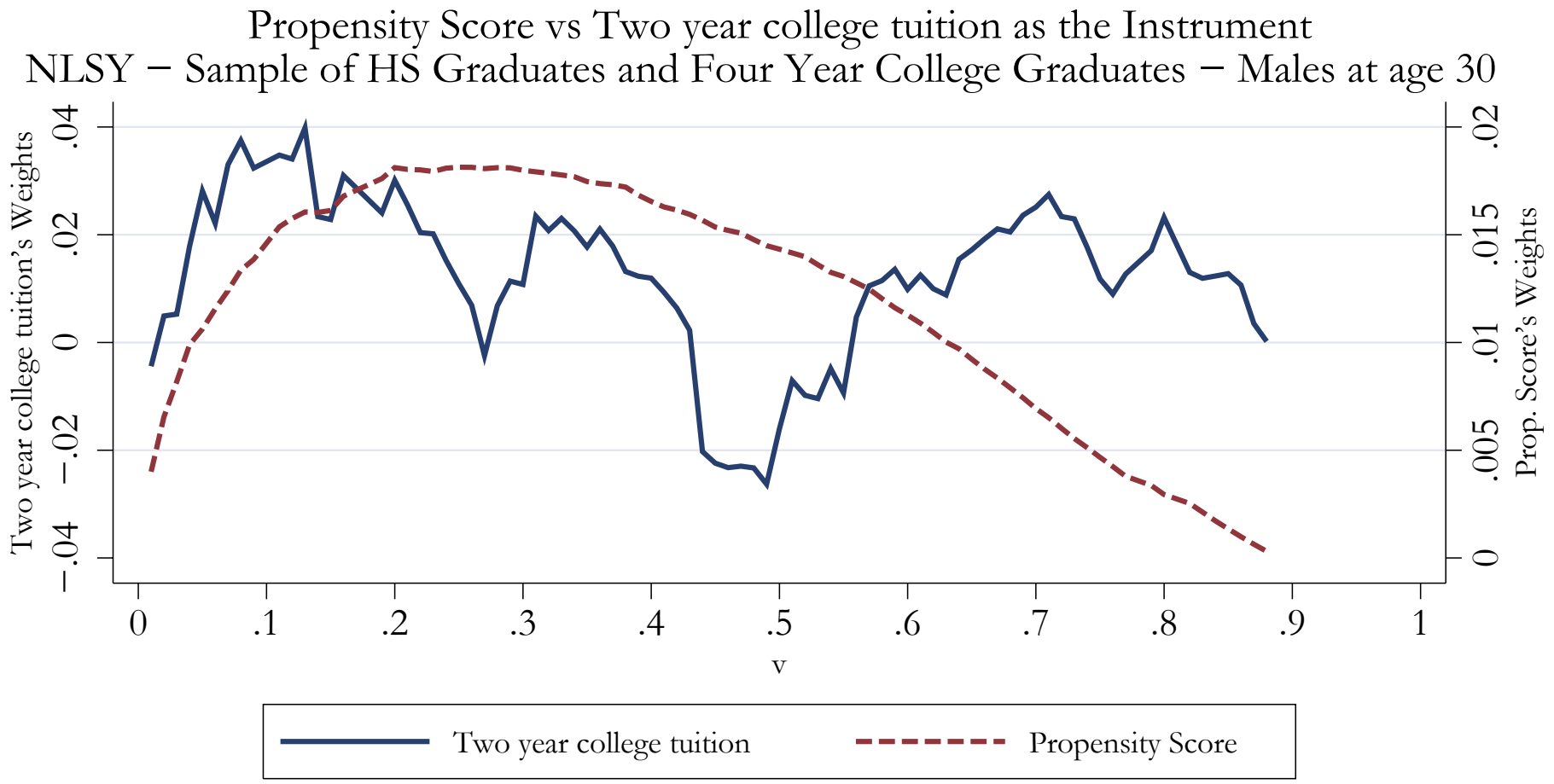
Source: Heckman, Urzua and Vytlacil (2004).
The dependent variable in the outcome equation is hourly earnings at age 30. The controls in the outcome equations are tenure, tenure squared, experience, corrected AFQT, black (dummy), hispanic (dummy), marital status, and years of schooling. Let D=0 denote dropout status, and D=1 denote GED status. The model for D (choice model) includes as controls the corrected AFQT, number of siblings, father's education, mother's 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.

Figure 15a. IV Weights



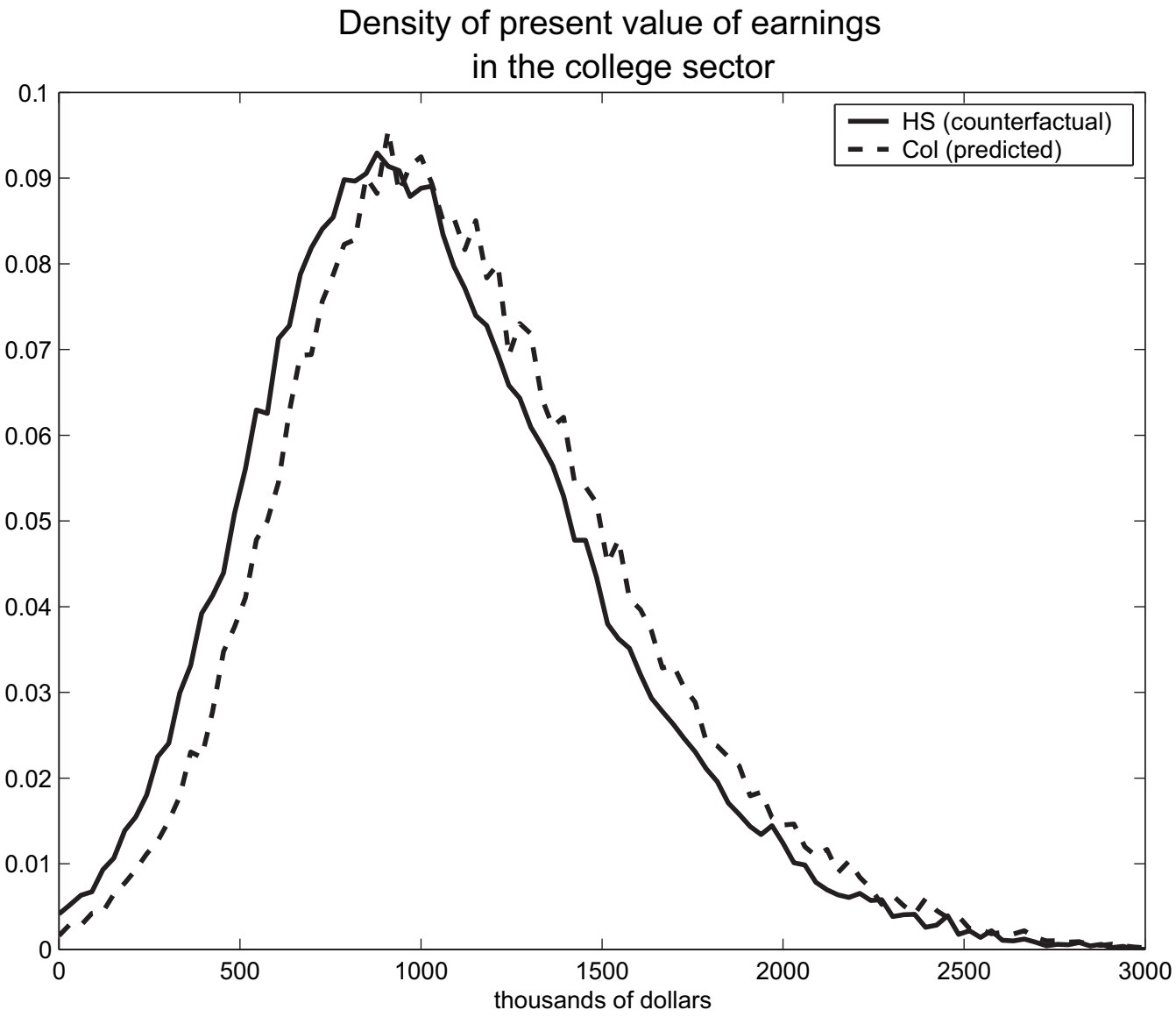
Source: Heckman, Urzua and Vytlacil (2004).
The dependent variable in the outcome equation is hourly earnings at age 30. The controls in the outcome equations are tenure, tenure squared, experience, corrected AFQT, black (dummy), hispanic (dummy), marital status, and years of schooling. Let D=0 denote dropout status, and D=1 denote GED status. The model for D (choice model) includes as controls the corrected AFQT, number of siblings, father's education, mother's 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.

Figure 15b. IV Weights



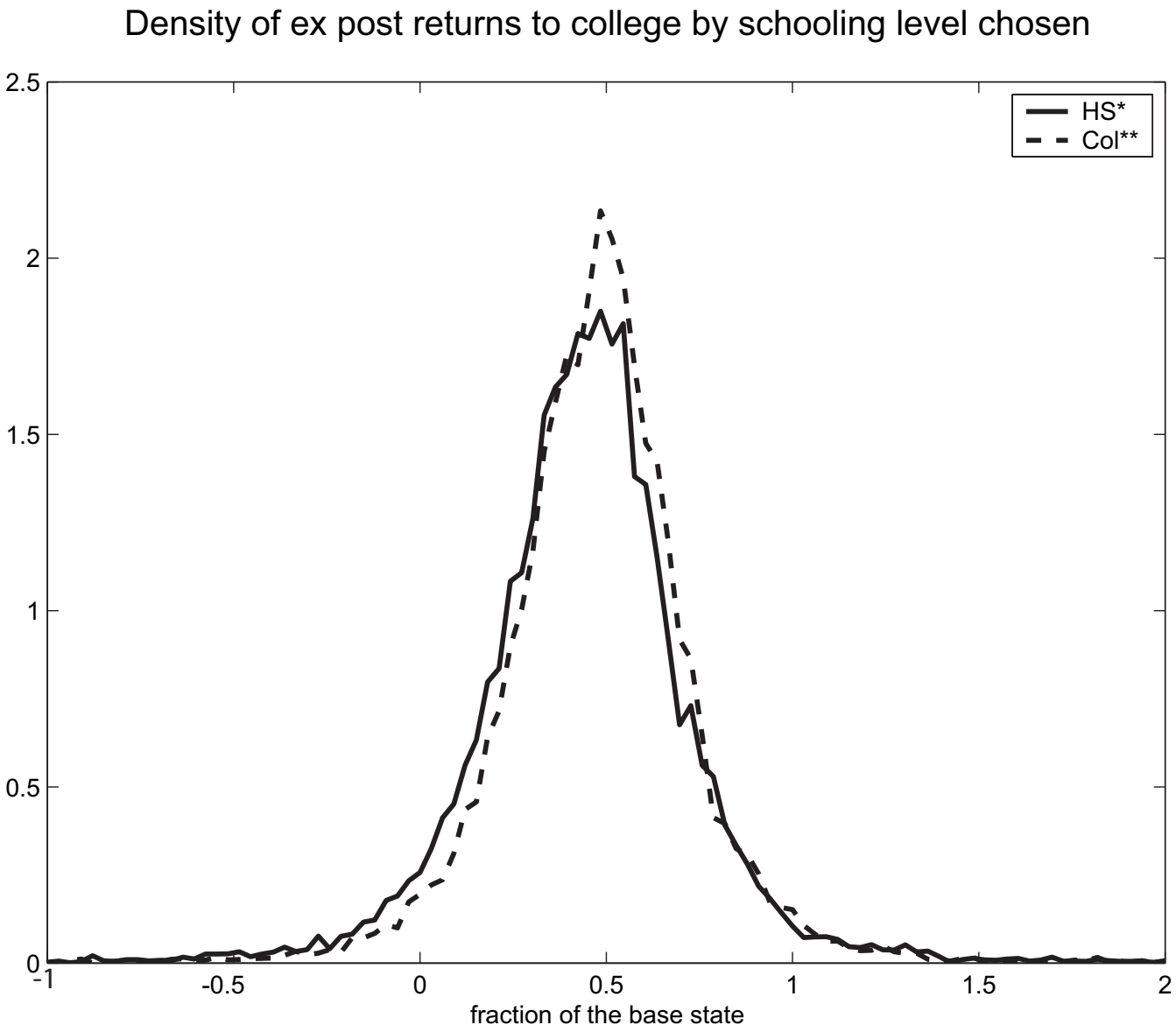
Source: Heckman, Urzua and Vytlacil (2004).
The dependent variable in the outcome equation is hourly earnings at age 30. The controls in the outcome equations are tenure, tenure squared, experience, corrected AFQT, black (dummy), hispanic (dummy), marital status, and years of schooling. Let D=0 denote dropout status, and D=1 denote GED status. The model for D (choice model) includes as controls the corrected AFQT, number of siblings, father's education, mother's 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.

Figure 16



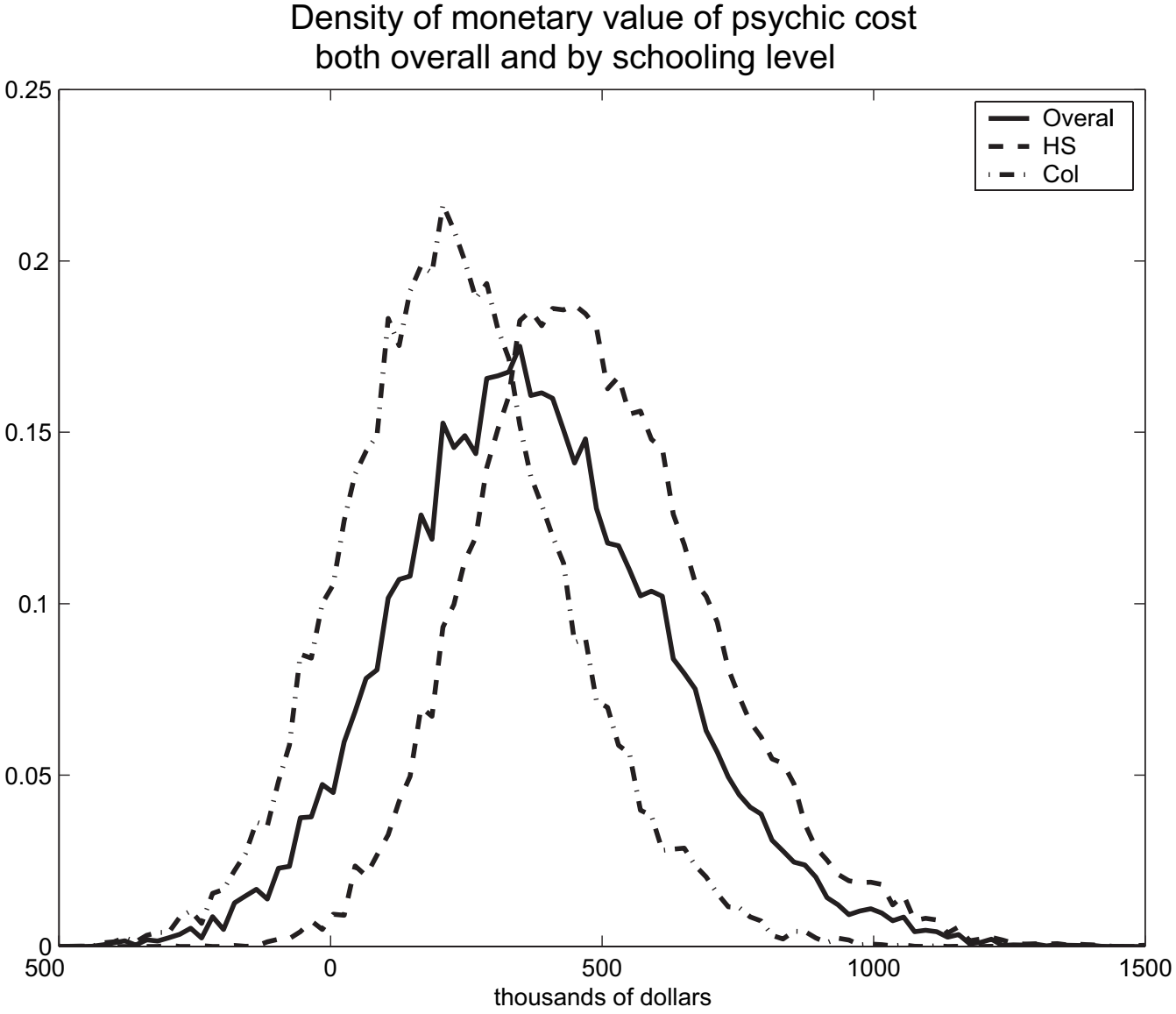
Let Y_1 denote present value of earnings (discounted at a 3% interest rate) in the college sector. Let $f(y_1)$ denote its density function. The dashed line plots the predicted Y_1 density conditioned on choosing college, that is, $f(y_1|S=1)$, while the solid line shows the counterfactual density function of Y_1 for those agents that are actually high school graduates, that is, $f(y_1|S=0)$.
Source: Cunha, Heckman and Navarro (2005c)

Figure 17



Let Y_0, Y_1 denote the present value of earnings in high school and college sectors, respectively. Define *ex post* returns to college as the ratio $R = \frac{(Y_1 - Y_0)}{Y_0}$. Let $f(r)$ denote the density function of the random variable R. The solid line is the density of *ex post* returns to college for high school graduates, that is, $f(r|S = 0)$. The dashed line is the density of *ex post* returns to college for college graduates, that is, $f(r|S = 1)$.
Source: Cunha, Heckman and Navarro (2005c)

Figure 18

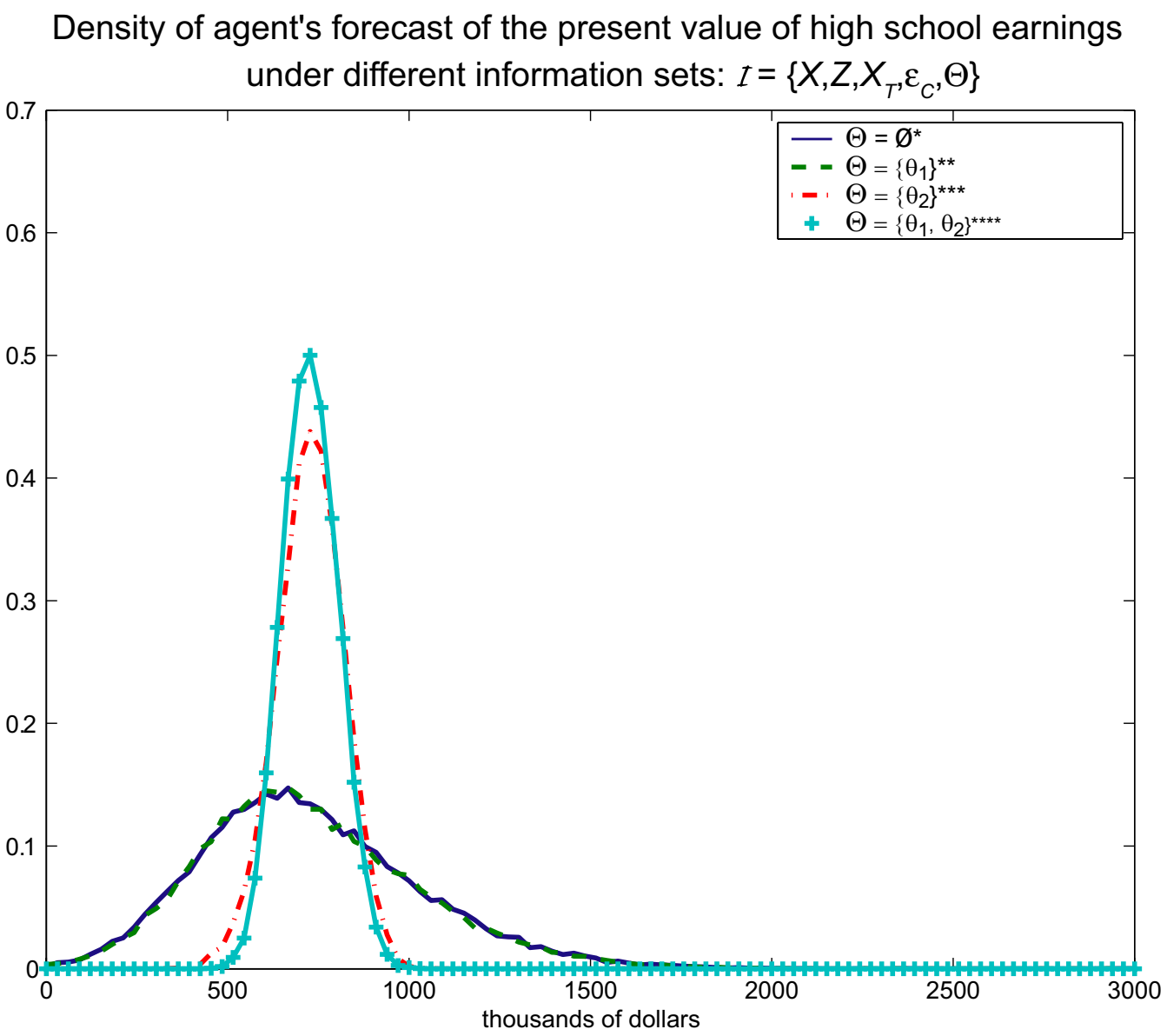


In this figure we plot the monetary value of psychic costs. Let C denote the monetary value of psychic costs. The monetary value of psychic costs is given by:

$$C = Z\gamma + \theta_1 \alpha_{C1} + \theta_2 \alpha_{C2} + \varepsilon_C$$

The contribution of ability to the costs of attending college, in monetary value, is $\theta_1 \alpha_{C1}$.
Source: Cunha, Heckman and Navarro (2005c)

Figure 19



Let Y_0 denote the agent's forecast of present value of earnings in the high school sector. These are formed over the whole population, not just the subpopulation who go to high school. We assume that agents know all coefficients. Let $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$ denote the agent's information set. Let $f(y_0|\mathcal{I})$ denote the density of the agent's forecast of present value of earnings in high school conditioned on the information set $\mathcal{I}$. Then:

*Plot of $f(y_0|\mathcal{I})$ under no element θ in the information set, i.e., $\Theta = \emptyset$.

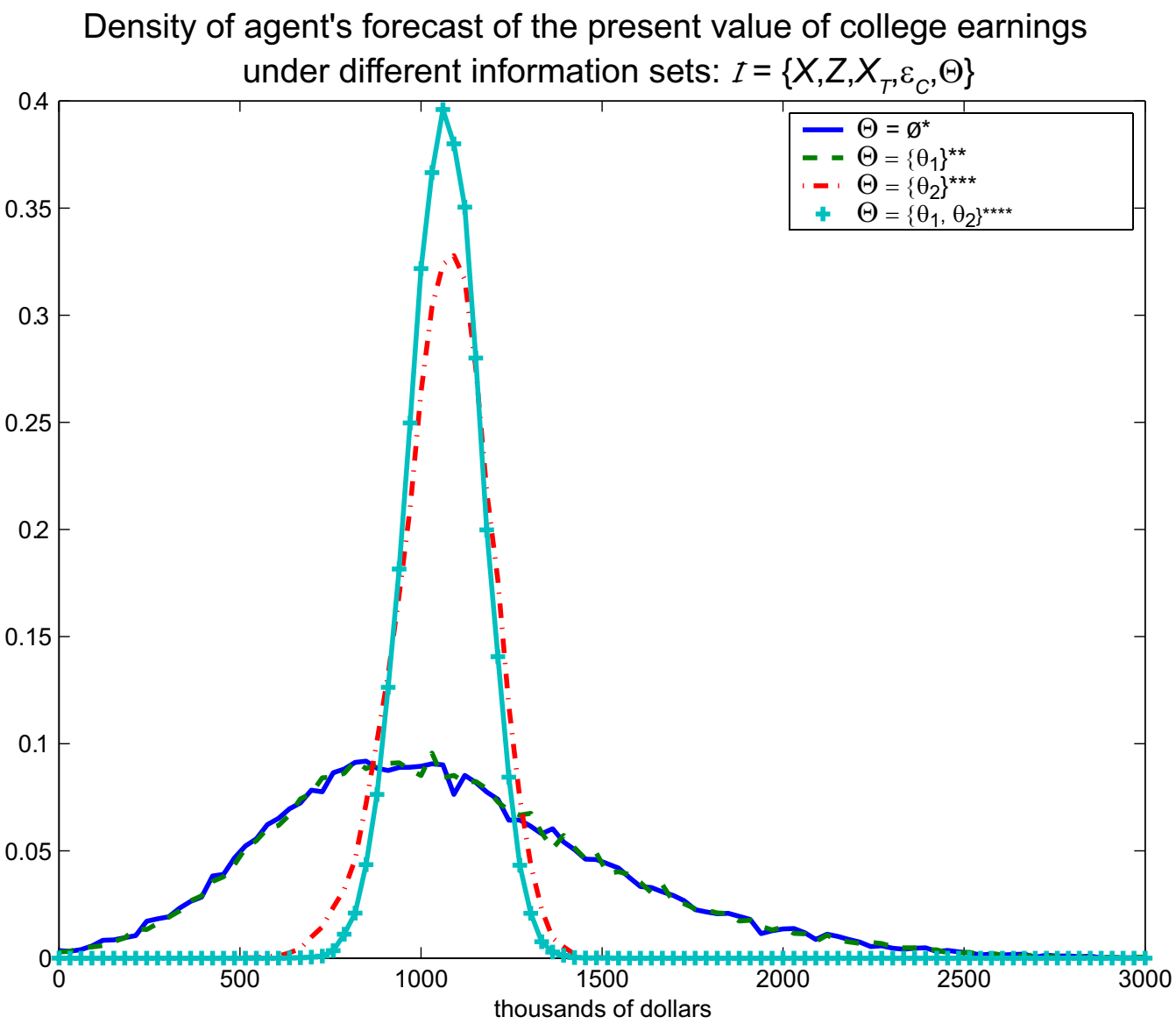
**Plot of $f(y_0|\mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.

***Plot of $f(y_0|\mathcal{I})$ when only factor 2 is in the information set, i.e.,$\Theta = \{\theta_2\}$.

****Plot of $f(y_0|\mathcal{I})$ when both factors are in the information set, i.e., $\Theta = \{\theta_1, \theta_2\}$.

Source: Cunha, Heckman and Navarro (2005c)

Figure 20



Let Y_1 denote the agent's forecast of present value of earnings in the college sector. These are formed over the whole population, not just the subpopulation who go to college. We assume that agents know all coefficients. Let $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$ denote the agent's information set. Let $f(y_1|\mathcal{I})$ denote the density of the agent's forecast of present value of earnings in college conditioned on the information set $\mathcal{I}$. Then:

*Plot of $f(y_1|\mathcal{I})$ under no element θ in the information set, i.e., $\Theta = \emptyset$.

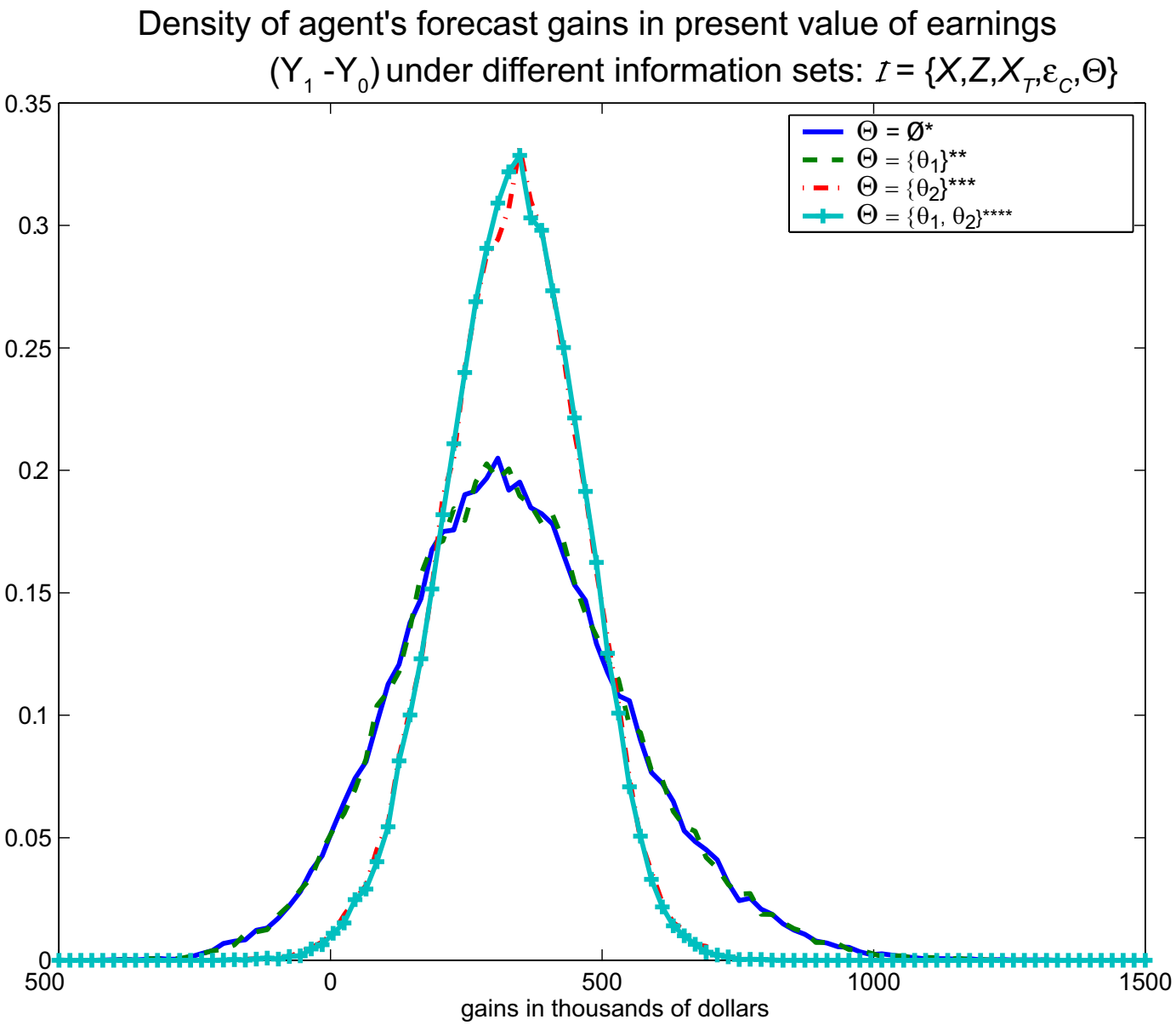
**Plot of $f(y_1|\mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.

***Plot of $f(y_1|\mathcal{I})$ when only factor 2 is in the information set, i.e.,$\Theta = \{\theta_2\}$.

****Plot of $f(y_1|\mathcal{I})$ when both factors are in the information set, i.e., $\Theta = \{\theta_1, \theta_2\}$.

Source: Cunha, Heckman and Navarro (2005c)

Figure 21



Let Y_0, Y_1 denote the agent's forecast of present value of earnings in the high school and college sectors, respectively. We define the difference in present value of earnings as $\Delta = Y_1 - Y_0$. We assume that agents know all coefficients. Let $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$, $f(\Delta|\mathcal{I})$ denote the agent's information set and the density of the agent's forecast of gains in present value of earnings in choosing college, conditioned on the information set $\mathcal{I}$. These are defined over the entire population. Then:
*Plot of $f(\Delta|\mathcal{I})$ under no element θ in the information set, i.e., $\Theta = \emptyset$.
**Plot of $f(\Delta|\mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.
***Plot of $f(\Delta|\mathcal{I})$ when only factor 2 is in the information set, i.e.,$\Theta = \{\theta_2\}$.
****Plot of $f(\Delta|\mathcal{I})$ when both factors are in the information set, i.e., $\Theta = \{\theta_1, \theta_2\}$.
Source: Cunha, Heckman and Navarro (2005c)