Cunha, Flavio; Heckman, James Joseph; Navarro, Salvador

Working Paper

Separating Uncertainty from Heterogeneity in Life Cycle Earnings

IZA Discussion Papers, No. 1437

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Cunha, Flavio; Heckman, James Joseph; Navarro, Salvador (2004) : Separating Uncertainty from Heterogeneity in Life Cycle Earnings, IZA Discussion Papers, No. 1437, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/20736

DISCUSSION PAPER SERIES

IZA DP No. 1437

Separating Uncertainty from Heterogeneity in Life Cycle Earnings

Flavio Cunha
James Heckman
Salvador Navarro

December 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Separating Uncertainty from Heterogeneity in Life Cycle Earnings

Flavio Cunha
University of Chicago

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

Salvador Navarro
University of Chicago

Discussion Paper No. 1437
December 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 1437
December 2004

ABSTRACT

Separating Uncertainty from Heterogeneity in Life Cycle Earnings

This paper develops and applies a method for decomposing cross section variability of earnings into components that are forecastable at the time students decide to go to college (heterogeneity) and components that are unforecastable. About 60% of variability in returns to schooling is forecastable. This has important implications for using measured variability to price risk and predict college attendance.

JEL Classification: C33, D84, I21

Keywords: uncertainty, lifecycle earnings, schooling, heterogeneity, counterfactuals

Corresponding author:

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

1 Introduction

This lecture commemorates the 100th anniversary of the birth of Sir John Hicks. In most of his work, Hicks relied on the Marshallian fiction of a representative agent and abstracted from heterogeneity and variability among people and firms. Economic theory now recognizes the importance of accounting for heterogeneity among agents in explaining a variety of phenomena. See the survey in Browning *et al.* (1999). A major discovery of microeconometrics is that diversity among agents is a central feature of economic life (see Heckman, 2001). While Hicks generally ignored heterogeneity, he did discuss uncertainty. The distinction between *ex ante* and *ex post* income played a central role in his analysis of economic dynamics (see Hicks, 1946, p.178). It is featured in our analysis.

This paper develops and implements a method for estimating the importance of uncertainty about lifetime earnings facing agents at the stage of their life cycles when they make their college-going decisions. We estimate what components of measured lifetime income variability among persons are due to uncertainty realized after that stage and discuss what assumptions must be maintained to identify the distributions of these components. In accomplishing this task, we distinguish unobservables from the point of view of the econometrician from unobservables from the point of view of the agents being studied. We distinguish components of outcome variability that are forecastable and acted on at a given stage of the life cycle from unpredictable components. If agents act on (make choices based on) all forecastable information, under the conditions specified in this paper, we can estimate components of intrinsic uncertainty and distinguish them from components of forecastable uncertainty. Using the tools presented here, analysts can determine how much of lifetime

earnings variability or inequality is forecastable at a given age and how much is unforecastable 'luck.' With concavity in utility and lack of full insurance, at the same level of mean income, the greater the fraction of variability in lifetime incomes that is unforecastable, the lower the welfare of agents. Like Hicks, we distinguish *ex ante* returns from *ex post* returns.

We build on, and extend, methods developed in Carneiro *et al.* (2003) who separate earnings heterogeneity (defined here as information about future earnings known to agents and acted on in their choices) from unforecastable (at the date choices are made) uncertainty. They assume an environment of complete autarky. In this paper, we consider a complete markets environment. A companion paper, Cunha *et al.* (2004), considers an environment with partial insurance of the type analyzed by Aiyagari (1994) and Laitner (1992).

A major theoretical issue discussed in this paper is the difficulty in separately identifying the market structure facing an agent from the agent's information set. We develop methods for distinguishing components of future outcomes that are both forecastable and are acted on from those components that cannot be acted on. What can be acted on and the magnitude of the effects of the actions depends upon the market structure facing agents and their preferences.

A major empirical finding reported in all three of our papers is that across a variety of market environments and for different assumptions about and estimates of risk aversion, a substantial part of the variability in the *ex post* returns to schooling is predictable and acted on by agents. Variability cannot be equated with uncertainty and this has important empirical consequences.

The plan of the rest of this paper is as follows. Section 2 states the problem of distinguishing between predictable earnings heterogeneity and unpredictable uncertainty for a

specified market environment and presents the empirical strategy used in this paper. Section 3 motivates the econometric method we use. This part of the paper is an intuitive summary of the methods formally developed in Carneiro *et al.* (2003) and our extensions of it. Section 4 discusses the fundamental problem of separating preferences from market structures and information. Section 5 presents our empirical analysis and simulations of the model and discusses the implications of the findings. Section 6 concludes. Two appendices describe our approach to identification and how we pool data sets to create synthetic life cycles. A third appendix posted at http://jenni.uchicago.edu/Hicks2004/ describes our data.

2 Distinguishing between heterogeneity and uncertainty

In the literature on earnings dynamics (*e.g.*, Lillard and Willis, 1978), it is common to estimate an earnings equation of the sort

$$Y_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + S_i\tau + v_{i,t}, \tag{1}$$

where $Y_{i,t}, \boldsymbol{X}_{i,t}, S_i, v_{i,t}$ 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. We use bold characters to denote vectors and distinguish them from scalars. 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.

Often the error term $v_{i,t}$ is decomposed into two or more components. For example, it is

common to specify that

$$v_{i,t} = \phi_i + \varepsilon_{i,t}. \tag{2}$$

The term ϕ_i is a person-specific effect. The error term $\varepsilon_{i,t}$ is generally assumed to follow an ARMA (p, q) process (see, *e.g.*, MaCurdy, 1982) such as $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + m_{i,t}$, where $m_{i,t}$ is a mean zero innovation independent of $\boldsymbol{X}_{i,t}$ and the other error components. The components $\boldsymbol{X}_{i,t}, \phi_i$, and $\varepsilon_{i,t}$ 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 measured from a given age—what Jencks *et al.* (1972) call 'luck.'

An alternative specification of the error process postulates a factor structure for earnings,

$$v_{i,t} = \boldsymbol{\theta}_i \boldsymbol{\alpha}_t + \delta_{i,t}, \tag{3}$$

where $\boldsymbol{\theta}_i$ is a vector of skills (*e.g.*, ability, initial human capital, motivation, and the like), $\boldsymbol{\alpha}_t$ is a vector of skill prices, and the $\delta_{i,t}$ are mutually independent mean zero shocks independent of $\boldsymbol{\theta}_i$. See Hause (1978) and Heckman and Scheinkman (1987) for analysis of such a model. Any process in the form of equation (2) can be written in terms of (3). The latter specification is more directly interpretable as a pricing equation than (2) and is a natural starting point for human capital analyses. It is the one used in this paper.

Depending on the available market arrangements for coping with risk, the predictable components of $v_{i,t}$ will have a different effect on choices and economic welfare than the un-

predictable components, if people are risk averse and cannot fully insure against uncertainty. Statistical decompositions based on (1), (2), and (3) or versions of them describe *ex post* variability but tell us nothing about which components of (1) or (3) are forecastable by agents *ex ante*. Is ϕ_i unknown to the agent? $\varepsilon_{i,t}$? Or $\phi_i + \varepsilon_{i,t}$? Or $m_{i,t}$? In representation (3), the entire vector $\boldsymbol{\theta}_i$, components of the $\boldsymbol{\theta}_i$, the $\delta_{i,t}$, or all of these may or may not be known to the agent at the time schooling choices are made.

The methodology presented in this paper 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. The essential idea of the method can be illustrated in the case of educational choice, the problem we study in our empirical work. In order to choose between high school and college, say at age 19, agents forecast future earnings (and other returns and costs) for each schooling level. Using information about educational choices at age 19, 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 at age 19. 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 the wealth maximizing decision, it is not necessary to know agent preferences to decompose realized earnings outcomes in this fashion. Our method uses choice information to extract *ex ante* or forecast components of earnings and to distinguish them from realized earnings. The difference between forecast and realized earnings allows us to identify the distributions of the components of uncertainty

facing agents at the time they make their schooling decisions.

To be more precise, consider a version of the generalized Roy (1951) economy with two sectors.[1] 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).

$Y_{1,i}$ is the *ex post* present value of earnings in the college sector, discounted over horizon T for a person choosing at a fixed age, assumed for convenience to be zero,

$$Y_{1,i} = \sum_{t=0}^{T} \frac{Y_{1,i,t}}{(1+r)^t},$$

and $Y_{0,i}$ is the *ex post* present value of earnings in the high-school sector at age zero,

$$Y_{0,i} = \sum_{t=0}^{T} \frac{Y_{0,i,t}}{(1+r)^t},$$

where r is the one-period risk-free interest rate. $Y_{1,i}$ 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

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

the other of these streams. This partial observability creates a fundamental identification problem which we address in this paper.

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

$$S_i = \begin{cases} 1, \text{ if } E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right) \geq 0 \\ 0, \text{ otherwise.}^2 \end{cases} \tag{4}$$

Under perfect foresight, the postulated information set would include $Y_{1,i}, Y_{0,i}$, and C_i. In 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. Under either set of assumptions, a separation theorem governs choices. Agents maximize expected wealth independently of how they consume it.

The decision rule is more complicated in the absence of full risk diversifiability and depends on the curvature of utility functions, the availability of markets to spread risk, and possibilities for storage. (See Cunha *et al.*, 2004, and Navarro, 2004, for a more extensive discussion.) In 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

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

paper we use a model with a simple market setting to motivate the identification analysis of a more general environment we analyze elsewhere (and is analyzed in Carneiro *et al.*, 2003).

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. To fix ideas, we start the analysis discussing identification of $\mathcal{I}_{i,0}$ for each person but in our empirical work we only partially identify person-specific $\mathcal{I}_{i,0}$ and instead identify the distributions of the remaining unobserved components.

To motivate the objectives of our analysis we offer the following heuristic discussion. We seek to decompose the 'returns coefficient' in an earnings-schooling model into components that are known at the time schooling choices are made and the components that are not known. For simplicity we assume that, for person i, returns are the same at all levels of schooling. Write discounted lifetime earnings of person i as

$$Y_i = \rho_0 + \rho_{1,i} S_i + J_i, \tag{5}$$

where $\rho_{1,i}$ is the person-specific *ex post* return, S_i is years of schooling, and J_i is a mean zero unobservable. We seek to decompose $\rho_{1,i}$ into two components $\rho_{1,i} = \eta_i + \nu_i$, 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, $S_i = \lambda(\eta_i, \boldsymbol{Z}_i, \tau_i)$, where the $\boldsymbol{Z}_i$ are other observed determinants of schooling and τ_i represents additional factors unobserved by the analyst

but known to the agent. We seek to determine what components of *ex post* school lifetime earnings Y_i enter the schooling choice equation.

If η_i is known, it enters λ. 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 us to identify components known to agents making their schooling decisions. Even if we cannot identify $\rho_{1,i}$, η_i, or ν_i for each person, under conditions specified in this paper we can identify their distributions.

Suppose that the model for schooling can be written in linear in parameters form:

$$S_i = \lambda_0 + \lambda_1 \eta_i + \lambda_2 \nu_i + \boldsymbol{\lambda}_3 \boldsymbol{Z}_i + \tau_i, \tag{6}$$

where τ_i has mean zero and is independent of $\boldsymbol{Z}_i$. $\boldsymbol{Z}_i$ is assumed to be independent of η_i and ν_i. The $\boldsymbol{Z}_i$ and the τ_i proxy costs and may also be correlated with J_i in (5).[3] In this framework, the goal of the analysis is to determine if $\lambda_2 = 0$, *i.e.*, to determine if agents pick schooling based on *ex post* shocks to returns and, if they do, the relative magnitude of the variance of η_i to that of ν_i.

Application of $\boldsymbol{Z}_i$ as an instrument for S_i in outcome equation (5) does not enable us to decompose $\rho_{1,i}$ into forecastable and unforecastable components. Only if agents do not use η_i in making their schooling decisions does the instrumental variable (IV) method recover

[3]Card (2001) presents a model that can be written in this form.

the population mean of $\rho_{1,i}$. In that case, standard random coefficient models can identify the variance of $(\eta_i + \nu_i)$ which is assumed to be independent of S_i.[4]

Notice that even under the most favorable conditions for applications of the IV method, we are only able to recover the *ex post* mean and total *ex post* variability of $\rho_{1,i} = \eta_i + \nu_i$. We cannot, however, decompose $Var\,(\eta_i + \nu_i)$ into its components. That is, we are not able to assign the proportion of the variance in the return that is due to η_i and that due to ν_i. Since we cannot identify how much of the *ex post* return to schooling is unknown to the agent at the time he makes his decision, we cannot solve the stated problem using just the instrumental variable method.

Our procedure is not based on the method of instrumental variables. Rather, it exploits certain covariances that arise under different information structures. To see how the method works, simplify the model down to two 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, we could also construct $E\,(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0})$. Under the correct model of expectations, we could form the residual

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

[4]One can use the residuals from $Y_i - \widehat{\rho}_0 - \widehat{\rho}_1 S_i = \widehat{U}_i$ to decompose the variance components, where instrumental variables are used to generate the coefficient estimates.

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 (5) and (6), 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 $V_{\widetilde{\mathcal{I}}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E\left(Y_{1,i} - Y_{0,i} - C_i \mid \widetilde{\mathcal{I}}_{i,0}\right)$. More precisely, the information set is valid if $S_i \perp\!\!\!\perp V_{\widetilde{\mathcal{I}}_{i,0}} \mid \widetilde{\mathcal{I}}_{i,0}$, where $X \perp\!\!\!\perp Y \mid Z$ means X is independent of Y given Z. In terms of the simple model of (5) and (6), 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. We 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 for the specified information set.[5] 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 paper.

Usually, we cannot determine the exact content of $\mathcal{I}_{i,0}$ known to each agent. If we could, we would perfectly predict S_i given our decision rule. More realistically, we might find variables that proxy $\mathcal{I}_{i,0}$ or their distribution. Thus, in the example of equations (5) and (6) we would seek to determine the distribution of ν_i and allocation of the variance of $\rho_{1,i}$ to η_i and ν_i rather than trying to estimate $\rho_{1,i}$, η_i, or ν_i for each person. This is the strategy pursued in this paper for a two-choice model of schooling.

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

Inference

The procedure just described is not practical for general models of educational outcomes. We do not know all of the information possessed by the agent. We do not observe $Y_{1,i,t}$ and $Y_{0,i,t}$ together for anyone. We must solve the problem of constructing counterfactuals. This entails solving the selection problem.

One conventional way to solve the selection problem is to invoke a 'common coefficient' assumption,

$$Y_{1,i,t} = \varphi_t\left(\boldsymbol{X}_{i,t}\right) + Y_{0,i,t}, \qquad t = 0, \ldots, T,$$

where $\varphi_t\left(\boldsymbol{X}_{i,t}\right)$ is the same for everyone with the same $\boldsymbol{X}_{i,t}$. A special case is where $\varphi_t\left(\boldsymbol{X}_{i,t}\right) = \varphi$, a constant. This specification assumes that for each person i, the earnings in college at age t differ from the earnings in high school by a constant, or a constant conditional on $\boldsymbol{X}_{i,t}$. Under standard assumptions, conventional econometric methods such as matching, instrumental variables, or control functions recover $\varphi_t\left(\boldsymbol{X}_{i,t}\right)$ for everyone (see Heckman and Robb, 1986, reprinted 2000, for discussions of alternative assumptions).

A common coefficient returns to schooling assumption for all groups with the same values of $\boldsymbol{X}_{i,t}$ rules out comparative advantage in the labor market that has been shown to be empirically important (see Carneiro *et al.*, 2004, and Heckman, 2001). This assumption can be tested nonparametrically and is decisively rejected (Heckman *et al.*, 1997). An alternative and weaker assumption is that ranks in the distribution of $Y_{1,i,t}$ can be mapped into ranks in the distribution of $Y_{0,i,t}$ (*e.g.*, the best in the $Y_{1,i,t}$ distribution is the best in the $Y_{0,i,t}$ distribution or the best in one is the worst in the other). We present evidence against that

assumption below.

An alternative approach is to use matching. Given matching variables $\boldsymbol{Q}_i$, we can form counterfactual marginal distributions from observed distributions using the matching assumption that

$$\begin{aligned} F\left(Y_{1,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 1, \boldsymbol{Q}_i\right) &= F\left(Y_{1,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 0, \boldsymbol{Q}_i\right) \\ &= F\left(Y_{1,i,t} \mid \boldsymbol{X}_{i,t}, \boldsymbol{Q}_i\right), \quad t = 0, \ldots, T. \end{aligned}$$

If the matching assumptions are valid, we can construct counterfactuals for everyone since the first distribution is observed and the second is the distribution of the counterfactual (what persons who do not attend college would have earned if they had attended college). By a parallel analysis of $F\left(Y_{0,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 0, \boldsymbol{Q}_i\right)$, we can construct $F\left(Y_{0,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 1, \boldsymbol{Q}_i\right) = F\left(Y_{0,i,t} \mid \boldsymbol{X}_{i,t}, \boldsymbol{Q}_i\right)$ for everyone, $t = 0, \ldots, T$. This is the distribution of high school outcomes for those who attend college. The marginal distributions acquired from matching are not enough to construct the distribution of returns $Y_{1,i} - Y_{0,i}$ because they do not identify the covariance or dependence between $Y_{1,i,t}$ and $Y_{0,i,t}$, unless it is assumed that the only dependence across the $Y_{1,i,t}$ and $Y_{0,i,t}$ is due to $\boldsymbol{Q}_i$ and/or $\boldsymbol{X}_{i,t}$, and the parameters of this dependence can be determined from the marginal distributions, or else special assumptions about dependence across outcomes are invoked.

Matching makes strong assumptions about the richness of the data available to analysts and does not, in general, identify joint distributions of counterfactual returns and hence the distribution of the rate of return. It assumes that the return to the marginal person is the same as the return to the average person (Heckman and Navarro, 2004).

Either matching or *IV* solves the selection problem under their assumed identifying conditions. Neither method provides a way for identifying the information agents act on *ex ante*. In this paper, we build on Carneiro *et al.* (2003) and use the factor structure representation (3) to construct the missing counterfactual earnings data without invoking either type of *ad hoc* identifying assumption.

To understand the essential idea underlying our method, consider the following linear in parameters model:

$$\begin{aligned} Y_{0,i,t} &= \boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + v_{0,i,t}, \qquad t = 0, \ldots, T, \\ Y_{1,i,t} &= \boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + v_{1,i,t}, \\ C_i &= \boldsymbol{Z}_i\boldsymbol{\gamma} + v_{i,C}. \end{aligned}$$

We 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 our method.

Suppose that there exists a vector of factors $\boldsymbol{\theta}_i = (\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,L})$ such that $\theta_{i,k}$ and $\theta_{i,j}$ are mutually independent random variables for $k, j = 1, \ldots, L, k \neq j$. Assume we can represent the error term in earnings at age t for agent i in the following manner:

$$\begin{aligned} v_{0,i,t} &= \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}, \\ v_{1,i,t} &= \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}, \end{aligned}$$

where $\boldsymbol{\alpha}_{0,t}$ and $\boldsymbol{\alpha}_{1,t}$ are vectors and $\boldsymbol{\theta}_i$ 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 $\boldsymbol{\theta}_i$. We

can also decompose the cost function C_i in a similar fashion:

$$C_i = \boldsymbol{Z}_i\boldsymbol{\gamma} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_C + \varepsilon_{i,C}.$$

All of the statistical dependence across potential outcomes and costs is generated by $\boldsymbol{\theta}$, $\boldsymbol{X}$, and $\boldsymbol{Z}$. Thus, if we could match on $\boldsymbol{\theta}_i$ (as well as $\boldsymbol{X}$ and $\boldsymbol{Z}$), we could use matching to infer the distribution of counterfactuals, and capture all of the dependence across the counterfactual states through the $\boldsymbol{\theta}_i$. However, in general, not all of the required elements of $\boldsymbol{\theta}_i$ are observed.

The parameters $\boldsymbol{\alpha}_C$ and $\boldsymbol{\alpha}_{s,t}$ for $s = 0, 1$, and $t = 0, \ldots, T$ are the factor loadings. $\varepsilon_{i,C}$ is independent of the $\boldsymbol{\theta}_i$ and the other ε components. In this notation, the choice equation can be written as:

$$\begin{aligned} I_i &= E\left(\sum_{t=0}^{T} \frac{\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - \left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+r)^t} - \left(\boldsymbol{Z}_i\boldsymbol{\gamma} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_C + \varepsilon_{iC}\right) \;\middle|\; \mathcal{I}_{i,0}\right) \\ S_i &= 1 \text{ if } I_i \geq 0;\ S_i = 0 \text{ otherwise.} \end{aligned} \tag{7}$$

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 cost. Constructing (7) 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 $\boldsymbol{X}_{i,t}$, the $\boldsymbol{\theta}_i$, 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.

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 our 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 we have defined it.[6]

To formally characterize our empirical procedure, it is useful to introduce some additional notation. Let $\odot$ denote the Hadamard product ($\mathbf{a} \odot \mathbf{b} = (\mathrm{a}_1 b_1, \ldots, \mathrm{a}_L b_L)$) for vectors $\mathbf{a}$ and $\mathbf{b}$ of length L. Let Δ_X, Δ_Z, Δ_θ, Δ_{ε_C}, Δ_{ε_t}, $t = 0, \ldots, T$, denote coefficient vectors associated with the $\boldsymbol{X}$, the $\boldsymbol{Z}$, the $\boldsymbol{\theta}$, 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 there is a deviation between the proposed information set and 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 act we can define the proposed choice index $\widetilde{I}_i$ in the following way:

$$
\begin{aligned}
\widetilde{I}_i \;=\; & \sum_{t=0}^{T} \frac{E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t}\left(\boldsymbol{\beta}_{1,t} - \boldsymbol{\beta}_{0,t}\right) + \sum_{t=0}^{T} \frac{\left[\boldsymbol{X}_{i,t} - E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t}\left(\boldsymbol{\beta}_{1,t} - \boldsymbol{\beta}_{0,t}\right) \odot \Delta_X \\
& + E(\boldsymbol{\theta}_i \mid \widetilde{\mathcal{I}}_{i,0}) \left[\sum_{t=0}^{T} \frac{(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t})}{(1+r)^t} - \boldsymbol{\alpha}_C\right] + \left[\boldsymbol{\theta}_i - E\left(\boldsymbol{\theta}_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \left\{\left[\sum_{t=0}^{T} \frac{(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t})}{(1+r)^t} - \boldsymbol{\alpha}_C\right] \odot \Delta_\theta\right\} \\
& + \sum_{t=0}^{T} \frac{E\left(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} + \sum_{t=0}^{T} \frac{\left[(\varepsilon_{1,i,t} - \varepsilon_{0,i,t}) - E\left(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t} \Delta_{\varepsilon_t} \\
& - E\left(\boldsymbol{Z}_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\boldsymbol{\gamma} - \left[\boldsymbol{Z}_i - E\left(\boldsymbol{Z}_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]\boldsymbol{\gamma} \odot \Delta_Z - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right) - \left[\varepsilon_{iC} - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \Delta_{\varepsilon_C},
\end{aligned}
\tag{8}
$$

[6] As pointed out to us by Lars Hansen, the term 'heterogeneity' is somewhat unfortunate. Under this term, we include trends common across all people (*e.g.*, macrotrends). The real distinction we 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.

To conduct our test, we fit a schooling choice model based on the proposed model (8). We estimate the parameters of the model including the Δ parameters. This decomposition for $\widetilde{I}_i$ assumes that agents know the $\boldsymbol{\beta}$, the $\boldsymbol{\gamma}$, and the $\boldsymbol{\alpha}$. We discuss this assumption in section 5. 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 $\Delta_X = 0$, $\Delta_\theta = 0$, $\Delta_Z = 0$, $\Delta_{\varepsilon_C} = 0$, and $\Delta_{\varepsilon_t} = 0$, $t = 0, \ldots, T$. That is, when $\widetilde{\mathcal{I}}_{i,0} = \mathcal{I}_{i,0}$ then $\Delta_X = 0$, $\Delta_\theta = 0$, $\Delta_Z = 0$, $\Delta_{\varepsilon_C} = 0$, $\Delta_{\varepsilon_t} = 0$, $t = 0, \ldots, T$, and the proposed choice index $\widetilde{I}_i = I_i$.

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 our method, assume for simplicity that the $\boldsymbol{X}_{i,t}$, the $\boldsymbol{Z}_i$, the $\varepsilon_{i,C}$, the $\boldsymbol{\beta}_{1,t}, \boldsymbol{\beta}_{0,t}$, the $\boldsymbol{\alpha}_{1,t}, \boldsymbol{\alpha}_{0,t}$, and $\boldsymbol{\alpha}_C$ 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 $\boldsymbol{\theta}_i$ are known and acted on in making schooling decisions if we postulate that some components of $\boldsymbol{\theta}_i$ 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 $\boldsymbol{\theta}_i$, 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 knew the unobservable captured by factor 2 that affects future earnings. If $\theta_{i,2}$ does not enter the choice equation but explains future earnings, then $\theta_{i,2}$ is unknown (not predictable by the agent) at the age schooling decisions are made. An alternative interpretation is that

the second component of $\left[\sum_{t=0}^{T} \frac{(\boldsymbol{\alpha}_{1,t}-\boldsymbol{\alpha}_{0,t})}{(1+r)^t} - \boldsymbol{\alpha}_C\right]$ is zero, *i.e.*, that even if the component is known, it is not acted on. Thus, we can only test for what the agent knows and acts on.

One plausible scenario is that $\varepsilon_{i,C}$ is known 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 beyond the dependence induced by the $\boldsymbol{\theta}_i$. Under a perfect foresight assumption we can identify this extra dependence. We develop this point further in section 3 after we introduce additional helpful notation. Our procedure can be generalized to consider all components of (8). We can test the predictive power of each subset of the overall possible information set at the date the schooling decision is being made.

The intuition underlying our testing procedure is thus very simple. The components that are forecastable and acted on in making schooling choices are captured by the components of *ex post* realizations that are known by the agents when they make their educational choices. In terms of the simple model of equations (5) and (6), and by decomposing $\rho_{1,i}$ into η_i and ν_i so $\rho_{1,i} = \eta_i + \nu_i$, we determine how much of the *ex post* variability in $\rho_{1,i}$ is due to forecastable η_i and unforecastable ν_i. The predictable components will be estimated to have nonzero coefficients in the schooling choice equation. The uncertainty at the date the decision about college is being made is captured by the factors that the agent does not act on when making the decision of whether or not to attend college.[7]

A similar but distinct idea motivates Flavin's (1981) test of the permanent income hypothesis and her measurement of unforecastable income innovations. She picks a particular

[7]This test has been extended to a nonlinear setting, allowing for credit constraints, preferences for risk, and the like. See Cunha *et al.* (2004) and Navarro (2004).

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 *et al.* (2004) extend her analysis but, like her, maintain an *a priori* specification of the stochastic process generating $\mathcal{I}_{i,0}$. Blundell *et al.* (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.

In the context of our factor structure representation, the contrast between our approach to identifying components of intrinsic uncertainty and the approach followed in the literature is as follows. The traditional approach would assume that the $\boldsymbol{\theta}_i$ are known to the agent while the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are not.[8] Our approach allows us to determine which components of $\boldsymbol{\theta}_i$ and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known and acted on at the time schooling decisions are made.

Assuming that the problems raised by selection on S_i are solved by the methods exposited in the next section and their vector generalizations, we can estimate the distributions of the components of (3) and the coefficients on the factors $\boldsymbol{\theta}_i$ from panel data on earnings. This statistical decomposition does not tell us which components of (3) are known at the time

[8]The analysis of Hartog and Vijverberg (2002) exemplifies this approach and uses variances of *ex post* income to proxy *ex ante* variability.

agents make their schooling decisions. If some of the components of $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known to the agent at the date schooling decisions are made and enter (8), 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 $\boldsymbol{\theta}_i$, would be estimated.

It is important to contrast the dependence between S_i and future $Y_{0,i,t}, Y_{1,i,t}$ arising from $\boldsymbol{\theta}_i$ from the dependence between S_i and the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$. Some of the $\boldsymbol{\theta}_i$ in the *ex post* earnings equation may not appear in the choice equation, and some additional dependence between S_i and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ may appear under certain information sets. The contrast between the sources generating earnings outcomes and the sources generating dependence between S_i and $\boldsymbol{\varepsilon}_i$ income realized is the essential idea in this paper. The method can be generalized to deal with nonlinear preferences and imperfect market environments.[9] A central issue, discussed in section 4, is how far one can go in identifying income information processes without specifying preferences, insurance, and market environments.

[9]In a model with complete autarky with preferences G, ignoring costs,

$$I_i = \sum_{t=0}^{T} E\left[\frac{G\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - G\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \;\middle|\; \widetilde{\mathcal{I}}_{i,0} \right],$$

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

$$\begin{aligned} V_{\widetilde{\mathcal{I}}_{i,0}} \;=\; & \sum_{t=0}^{T} \frac{G\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - G\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} - \\ & \sum_{t=0}^{T} E\left[\frac{G\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - G\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \;\middle|\; \widetilde{\mathcal{I}}_{i,0} \right]. \end{aligned}$$

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

3 Identifying counterfactual distributions and extracting components of unpredictable uncertainty using factor models

To motivate our econometric procedures, it is useful to work with a slightly more abstract notation and a simpler set up. Omit the individual i subscript to simplify the notation and suppose that there is one period only ($T = 0$) so $Y_1 = Y_{1,0}$, $Y_0 = Y_{0,0}$. We relax this assumption later in this section but initially use this framework to focus on the main econometric ideas motivating our solution of the selection problem. Assume that (Y_0, Y_1) have finite means and can be expressed in terms of conditioning variables $\boldsymbol{X}$. Write

$$Y_0 = \mu_0(\boldsymbol{X}) + U_0, \tag{9a}$$

$$Y_1 = \mu_1(\boldsymbol{X}) + U_1, \tag{9b}$$

where $E(U_0 \mid \boldsymbol{X}) = E(U_1 \mid \boldsymbol{X}) = 0$, $E(Y_0 \mid \boldsymbol{X}) = \mu_0(\boldsymbol{X})$, and $E(Y_1 \mid \boldsymbol{X}) = \mu_1(\boldsymbol{X})$. The *ex post* gain for an individual who moves from $S = 0$ to $S = 1$ is $Y_1 - Y_0$.

Write index I as a net utility,

$$I = Y_1 - Y_0 - C, \tag{10}$$

where C is the cost of participation in sector 1. We write $C = \mu_C(\boldsymbol{Z}) + U_C$, where the $\boldsymbol{Z}$ are determinants of cost. We may write

$$I = \mu_I(\boldsymbol{X}, \boldsymbol{Z}) + U_I. \tag{11}$$

Under perfect certainty,

$$\mu_I(\boldsymbol{X}, \boldsymbol{Z}) = \mu_1(\boldsymbol{X}) - \mu_0(\boldsymbol{X}) - \mu_C(\boldsymbol{Z}) \quad \text{and} \quad U_I = U_1 - U_0 - U_C.$$

More generally, we define U_I as the error in the choice equation and it may or may not include all future U_1, U_0, or U_C. Similarly, $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ may only be based on expectations of future $\boldsymbol{X}$ and $\boldsymbol{Z}$ at the time schooling decisions are made. We write

$$S = 1 \text{ if } I \geq 0; \; S = 0 \text{ otherwise.} \tag{12}$$

A major advantage of our approach over previous work on estimating components of uncertainty facing agents is that we control for the econometric consequences of endogeneity in the choice of S and thereby avoid self-selection biases. The choice equation is also a source of identifying information for extracting forecastable components. This paper builds on recent research by Carneiro *et al.* (2003) that solves the problem of constructing counterfactuals by identifying the joint distribution of (Y_0, Y_1) conditional on S (or I) using a factor structure model. These models generalize the LISREL models of Jöreskog (1977) and the MIMIC models of Jöreskog and Goldberger (1975) to produce counterfactual distributions. We now exposit the main idea underlying our method, working with a one-factor model to simplify

the exposition. Carneiro *et al.* (2003) develop the general multifactor model we use in our empirical analysis.

3.1 Identifying counterfactual distributions

Identifying the joint distribution of potential outcomes is a difficult problem because we do not observe both components of (Y_0, Y_1) for anyone. Thus, one cannot directly form the joint distribution of potential outcomes (Y_0, Y_1). Heckman and Honoré (1990) show that if (i) $C = 0$ for every person, (ii) decision rule (12) applies in an environment of perfect certainty, (iii) there are distinct variables in $\mu_1(\boldsymbol{X})$ and $\mu_0(\boldsymbol{X})$, (iv) $\boldsymbol{X}$ is independent of (U_1, U_0), and other mild regularity restrictions are satisfied, then one can identify the joint distribution of (Y_0, Y_1) given $\boldsymbol{X}$, even without additional $\boldsymbol{Z}$ variables. In this case the agents choose S solely in terms of the differences in potential outcomes. However, in an environment of uncertainty or if C varies across people and contains some variables unobserved by the analyst, this method breaks down. We present a more general analysis without maintaining the perfect certainty assumption.

As shown by Heckman (1990), Heckman and Smith (1998), and Carneiro *et al.* (2003), under the assumptions that (i) $(\boldsymbol{Z}, \boldsymbol{X})$ are statistically independent from (U_0, U_1, U_I), (ii) $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ is a nontrivial function of $\boldsymbol{Z}$ given $\boldsymbol{X}$, (iii) $\mu_0(\boldsymbol{X}), \mu_1(\boldsymbol{X})$, and $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ have full support, and (iv) the elements of the pairs $(\mu_0(\boldsymbol{X}), \mu_I(\boldsymbol{X}, \boldsymbol{Z}))$ and $(\mu_1(\boldsymbol{X}), \mu_I(\boldsymbol{X}, \boldsymbol{Z}))$ can be varied independently of each other, then one can identify the joint distributions of (U_0, U_I), (U_1, U_I) up to a scale σ_I^* for U_I and also $\mu_0(\boldsymbol{X}), \mu_1(\boldsymbol{X})$, and $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$, the last expression up to scale σ_I.[10] Thus, one can identify the joint distributions of (Y_0, I^*) and

[10]Full support means that the support of $\mu_1(\boldsymbol{X})$ matches (or contains) the support of U_1; the support

(Y_1, I^*) given $\boldsymbol{X}$ and $\boldsymbol{Z}$ where $I^* = I/\sigma_I$. As a by-product we identify the mean functions. One cannot recover the joint distribution of (Y_0, Y_1) or (Y_0, Y_1, I^*) given $\boldsymbol{X}$ and $\boldsymbol{Z}$ without further assumptions. We provide an intuitive motivation for why $F(Y_0, I^*)$ and $F(Y_1, I^*)$ are identified in Appendix 1. Once we estimate these distributions, we perform factor analysis on (Y_0, I^*) and (Y_1, I^*).

The factor structure approach provides a solution to the problem of constructing counterfactual distributions. We show the essential ideas. Suppose that the unobservables follow a one-factor structure (*i.e.*, θ is a scalar). Carneiro *et al.* (2003) generalize these methods to the multifactor case. We can extend these methods to nonseparable models using the analysis reported in Heckman, Matzkin, Navarro, and Urzua (2004), but we do not do so in this paper.

We assume that all of the dependence across (U_0, U_1, U_{I^*}) is generated by a scalar factor θ,

$$\begin{aligned} U_0 &= \theta\alpha_0 + \varepsilon_0, \\ U_1 &= \theta\alpha_1 + \varepsilon_1, \\ U_{I^*} &= \theta\alpha_{I^*} + \varepsilon_{I^*}. \end{aligned}$$

We assume that θ is statistically independent of $(\varepsilon_0, \varepsilon_1, \varepsilon_I)$ and satisfies $E(\theta) = 0$, and $E(\theta^2) = \sigma_\theta^2$. All the ε's are mutually independent with $E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_{I^*}) = 0$, $Var(\varepsilon_0) = \sigma_{\varepsilon_0}^2$, $Var(\varepsilon_1) = \sigma_{\varepsilon_1}^2$, and $Var(\varepsilon_I) = \sigma_{\varepsilon_I}^2$ (the ε terms are called uniquenesses in

of $\mu_0(\boldsymbol{X})$ matches (or contains) the support of U_0 and the support of $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ matches (or contains) the support of U_I. (See Heckman and Honoré, 1990, and Carneiro *et al.*, 2003, for more precise formulations of these conditions.) The support of a random variable is the set of values where it has a positive density.

factor analysis). Because the factor loadings may be different, the factor may affect outcomes and choices differently and may even have different signs in different equations.

To show how one can recover the joint distribution of (Y_0, Y_1) using factor models, we break the argument into two parts. First we show how to recover the factor loadings, factor variance, and the variances of the uniquenesses. This part is like traditional factor analysis except that some latent variables (*e.g.*, I^*) are only observed up to scale so their scale must be normalized. Then, we show how to construct joint distributions of counterfactuals.

3.2 Recovering the factor loadings

We consider identification of the model when the analyst has different types of information about the choices and characteristics of the agent.

3.2.1 The case when there is information on Y_0 for $I < 0$ and Y_1 for $I > 0$ and the decision rule is (12)

Under the conditions stated in section 3.1 and the papers referenced there, after conditioning on $\boldsymbol{X}$ and controlling for selection, one can identify $F(U_0, U_{I^*})$ and $F(U_1, U_{I^*})$. From these distributions one can identify the left hand side of

$$Cov(U_0, U_{I^*}) = \alpha_0 \alpha_{I^*} \sigma_\theta^2$$

and

$$Cov(U_1, U_{I^*}) = \alpha_1 \alpha_{I^*} \sigma_\theta^2.$$

The scale of the unobserved I is normalized, a standard condition for discrete choice models. A second normalization that we need to impose is $\sigma_\theta^2 = 1$. This is required since the factor is not observed and we must set its scale. That is, since $\alpha\theta = k\alpha\frac{\theta}{k}$ for any constant k, we need to set the scale by normalizing the variance of θ. We could alternatively normalize some α_j to one. Finally, we set $\alpha_{I^*} = 1$, an assumption we can relax, as noted below.

Under these conditions, we can identify α_1 and α_0 from the known covariances above. From the first covariance, we identify α_0. From the second, we identify α_1. From the normalization, we know σ_θ^2. Since

$$Cov\left(U_1, U_0\right) = \alpha_1\alpha_0\sigma_\theta^2,$$

we can identify the covariance between Y_1 and Y_0 even though we do not observe the pair (Y_1, Y_0) for anyone. We then use the variances $Var\left(U_1\right), Var\left(U_0\right)$ and the normalization $Var\left(U_{I^*}\right) = 1$ to recover the variance of the uniquenesses $\sigma_{\varepsilon_0}^2, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_{I^*}}^2$.

The fact that we needed to normalize both $\sigma_\theta^2 = 1$ and $\alpha_{I^*} = 1$ is a consequence of our assumption that we have only one observation for Y_1 and Y_0. If we have access to more observations on life cycle earnings from panel data, as we do in our empirical work, we can use $(Y_{0,0}, \ldots Y_{0,T}, Y_{1,0}, \ldots, Y_{1,T})$ to relax one normalization, say $\sigma_\theta^2 = 1$, since then we can form, conditional on $\boldsymbol{X}$ and $\boldsymbol{Z}$, the left hand side of

$$\frac{Cov\left(U_{1,t'}, U_{1,t}\right)}{Cov\left(U_{1,t'}, U_{I^*}\right)} = \alpha_{1t}$$

and

$$\frac{Cov\left(U_{0,t'}, U_{0,t}\right)}{Cov\left(U_{0,t'}, U_{I^*}\right)} = \alpha_{0t},$$

and recover σ_θ^2 from, say, $Cov\left(U_{1,t}, U_{I^*}\right) = \alpha_{1,t}\sigma_\theta^2$. Identification of the variances of the uniquenesses follows as before.

The central idea motivating our identification strategy is that even though we never observe (Y_0, Y_1) as a pair, both Y_0 and Y_1 are linked to S through the choice equation. From S we can generate I^*, using standard methods in discrete choice analysis. From this analysis we effectively observe (Y_0, I^*) and (Y_1, I^*). The common dependence of Y_0 and Y_1 on I^* secures identification of the joint distribution of Y_0, Y_1, I^*. We next develop a complementary strategy based on the same idea where, in addition to a choice equation, we have a measurement equation observed for all observations whether or not Y_1 or Y_0 is observed. The measurement may be a test score which is a proxy for 'ability' θ. This measurement plays the role of I^* and, in certain respects, identification with a measurement of this type is more transparent and more traditional.

3.2.2 Adding a measurement equation

Suppose that we have access to a measurement for θ that is observed whether $S = 1$ or $S = 0$ in addition to data on outcomes S and Y_0 or Y_1. In educational statistics, a test score is often used to proxy ability. Suppose that the analyst has access to one ability test M for

each person. Measured ability M is

$$M = \mu_M(\boldsymbol{X}) + U_M.$$

Assume that

$$U_M = \theta\alpha_M + \varepsilon_M,$$

where ε_M is mutually independent from $(\varepsilon_0, \varepsilon_1, \varepsilon_I)$, and θ.[11] We assume $\alpha_M \neq 0$. With this additional information we can form

$$\begin{aligned} Cov(M, Y_0|\boldsymbol{X}, \boldsymbol{Z}) &= Cov(U_M, U_0) = \alpha_M\alpha_0\sigma_\theta^2, \\ Cov(M, Y_1|\boldsymbol{X}, \boldsymbol{Z}) &= Cov(U_M, U_1) = \alpha_M\alpha_1\sigma_\theta^2, \\ Cov(M, I^*|\boldsymbol{X}, \boldsymbol{Z}) &= Cov(U_M, U_{I^*}) = \alpha_M\alpha_{I^*}\sigma_\theta^2. \end{aligned}$$

Conditioning on $(\boldsymbol{X}, \boldsymbol{Z})$, we can recover the error terms for the unobservables U_0, U_{I^*} and U_M using the preceding arguments. If we impose the normalization $\alpha_M = 1$, which can be interpreted as requiring that higher levels of measured ability are associated with higher levels of factor θ, we can form the ratio

$$\frac{Cov(U_0, U_{I^*})}{Cov(U_M, U_{I^*})} = \alpha_0$$

[11] For simplicity, we assume that this is a continuous measurement. Discrete measurements can also be used. See Carneiro *et al.* (2003).

and identify α_0. In a similar fashion, we can form

$$\frac{Cov\left(U_1, U_{I^*}\right)}{Cov\left(U_M, U_{I^*}\right)} = \alpha_1$$

and we can recover α_1. From

$$Cov\left(U_M, U_0\right) = \alpha_0 \sigma_\theta^2,$$

we can obtain σ_θ^2. Finally, we can identify α_{I^*} based on information from

$$Cov\left(U_M, U_{I^*}\right) = \alpha_{I^*} \sigma_\theta^2,$$

so we can obtain α_{I^*} up to scale. Thus, with one measurement, one choice equation and two outcomes we can identify σ_θ^2 and α_{I^*} up to scale. We can use the identified variances $Var\left(U_0\right), Var\left(U_1\right), Var\left(U_{I^*}\right) = 1$, and $Var\left(U_M\right)$ to recover the uniquenesses $\sigma_{\varepsilon_0}^2, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_{I^*}}^2$, and $\sigma_{\varepsilon_M}^2$. Thus, having access to a measurement (M) and choice data with decision rule (10)–(12) allows us to estimate the covariances among the counterfactual states.[12]

But how to identify the distributions? Traditional factor analysis assumes normality. We present a more general nonparametric analysis. Allowing for nonnormality is essential for getting acceptable empirical results as we note below.

[12]We cannot dispense with the choice equation unless we have data on $F(Y_0, M \mid \boldsymbol{X}, \boldsymbol{Z})$ and $F(Y_1, M \mid \boldsymbol{X}, \boldsymbol{Z})$. Recall that, in most cases, we observe data that allows us to construct $F\left(Y_0, M \mid \boldsymbol{X}, \boldsymbol{Z}, S = 0\right)$ and $F\left(Y_1, M \mid \boldsymbol{X}, \boldsymbol{Z}, S = 1\right)$. The required information for dispensing with the choice equation might be obtained when we have limit sets $\bar{\mathcal{Z}}_u$ and $\bar{\mathcal{Z}}_l$ such that $\Pr(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}) = 1$ for $\mathbf{z} \in \bar{\mathcal{Z}}_u$ and $\Pr(S = 0 \mid \boldsymbol{X}, \boldsymbol{Z}) = 0$ for $\mathbf{z} \in \bar{\mathcal{Z}}_l$. Then we can replace I with M and do factor analysis. (See Carneiro *et al.*, 2001.)

3.3 Recovering the distributions nonparametrically

Given the identification of factor loadings, factor variances, and uniquenesses, we show how to identify the marginal distributions of θ and $\varepsilon_0, \varepsilon_1, \varepsilon_{I^*}$ nonparametrically (the last one up to scale). The method is based on a theorem by Kotlarski (1967). For completeness, we state his theorem.

Theorem 1 Suppose that we have two random variables T_1 and T_2 that satisfy:

$$\begin{aligned} T_1 &= \theta + v_1 \\ T_2 &= \theta + v_2 \end{aligned}$$

with θ, v_1, v_2 mutually statistically independent, $E(\theta) < \infty$, $E(v_1) = E(v_2) = 0$, that the conditions for Fubini's theorem are satisfied for each random variable, and that the random variables possess nonvanishing (almost everywhere) characteristic functions. Then, the densities $f_\theta, f_{v_1}, f_{v_2}$ are identified.

Proof See Kotlarski (1967). □

Applied to the current context, we have a choice equation, two outcome equations, and a measurement equation.[13] Assume that we normalize $\alpha_M = 1$ so that all factor loadings,

[13] Again, for the sake of simplicity, we assume that M is continuous but our methods work for discrete measurements. (See Carneiro *et al.*, 2003).

factor variances, and variances of uniquenesses are known. The system is

$$\begin{aligned} I^* &= \mu_{I^*}(\boldsymbol{X}, \boldsymbol{Z}) + \theta\alpha_{I^*} + \varepsilon_{I^*}, \\ Y_0 &= \mu_0(\boldsymbol{X}) + \theta\alpha_0 + \varepsilon_0, \\ Y_1 &= \mu_1(\boldsymbol{X}) + \theta\alpha_1 + \varepsilon_1, \\ M &= \mu_M(\boldsymbol{X}) + \theta + \varepsilon_M. \end{aligned}$$

Note that this system can be rewritten as

$$\begin{aligned} \frac{I^* - \mu_{I^*}(\boldsymbol{X}, \boldsymbol{Z})}{\alpha_{I^*}} &= \theta + \frac{\varepsilon_{I^*}}{\alpha_{I^*}}, \\ \frac{Y_0 - \mu_0(\boldsymbol{X})}{\alpha_0} &= \theta + \frac{\varepsilon_0}{\alpha_0}, \\ \frac{Y_1 - \mu_1(\boldsymbol{X})}{\alpha_1} &= \theta + \frac{\varepsilon_1}{\alpha_1}, \\ M - \mu_M(\boldsymbol{X}) &= \theta + \varepsilon_M. \end{aligned}$$

Applying Kotlarski's theorem to any pair of equations, we conclude that we can identify the densities of $\theta, \frac{\varepsilon_{I^*}}{\alpha_{I^*}}, \frac{\varepsilon_0}{\alpha_0}, \frac{\varepsilon_1}{\alpha_1}, \varepsilon_M$. Since we know α_{I^*}, α_0, and α_1, we can identify the densities of $\theta, \varepsilon_{I^*}, \varepsilon_0, \varepsilon_1, \varepsilon_M$.[14] Thus, we can identify the distributions of all of the error terms. Finally, to recover the joint distribution of (Y_1, Y_0), note that

$$F(Y_1, Y_0 \mid \boldsymbol{X}) = \int F(Y_1, Y_0 \mid \theta, \boldsymbol{X})\, dF_\theta(\theta).$$

From Kotlarski's theorem, $F_\theta(\theta)$ is known. Because of the factor structure, Y_1, Y_0, and S are

[14] Recall that U_I is only known up to scale σ_I.

independent once we condition on θ, so it follows that

$$F\left(Y_1, Y_0 \mid \theta, \boldsymbol{X}\right) = F\left(Y_1 \mid \theta, \boldsymbol{X}\right) F\left(Y_0 \mid \theta, \boldsymbol{X}\right).$$

But $F\left(Y_1 \mid \theta, \boldsymbol{X}\right)$ and $F\left(Y_0 \mid \theta, \boldsymbol{X}\right)$ are identified once we condition on the factors since

$$\begin{aligned} F\left(Y_1 \mid \theta, \boldsymbol{X}, S=1\right) &= F\left(Y_1 \mid \theta, \boldsymbol{X}\right) \\ F\left(Y_0 \mid \theta, \boldsymbol{X}, S=0\right) &= F\left(Y_0 \mid \theta, \boldsymbol{X}\right). \end{aligned}$$

Note further that if θ were known to the analyst, our procedure would be equivalent to matching on θ which is equivalent, for identification, to matching on the propensity score $\Pr\left(S=1 \mid \boldsymbol{X}, \boldsymbol{Z}, \theta\right)$.[15] Our method generalizes matching by allowing the variables that would produce the conditional independence assumed in matching to be unobserved by the analyst.

The discussion in this section is for a one-factor model. In our empirical work, we use a multifactor model where the factors are used to characterize earnings dynamics and possible dependence between future ε and S. Carneiro *et al.* (2003) provide the analysis we need for the general multifactor case. The key idea is that, with enough measurements, outcomes and choice equations, we can identify the number of factors generating dependence among the Y_1, Y_0, C, S, and M and the distributions of the factors.[16]

[15] Carneiro *et al.* (2003) discuss the matching relationship between factor and matching models. For a discussion of factor models and control functions, see Heckman and Navarro (2004).

[16] A precise statement of what is 'enough' information is given in Carneiro *et al.* (2003). See their discussion of the Ledermann bound. The key idea is that the number of factors has to be small relative to the number of measurements, outcomes and choice equations. This bound can be relaxed if there are *a priori* restrictions on the factor loadings beyond innocuous normalizations. Using nonnormality one can also relax the Ledermann bound.

3.4 Models with multiple factors and tests for full insurance versus perfect certainty

Our empirical work is based on a 5 period ($t = 0, \ldots, 4$) version of equations (1) and (8). In fitting the model, we introduce the possibility of additional sources of dependence in the choice equation (8), distinct from the dependence arising from some or all of the components of $\boldsymbol{\theta}$. This additional dependence may be generated from future $(\varepsilon_{1,i,t}, \varepsilon_{0,i,t})$, $t = 0, \ldots, T$ that affect schooling choices.

From the covariances between S_i (or I_i^*) and $Y_{0,i,t}$ and $Y_{1,i,t}$, $t = 0, \ldots, T$, under certain information sets, we can identify additional sources of dependence between $(Y_{0,i,t}, Y_{1,i,t})$ and I_i^* apart from $\boldsymbol{\theta}_i$ arising from the dependence of $\varepsilon_{0,i,t}$ and $\varepsilon_{1,i,t}$ with $\sum_{t=0}^{T} \frac{E(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} | \widetilde{\mathcal{I}}_{i,0})}{(1+r)^t}$. In our empirical specification discussed below, there are multiple earnings outcomes in each schooling state, a choice equation and a vector of measurement equations to tie down the distribution of $\boldsymbol{\theta}_i$ and the distributions of the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$.

To see how additional sources of dependence might arise in fitting the data, consider a model with perfect foresight. Following the analysis in section 3.2 and in the papers cited there, we can estimate

$$Cov\left(Y_{j,i,t}, I_i^* \mid \boldsymbol{X}, \boldsymbol{Z}\right) = \frac{\boldsymbol{\alpha}_{j,t}'}{\sigma_I^*} \Sigma_\Theta \left[\frac{\sum_{t=0}^{T} \left(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t}\right)}{\left(1+r\right)^t} - \boldsymbol{\alpha}_C \right] + \left(\frac{1}{\sigma_I^*}\right) \frac{Var\left(\varepsilon_{j,i,t}\right)}{\left(1+r\right)^t}, \quad t = 0, \ldots, T; \; j = 0, 1,$$

where Σ_Θ is the variance-covariance matrix of the $\boldsymbol{\theta}_i$. Conditional on $\boldsymbol{X}$ and $\boldsymbol{Z}$, dependence between $Y_{j,i,t}$ and I_i^* can arise from two sources: from the $\boldsymbol{\theta}_i$ and from the $\varepsilon_{j,i,t}$. Under

complete markets, if the $\varepsilon_{j,i,t}$ are unknown at date $t = 0$ and have mean zero given $\mathcal{I}_{i,0}$, the second term on the right hand side vanishes and the factors $\boldsymbol{\theta}_i$ capture any dependence between $Y_{j,i,t}$ and S_i.

Using limit set arguments, as in Carneiro *et al.* (2003) and Cunha *et al.* (2004), we can identify the $\boldsymbol{\alpha}_{j,t}$, $j = 0, 1$, $t = 0, \ldots, T$, the distribution of $\boldsymbol{\theta}_i$ and the distributions of the $\varepsilon_{j,i,t}$ from earnings data alone in the limit sets.[17] Under either complete markets or under perfect foresight, we can identify $\boldsymbol{\alpha}_C$ up to scale σ_I^* from the covariances between $Y_{j,i,t}$, and I_i^*, provided a rank condition is satisfied. In the case of scalar θ_i, we can identify α_C for a fixed scale of I_i^* from the preceding equation for perfect foresight as

$$\frac{1}{\alpha_{j,t}\sigma_\theta^2}\left[-Cov\left(Y_{j,i,t}, I_i^* \mid \boldsymbol{X}, \boldsymbol{Z}\right) + \frac{Var\left(\varepsilon_{j,i,t}\right)}{\left(\sigma_I^*\right)\left(1+r\right)^t} + \frac{\alpha_{j,t}}{\sigma_I^*}\sigma_\theta^2 \frac{\sum_{t=0}^{T}\left(\alpha_{1,t}-\alpha_{0,t}\right)}{\left(1+r\right)^t}\right] = \frac{\alpha_C}{\sigma_I^*}.$$

Since we know all of the ingredients on the left hand side, we can identify α_C up to scale σ_I^*. If there is an element of $\boldsymbol{X}$ not in $\boldsymbol{Z}$, we can identify the scale σ_I^* (See equation (7)). Since α_C is overidentified if $T > 0$, we can test between a perfect foresight model and a complete contingent claims model by checking if the same α_C is estimated for different $Cov\left(Y_{j,i,t}, I^*\right)$ terms.[18] In the complete contingent claims model with uncertainty, the middle term in the brackets would be zero for all $\varepsilon_{j,i,t}$.[19]

[17] Footnote 12 defines the limit sets. See Carneiro *et al.* (2003) for a more complete discussion of identification in limit sets.

[18] This procedure would break down only if $\frac{\frac{Var\left(\varepsilon_{j,i,t}\right)}{\left(1+r\right)^t}}{\alpha_{j,t}\Sigma_\Theta \sum_{t=0}^{T} \frac{\left(\alpha_{1,t}-\alpha_{0,t}\right)}{\left(1+r\right)^t}}$ is constant across all t.

[19] This testing procedure generalizes to the case of vector $\boldsymbol{\theta}$ provided that a rank condition

$$\boldsymbol{\alpha}_{j,t}'\Sigma_\Theta \frac{\sum_{t=0}^{T}\left(\boldsymbol{\alpha}_{1,t}-\boldsymbol{\alpha}_{0,t}\right)}{\left(1+r\right)^t} \neq 0$$

holds for a collection of L terms of the covariances of $Y_{j,i,t}$ with I_i^* where L is the number of factors.

4 More general preferences and market settings

To focus on the main ideas regarding model identification in this paper, we have deliberately 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 we have used to this point breaks down.

If we postulate information processes *a priori*, and preferences up to some unknown parameters as in Flavin (1981), Blundell and Preston (1998), and Blundell *et al.* (2004), we can identify departures from specified market structures. In Cunha *et al.* (2004), we postulate an Aiyagari (1994) – Laitner (1992) economy with one asset and parametric preferences to identify the information processes in the agent's information set. We 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. In Cunha *et al.* (2004), we 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 their empirical results by Flavin (1981) and Blundell *et al.* (2004) may be a consequence of their misspecification of the agent's information set generating process. We discuss this point further in section 5 when

we present our estimates, to which we now turn.

5 Empirical results

We first describe our data and estimating equations. We then discuss the estimates obtained from our model, and their economic implications.

5.1 The data, equations, and estimation

Appendix 2 considers a practical problem that plagues life cycle analysis. Few data sets contain the full life cycle of earnings along with the test scores and schooling choices needed to directly estimate our model and extract components of uncertainty. We need to combine data sets. Otherwise, we can only obtain partial identification of the model. In our empirical analysis, we use a sample of white males from the NLSY data pooled with PSID data, as described in Appendix 3 (found at http://jenni.uchicago.edu/Hicks2004/), to produce life cycle data on earnings and schooling.

Following the preceding theoretical analysis, we consider only two schooling choices: high school and college graduation. From now on we use c, h to denote college and high school, respectively. 'c' corresponds to 1 and 'h' corresponds to 0 in the previous notation. For simplicity and familiarity, in this paper we assume complete contingent claims markets. Because we assume that all shocks are idiosyncratic, schooling choices are made on the basis of expected present value income maximization. Carneiro *et al.* (2003) assume the absence of any credit markets or insurance. One of the goals of this paper is to check whether their empirical findings about components of income inequality are robust to different assumptions

about the operation of the credit market and insurance markets. Cunha *et al.* (2004) estimate an Aiyagari-Laitner economy with a single asset and borrowing constraints and discuss risk aversion and the relative importance of uncertainty.

The method developed in this paper is based on the idea that some or all components of expected future earnings may affect current choices. In order to gain some preliminary insights on whether components of future earnings (and returns) affect current schooling choices, we present a simple empirical analysis in Table 1. Using the sample described below and in Appendix 3 (found at http://jenni.uchicago.edu/Hicks2004/), we regress log *ex post* earnings on schooling and schooling interacted with an ability test (ASVAB) to obtain an estimate of the *ex post* return to schooling under the assumption that, conditional on the test score, the *ex post* return is the same for everyone.[20] This is a form of matching estimator as described in Section 2. Assuming that the conditioning variable controls for selection, we use the estimated return to schooling and plug it into a schooling choice model to test whether future earnings affect college choices. In order to account for possible selection biases not controlled for by matching, we repeat the exercise using instrumental variables estimates of returns instead.[21] The matching (OLS) estimator is reported in the first row of Table 1. The IV estimator is reported in the second row. The estimated effects of these estimators on schooling choices are given in the third and fourth rows. For either estimation method, we find statistically significant evidence that estimated *ex post* returns affect current schooling choices. This evidence suggests that some components of future earnings may predict schooling.

[20]We only use the NLSY sample because of the availability of instruments in it.

[21]See Heckman and Navarro (2004) for an exposition of the strong conditions required for this to be a valid procedure.

However, this evidence is not decisive. The estimates do not clearly delineate what is unknown to the agent at the time schooling choices are made. They also do not distinguish between the role of ability in generating future earnings from the role of ability in reducing costs of schooling. The procedure developed in this paper makes these distinctions. We can also determine the information set facing agents using the method developed in the previous sections, which we now apply.[22]

Table 2.1 presents descriptive statistics of the data used to estimate the model. College graduates have higher present value of earnings than high school graduates. College graduates also have higher test scores, come from better family backgrounds, and are more likely to live in a location where college tuition is lower.

To simplify the empirical analysis, we divide the lifetimes of individuals into 5 periods. The first period covers ages 19 through 28, the second goes from 29 through 38, the third from 39 to 48, the fourth from 49 to 58, and the fifth from 59 to 65. For each schooling level s, $s \in \{c, h\}$, and for each period t, we calculate the present value of earnings as of age 19, $Y_{s,t}$.[23] To simplify notation drop the 'i' subscript. If $Y_{s,t}$ is generated by a three factor model, we would write:

$$Y_{s,t} = \boldsymbol{X}\boldsymbol{\beta}_{s,t} + \theta_1 \alpha_{s,t,1} + \theta_2 \alpha_{s,t,2} + \theta_3 \alpha_{s,t,3} + \varepsilon_{s,t} \text{ for } t = 0, 1, 2, 3, 4, \ s \in \{c, h\}. \tag{13}$$

It turns out that a three factor model is all that is required to fit the data. Since the scales

[22] A better test would be based on variables that more plausibly affect returns but not schooling, except through returns. Labor market prices for different schooling levels are one plausible candidate.

[23] In our empirical work we use a 3% interest rate. We assume it is constant. It would be useful to explore alternative time series of interest rates based on the data actually facing our cohorts. Alternative choices of constant interest rates do no affect the main qualitative findings about the relative importance of forecastable heterogeneity.

of the factors are unknown, it is necessary to normalize some loadings (the $\boldsymbol{\alpha}$). In this paper, we set $\alpha_{h,0,2} = \alpha_{h,2,3} = 1$. The normalization for ability (associated with the measurements $\boldsymbol{M}$ based on test scores) is presented in the next paragraph. Using the identification scheme of Carneiro *et al.* (2003) for the factor loadings, we also normalize $\alpha_{s,t,3} = 0$, for $t = 0$ and $t = 1$ and for $s = c$ and $s = h$. This normalization has the substantive interpretation that θ_3 affects earnings only in the third and subsequent periods. Thus, θ_3 is associated with mid-career wage developments.

For the measurement system for cognitive ability ($\boldsymbol{M}$ in the notation of section 3.2.2) we use five components of the ASVAB test battery: arithmetic reasoning, word knowledge, paragraph comprehension, math knowledge and coding speed. We dedicate the first factor (θ_1) to this test system, and exclude the others from it. This justifies our interpretation of θ_1 as ability. We include family background variables among the covariates $\boldsymbol{X}_M$ in the ASVAB test equations. In Table 2.2 we list the elements of $\boldsymbol{X}_M$. Formally, let M_j denote the test score j,

$$M_j = \boldsymbol{X}_M \boldsymbol{\omega}_j + \theta_1 \alpha_{test_j,1} + \varepsilon_{test_j}. \tag{14}$$

To set the scale of θ_1, we normalize $\alpha_{test_1,1} = 1$.

The cost function C is given by

$$C = \boldsymbol{Z}\boldsymbol{\gamma} + \theta_1 \alpha_{C,1} + \theta_2 \alpha_{C,2} + \varepsilon_C, \tag{15}$$

where the $\boldsymbol{Z}$ are variables that affect the costs of going to college and include variables that

do not affect outcomes $Y_{s,t}$ such as local tuition. Table 2.2 shows the full set of covariates used, and the exclusions (the variables in $\boldsymbol{Z}$ not in $\boldsymbol{X}$.) We include tuition among the elements of $\boldsymbol{Z}$ but allow for a more general notion of costs in our empirical work, including psychic costs.

The valuation or net utility function for schooling choice is

$$I = E_0\left(\sum_{t=0}^{4}\frac{Y_{c,t}-Y_{h,t}}{(1+r)^t}\right) - E_0\left(C\right), \tag{16}$$

where E_0 denotes the information set under $\mathcal{I}_0$ and r is the interest rate. Individuals go to college if $I > 0$. The individual decision maker is assumed to be the child although parental resources can affect C. Cost variable C also includes the effect of ability on reducing tuition costs. We test and do not reject the hypothesis that individuals, at the time they make college going decisions, know their cost functions, the $\boldsymbol{Z}$ and the $\boldsymbol{X}$, factors θ_1, θ_2, and unobservables in cost ε_C. However, they do not know factor θ_3, or $\varepsilon_{s,t}$, $s \in \{c,h\}$, $t \in \{0,1,2,3,4\}$, at the time they make their educational choices. Addition of these components to the choice equation does not improve the fit of the model to the data.[24]

We assume that each factor k, is generated by a mixture of J_k normal distributions,

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

here $\phi\left(\eta \mid \mu_j, \tau_j\right)$ is a normal density for η with mean μ_j and variance τ_j and $\sum_{j=1}^{J_k} p_{k,j} = 1$, and $p_{k,j} > 0$. As shown in Ferguson (1983), mixtures of normals with a large number of

[24]We use 't' statistics in the choice equation to determine whether additional factors enter the choice equation. We use χ^2 goodness of fit measures to determine if additional factors are required.

components approximate any distribution of θ_k arbitrarily well in the ℓ^1 norm. The $\varepsilon_{s,t}$ are also assumed to be generated by mixtures of normals. We estimate the model using Markov Chain Monte Carlo methods as described in Carneiro *et al.* (2003). In Tables $2.3 - 2.5$ we present estimated coefficients and factor loadings. For all factors, a two-component model ($J_k = 2$, $k = 1, \ldots, 3$) is adequate.[25]

5.2 Empirical results

5.2.1 How the model fits the data

To assess the validity of our estimates and to assess the number of factors we need and the number of components of the mixtures that are required, we perform a variety of checks of fit of predictions against the data. We first compare the proportions of people who choose each schooling level. In the NLSY data, 52.9% choose high school and 47.1% choose college. The model predicts roughly 53.2% and 46.8%, respectively. The model replicates the observed proportions remarkably well, and formal tests of equality of predicted and actual proportions cannot be rejected at the 5% significance level. This is also true when we partition the data on subsets of $\boldsymbol{X}$ and $\boldsymbol{Z}$.

Figures 1.1–1.5 show the densities of the predicted and actual present values of earnings for the overall sample of the pooled NLSY-PSID data sets.[26] The fit is good. When we perform formal tests of equality of predicted and actual overall distributions at the 5% level, the model marginally fails to fit the data for the overall sample for the first, third and

[25] Additional components do not improve the goodness of fit of the model to the data.

[26] The earnings are pretax. It would be better to use post-tax earnings and we propose to do so in subsequent work.

last periods (see Table 3a). However, addition of factors and additional components of the mixture of normals do not significantly improve the fit. Reducing the number of factors by one substantially reduces the overall fit (see Table 3b). Figures 2.1–2.5 and 3.1–3.5, show the same densities restricted to the sample of those who choose high school (sequence 2) and college (sequence 3). The fit is also good. The model fits the data better when we perform formal tests of equality of predicted and actual distributions for each schooling choice than it does overall, suggesting the failure of fit is due to the failure to predict mean differences. As is apparent from Table 3a, the only case in which the model does not pass the χ^2 goodness of fit test is for the high-school distribution of earnings in period 4. We conclude that a three factor model with our normalizations fits the data. From this analysis, we conclude that earnings innovations $\varepsilon_{s,t}$ in a three factor model are not in the agents' information sets at the time they are making schooling decisions. If they were, additional factors would be required to capture the full covariance between educational choices and future earnings.[27]

5.2.2 The factors: non-normality and evidence on selection

Figure 4 reveals that in order to fit the data, one must allow for non-normal factors. The figure plots the estimated densities of the factors along normal versions with the same mean and variance. None of the factors is normally distributed. A traditional assumption used in factor analysis (see Jöreskog, 1977) is violated. Our approach is more general and does not require normality.

Figure 5.1 plots the density of factor 1 conditional on educational choices. The solid line is the density of factor 1 for agents who are high school graduates, while the dashed line is

[27] Cunha *et al.* (2004) consider application of alternative testing and model selection criteria.

the density of the factor for agents who are college graduates. Since factor 1 is associated with cognitive tests, we can interpret it as an index of 'ability' . The agents who choose college have, on average, higher ability. Factor 1 is estimated from a test score equation that controls for parental background and level of education at the date the ASVAB tests are taken. Figure 5.1 shows that selection on ability is an important factor in explaining college attendance. A similar analysis of factor 2 that is presented in Fig. 5.2 reveals that schooling decisions are not very much affected by it, while we see no evidence of selection by schooling level on factor 3 (see Fig. 5.3). This evidence is consistent with the interpretation that at the time agents make their schooling decisions, they do not know factor 3. Agents cannot select on factors they do not know when they are making their schooling decisions.

5.2.3 Estimating joint distributions of *ex ante* and *ex post* counterfactuals: returns, costs, and ability as determinants of schooling

A major contribution of this paper is the identification and estimation of *ex ante* and *ex post* distributions of outcomes and returns without imposing special assumptions about the dependence across potential outcomes. Letting E_0 denote the expectation under the *ex ante* information set $\mathcal{I}_0$, we construct the distribution of (Y_0, Y_1) (*ex post*) and of $(E_0(Y_0), E_0(Y_1))$ *ex ante* conditional on $\boldsymbol{X}$. The $\boldsymbol{X}$ are assumed to be known both *ex ante* and *ex post*. The *ex post* gross return R (excluding cost) is

$$R = \frac{Y_1 - Y_0}{Y_0}$$

while the *ex ante* gross return is

$$E_0(R) = E_0\left(\frac{Y_1 - Y_0}{Y_0}\right).$$

Both population heterogeneity and uncertainty produce the randomness generating R. Population heterogeneity in $\mathcal{I}_0$ (information sets) produces the randomness generating $E_0(R)$. A standard argument shows that the means of R and $E_0(R)$ over the entire population, and on any conditioning subset, are the same.

In estimating the distribution of earnings in counterfactual schooling states within a policy regime (*e.g.*, the distributions of college earnings for people who actually choose to be high school graduates under a particular tuition policy), one standard approach is to assume that both distributions are the same except for an additive constant—the coefficient of a schooling dummy in an earnings regression possibly conditioned on the covariates. Recently developed methods relax this assumption by assuming preservation of ranks across potential outcome distributions, but do not freely specify the two outcome distributions (see Heckman *et al.*, 1997; Chernozhukov and Hansen, 2005; Shaikh and Vytlacil, 2004).

Table 4.1 presents the *ex post* conditional distribution of college earnings given high school earnings decile by decile. If the dependence across outcomes were perfect and positive, the diagonal elements would be '1' and the off diagonal elements would be '0.' There is negative dependence between the relative positions of individuals in the two distributions, and the dependence is far from perfect. For example, almost 10% of those who are at the sixth decile of the *ex post* high school distribution would be in the eighth decile of the *ex post* college distribution.

Note that this comparison is not made in terms of positions in the overall distribution of earnings. We can determine where individuals are located in the population distribution of potential high school earnings and the population distribution of potential college earnings although in the data we only observe individuals in either one or the other state. The assumption of perfect dependence across factual and counterfactual distributions that is often made in the literature is incorrect for the data we analyze.

While Table 4.1 is the conditional distribution of *ex post* earnings across people, Table 4.2 presents the conditional distribution of population *ex ante* college earnings on high school earnings decile by decile. These conditional distributions are produced by allowing $\boldsymbol{X}, \theta_1, \theta_2, \varepsilon_C$ to vary across persons as they do in the population, but integrating out the unknown $\varepsilon_{s,t}$, $s = c, h$, $t = 0, \ldots, 4$, and θ_2. (In Table 4.1, these components contribute to the measured variability.) The *ex ante* conditional distribution shows less dispersion than the distribution of *ex post* outcomes since components of future realizations are integrated out. *Ex ante*, agents forecast more negative dependence across counterfactual earnings states than the *ex post* dependence on realized earnings. Realized θ_3 and the $\{\varepsilon_{s,t}\}_{t=0}^{4}$ are forces toward positive dependence. The distinction between *ex ante* and *ex post* counterfactual distributions is a major contribution of this paper and demonstrates that information revelation is an important aspect of life cycle decision making.

Our ability to distinguish *ex ante* outcomes from *ex post* outcomes highlights a major advantage of our approach over conventional instrumental variable and matching approaches to estimating returns to education which focus on *ex post* returns. Decisions are made *ex ante*. Outcomes are measured *ex post*. It is the *ex ante* return that agents act on but the *ex*

post, or realized, return that empirical economists usually measure.[28]

Let $I = \sum_{t=0}^{4} \frac{\left(Y_{c,t} - Y_{h,t}\right)}{(1+r)^t} - C$. Using our empirical model, we present three sets of estimates: (i) *Ex ante* returns based on *ex ante* choices $E_0\left(R \mid E_0\left(I\right) \geq 0\right)$ and $E_0\left(R \mid E_0\left(I\right) < 0\right)$; (ii) *Ex post* returns based on choices made with *ex ante* information $\left(R \mid E_0\left(I\right) \geq 0\right), \left(R \mid E_0\left(I\right) < 0\right)$ (what is usually presented in the literature on 'program evaluation') and (iii) *Ex post* returns based on *ex post* choices $\left(R \mid I \geq 0\right), \left(R \mid I < 0\right)$. The last set of returns conveys how returns and choices would differ if agents could 'do it over again,' *i.e.*, make decisions based on hindsight. The same people are used to form measures (i) and (ii). For measure (iii), agents are allowed to change their schooling choices with hindsight.

Figures 6.1 and 6.2 present, respectively, the fitted and counterfactual marginal distributions of *ex post* earnings for high school and college graduates. Figure 6.1 reveals that high school graduates are more likely to be successful in the high school sector than those who attend college. In Fig. 6.2, we compare the densities of present value of earnings in the college sector for persons who choose college with the counterfactual distributions of college earnings for high school graduates. The density of the present value of earnings for college graduates is to the right of the counterfactual density of the present value of earnings of high school graduates if they were college graduates. The surprising feature of both figures is that the overlap of the distributions is substantial. *Ex post*, many high school graduates would have large earnings as college graduates. This suggests the importance of costs and expectational elements in explaining schooling decisions. The densities of *ex ante* earnings is more compressed than the densities of *ex post* earnings (see Figs 6.3 and 6.4) but the

[28] As Hicks (1946, p.179) puts it, '*Ex post* calculations of capital accumulation have their place in economic and statistical history; they are a useful measuring for economic progress; but they are of no use to theoretical economists who are trying to find out how the system works, because they have no significance for conduct.'

patterns are similar reflecting the fact that most of the measured variability in earnings is due to heterogeneity. The densities under perfect certainty (Figs 6.5 and 6.6) for high school and college, respectively, show a much sharper separation between the earnings in the choice taken and the counterfactual earnings. Using hindsight, people would make wiser choices, and separate out more sharply, but there is still considerable overlap between the two distributions for both schooling choices.

Tables 5.1–5.4 provide further evidence on the importance of distinguishing between *ex ante* and *ex post* returns. In Table 5.1, we report the estimated and counterfactual *ex post* present value of earnings for agents who choose high school. The typical high school student would earn \$605.92 thousand dollars over the life cycle. She would earn \$969.34 thousand if she had chosen to be a college graduate.[29] This implies a lifetime return of 117% to a college education over the whole life cycle (*i.e.*, a monetary gain of \$363.42 thousand dollars for four years of college). In Table 5.2, we note that the typical college graduate earns \$1,007.64 thousand dollars (above the counterfactual earnings of a typical high school student), and would make only \$536.43 thousand dollars over her lifetime if she chose to be a high school graduate instead. The lifetime 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 133%, above that of the return for a high school graduate.

Table 5.3 reports the *ex post* earnings in high school and college and returns to college for people indifferent between college and high school. Not surprisingly, people on the margin of indifference have returns that are intermediate between those who go to college and those who go to high school.

[29]These numbers may appear to be large but are a consequence of using only a 3% discount rate.

Table 5.4 presents rates of return to college under different assumptions about agent information, for people who choose high school, for people who choose college and for those at the margin of indifference between going to college or not. The persons at the margin are more likely to be affected by a policy that encourages college attendance, and their returns should be used to compute the marginal benefit of policies that induce people into schooling.[30]

Ex ante and *ex post* mean returns must be the same for any subpopulation if agents use the information available to them. The mean returns under perfect certainty are different from the other returns because of resorting by persons into schooling in response to the information revealed after initial college choices are made. Some people would choose different levels of schooling if they had hindsight. Returns to college for those choosing high school in hindsight would be lower; returns to college would be higher. For those on the margin of indifference, the returns are about the same under perfect certainty as they are in the other two experiments reported in the table.

While *ex ante* and *ex post* mean returns must be identical, the *ex ante* and *ex post* distributions are not.[31] Figure 7.1 plots the density of *ex post* returns to education for agents who are high school graduates (the solid curve), and the density of *ex post* returns to education for agents who are college graduates (the dashed curve). College graduates

[30]Heckman and Vytlacil (1999, 2005) develop an alternative method for estimating the *ex post* return to persons at the margin of attending school.

[31]Let $W_1 = \mu(\eta, \nu_1)$ be the outcome in period '1.' The agent in period '0' knows (η, ν_0). The *ex ante* mean value of W_1 given η and ν_0 is

$$E_0(W_1 \mid \eta, \nu_0) = \int \mu(\eta, \nu_1)\, dF(\nu_1 \mid \eta, \nu_0)$$

where $F(a \mid b)$ is the distribution of a given b. The *ex post* mean of W_1 given (η, ν_1) is $\mu(\eta, \nu_1)$. The *ex post* mean of W_1 given (η) is $E(W_1 \mid \eta) = \int \mu(\eta, \nu_1)\, dF(\nu_1 \mid \eta)$ averaging over (η, ν_0) and $E(W_1, \eta)$ over η produces the same mean outcome. This is true for any central moment.

have returns distributed 'to the right' of high school graduates, so the difference is not only a difference for the mean individual but is actually present over the entire distribution. Agents who choose a college education are the ones who tend to gain more from it.

Figure 7.2 presents the *ex ante* returns for college and high school students. These densities are not much different from that of the *ex post* densities. Figure 7.3 shows the densities of returns for those who would choose high school and college in an environment of perfect certainty. Clearly, the distributions are more sharply separated. Uncertainty reduces the force of comparative advantage emphasized by Roy (1951).

Figure 8 shows the estimated densities of the monetary value of cost, both overall and by schooling level. College is less costly for those who attend college. 'Psychic costs' can stand in for expectational errors and attitudes towards risk. We do not distinguish among these explanations in this paper. The estimated costs are too large to be due to tuition costs alone.

It is important to note that our cost estimates are critically dependent on the assumption that the $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are known by the agent. If the agent cannot accurately forecast future prices, and the prices are random variables but statistically independent of the $\boldsymbol{\theta}$ (as would be plausible, since the prices are set in national markets and the $\boldsymbol{\theta}$ are individual specific), then what we are calling estimated costs include expectational errors (see Carneiro *et al.*, 2003).[32] In the absence of cost data, and data on expectations, this ambiguity is intrinsic and highlights the importance of maintained assumptions in interpreting evidence on schooling

[32]This is obvious from expression (2.8). If the $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are random variables from the point of view of the agent using information set $\widetilde{\mathcal{I}}_{i,0}$, and are independent of $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{\theta}$, then expectational errors enter symmetrically with cost shocks. Thus, consider the first two terms in (2.8) associated with the $\boldsymbol{X}$ and $\boldsymbol{\beta}$.

Analyzing the contribution of expectations about $\boldsymbol{\beta}$ to the total error term in the schooling choice index,

choices.

In the human capital literature, a conventional maintained assumption used when computing rates of return from measured earnings data is that direct costs are only a small fraction of total earnings (see Heckman, Lochner, and Todd, 2004). Our evidence casts doubt on the validity of this assumption. Psychic costs (including expectational forecast errors) are a sizeable component of the net return, and they explain why agents who face high gross returns do not go to college. Ignoring direct costs overstates the rates of return. The existence of large *ex post* returns that could be realized by high school students who do not attend college are attributable in our model to psychic costs and expectational errors in some unknown proportion.

---

we obtain four components

$$
\begin{aligned}
&\sum_{t=0}^{T} \frac{E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} E\left(\boldsymbol{\beta}_{1,i,t} - \boldsymbol{\beta}_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right) \\
&+ \sum_{t=0}^{T} \frac{E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} \left[\boldsymbol{\beta}_{1,i,t} - \boldsymbol{\beta}_{0,i,t} - \left[E\left(\boldsymbol{\beta}_{1,i,t} - \boldsymbol{\beta}_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]\right] \odot \boldsymbol{\Delta}_{\beta} \\
&+ \sum_{t=0}^{T} \frac{\left(\boldsymbol{X}_{i,t} - E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right)}{(1+r)^t} E\left(\boldsymbol{\beta}_{1,i,t} - \boldsymbol{\beta}_{0,j,t} \mid \widetilde{\mathcal{I}}_{i,0}\right) \odot \boldsymbol{\Delta}_{X} \\
&+ \sum_{t=0}^{T} \frac{\left(\boldsymbol{X}_{i,t} - E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right)}{(1+r)^t} \left[\boldsymbol{\beta}_{1,i,t} - \boldsymbol{\beta}_{0,i,t} - \left[E\left(\boldsymbol{\beta}_{1,i,t} - \boldsymbol{\beta}_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]\right] \odot \boldsymbol{\Delta}_{X,\beta}
\end{aligned}
$$

where, as before, $\odot$ is a Hadamard product, and $\boldsymbol{\Delta}_{\beta}$ and $\boldsymbol{\Delta}_{X,\beta}$ are defined as coefficients analogous to the coefficients used in (2.8). A comparable expression can be derived for the other coefficients if they are random. The expectational errors about the coefficients are an additional source of variability in outcomes that cannot be distinguished from variations due to the expectational errors in the $\boldsymbol{X}$ without using additional information. See the second and fourth terms and note they they would enter ε_C as we have defined it in the previous sections and would hence be conflated with costs.

5.2.4 How well can agents predict future earnings?

In Figs 9.1 through 9.3, we separate the effect of heterogeneity (total unobserved variance) from uncertainty in earnings. These calculations are reported for the population as a whole. Figure 9.1 plots the densities of the present value of earnings for the agent, using different information sets, denoted by $\mathbf{\Theta}$. First, consider the case in which the agent has no information about the $\boldsymbol{\theta}$ or the $\{\varepsilon_{0,t}, \varepsilon_{1,t}\}_{t=0}^{T}$. The $\boldsymbol{Z}, \boldsymbol{X}, \varepsilon_C$, and the model coefficients are assumed to be known in all of these simulations. They are set at mean values. The choice of means affects the locations but not the shapes of the densities. The $\varepsilon_{s,t}$ are unknown and various assumptions about which the agent knows are tested. Note that the density has a large variance, if the agent knows only factor 1, *i.e.*, the factors in the information set are $\mathbf{\Theta} = \{\theta_1\}$.[33] In this case, the reduction in the forecast from knowing ability only from knowledge of her cognitive ability adds little to the forecast of her future earnings. Now, assume that the agent is given knowledge of factor 2 as well, so that $\mathbf{\Theta} = \{\theta_1, \theta_2\}$. Note that knowledge of factor 2 causes a substantial reduction in the variance of the present value of earnings in high school. Thus, while factor 2 does not greatly affect college choices, it greatly informs the agent about her future earnings. When the agent is given knowledge of factors 1, 2, and 3, that is, $\mathbf{\Theta} = \{\theta_1, \theta_2, \theta_3\}$, she can forecast earnings somewhat better. However, our analysis suggests that agents do not know factor 3. Figure 9.2 reveals much the same story about college earnings, except that knowledge of factor 3 now substantially increases the predictability of college earnings.

Knowledge of the factors enables agents to make better forecasts of returns. Figure 9.3

[33] As opposed to the econometrician who never gets to observe the $\mathbf{\Theta}$.

presents the same type of exercise regarding information sets available to the agent for returns to college ($Y_1 - Y_0$). Knowledge of factor 2 also helps the agents forecast their gains better. Almost 48% of the variability in returns is forecastable at age 19. Knowledge of factor 3, which is not known at age 19, would greatly improve predictability of future earnings.

Table 6.1 presents the variance of potential lifetime earnings in each state, and returns under different information sets available to the agent. Tables 6.1–6.6 are calculated for the entire population. Note that in Table 6.1 knowledge of factor 2 is quantitatively important in reducing forecast variance of lifetime earnings for college and high school. Factor 3 is more powerful but, according to our estimates, it is not known by the agent at age 19. Tables 6.2–6.6 show the period by period predictability of discounted earnings from the vantage point of age 19 when the agent knows only θ_1 and θ_2. Earnings in later periods are less predictable than earnings in earlier periods using only factors one and two. Quantitatively, factors 2 and 3 are important in predicting future earnings and returns whereas ability (factor 1) is not.

This discussion sheds light on the issue of distinguishing predictable components of heterogeneity from uncertainty. We have demonstrated 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. The remaining dispersion is due to luck (uncertainty) or unforecastable errors regarding the coefficients as of age 19. Since any measurement errors in *ex post* earnings are allocated to uncertainty, our estimates arguably underestimate the degree of predictability of future earnings known to the agents at age 19.

5.2.5 *Ex ante* choices versus choices under perfect certainty

Once the distinction between heterogeneity and uncertainty is made, we can talk meaningfully about the distinction between *ex ante* and *ex post* decision making. From our analysis, we conclude that, at the time agents pick their schooling, $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ and θ_3 in their earnings equations are unknown to them. These are the components that correspond to 'luck' as defined by Jencks *et al.* (1972). It is clear that schooling choices would be different, at least for some individuals, if they knew the realized components of earnings. If agents knew these luck components when choosing schooling levels, decision rule (10)–(12) would now be

$$\begin{aligned} I &= \sum_{t=0}^{4} \frac{(Y_{c,t} - Y_{h,t})}{(1+r)^t} - C > 0 \\ S &= 1 \text{ if } I > 0;\ S = 0 \text{ otherwise,} \end{aligned}$$

where no expectation is taken to calculate I since all components are known with certainty by the agents.

In our empirical model, if individuals could pick their schooling level using their *ex post* information (*i.e.*, after learning their luck components in earnings), 25.19% of high school graduates would rather be college graduates and 31.40% of college graduates would have stopped at the high school level. Uncertainty about future outcomes greatly affects schooling choices, and there is plenty of scope for *ex post* regret.[34]

[34]In a companion paper (Cunha *et al.*, 2005), we address issues similar to the ones addressed in this paper but use a more *ad hoc* approach to pooling data across samples to construct a life cycle data set. That procedure follows Carneiro *et al.* (2003) rather than the more rigorous methodology derived in Appendix 2. That paper shows even less uncertainty than we have shown here and establishes a strong correlation across latent skill levels, which is positive. We are much more confident in the empirical results in this paper than

6 Summary and conclusions

This paper discusses the problem of separating heterogeneity from uncertainty. We develop and apply a method for estimating both heterogeneity and uncertainty from *ex post* earnings data and from schooling choices. We estimate substantial predictable and unpredictable components of earnings as of age 19. Agents have greater difficulty in predicting outcomes in later periods of their life cycles than they do in earlier periods. Procedures that equate variability with uncertainty overstate risk and, hence, understate the pricing of risk. If agents knew their *ex post* earnings outcomes resulting from their schooling choices, a substantial fraction (around 30%) would change their schooling decisions. Hicks' distinction between *ex ante* and *ex post* is an empirically important one.

This paper takes a first step toward resolving an empirical puzzle in the labor economics literature. *Ex post* returns to college are high for those who stop at high school. Our evidence is that, within a complete markets setting, psychic costs of schooling (and expectational errors in a more general model) account for this phenomenon. This evidence has importance implications for the conventional human capital literature that ignores these costs in computing rates of return to schooling. However, a story that relies on psychic costs to explain the puzzle is not entirely satisfactory. One needs to account more systematically for borrowing constraints and risk aversion, and we do so elsewhere in Carneiro *et al.* (2003), Cunha *et al.* (2004), and Navarro (2004).

Throughout this paper we have maintained the assumption of complete markets for idiosyncratic components of risk. An open question which we address, but do not solve, is how

in the results reported in the previous paper.

to simultaneously identify constraints (market structure), preferences and information confronting agents. Different scholars focus on different aspects of the decision problem facing agents. Those who postulate specific information structures and the preferences of agents test for alternative market structures (*e.g.*, partial insurance). In this paper, we have estimated information structures, making assumptions about market structures and constraints that neutralize the effects of risk preferences and uncertainty on schooling choices.

In Cunha *et al.* (2004), we build on the analysis of this paper to estimate an Aiyagari (1994) – Laitner (1992) economy and simultaneously identify preferences (within a parametric family) and information sets allowing for market incompleteness. We extend the analysis of Carneiro *et al.* (2003) by considering more flexible parameterizations of preferences against risk aversion and allowing for restricted lending and borrowing. (They assume an environment of complete autarky). A robust finding across all environments we have studied is that uncertainty is empirically important. Hicks' important distinction between *ex ante* and *ex post* income receives substantial empirical support in the data on schooling choice and earnings, and changes the way we interpret a vast empirical literature on *ex post* returns to schooling.

Acknowledgements

We have greatly benefitted from comments on previous drafts received from Lars Hansen, David Hendry, John Muellbauer, Robert Townsend, and participants at the Gorman Memorial Conference, London, June 2004. Martin Browning, Lars Hansen, Annette Vissing-Jorgenson, and Sergio Urzua provided very helpful comments on this revision. Paul Schrimpf

provided excellent research assistance and made many useful comments on both drafts. We thank two anonymous referees for their comments on this paper and Jennifer Boobar and Weerachart Kilenthong for very close and helpful readings of this paper.

This paper was presented as the Hicks Lecture at Oxford University on April 27, 2004. This research was supported by NIH R01-HD043411 and NSF grant SES-0241858. Cunha acknowledges support from CAPES grant 1430/99-8. Navarro received support from Conacyt, Mexico and from the George G. Stigler and the Esther and T.W. Schultz fellowships at the University of Chicago.

Appendix 1 A motivation for the nonparametric identification of the joint distribution of outcomes and the binary choice equation

The following intuition motivates conditions under which $F(Y_0, I^* \mid \boldsymbol{X}, \boldsymbol{Z})$ is identified. A formal proof is given in Carneiro *et al.* (2003). A parallel argument holds for $F(Y_1, I^* \mid \boldsymbol{X}, \boldsymbol{Z})$. First, under the conditions given in Cosslett (1983), Manski (1988), and Matzkin (1992), we can identify $\frac{\mu_I(\boldsymbol{X},\boldsymbol{Z})}{\sigma_I}$ from $\Pr(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}) = \Pr(\mu_I(\boldsymbol{X}, \boldsymbol{Z}) + U_I \geq 0 \mid \boldsymbol{X}, \boldsymbol{Z})$. We can also identify the distribution of $\frac{U_I}{\sigma_I}$.[35] Second, from this information and $F(Y_0 \mid S = 0, \boldsymbol{X}, \boldsymbol{Z}) =$

[35] An alternative to the conventional approach, which requires large support conditions, postulates that $\mu_I(\boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{\gamma}_{\boldsymbol{X}} + \boldsymbol{Z}\boldsymbol{\gamma}_{\boldsymbol{Z}}$ and normalizes one coefficient on a continuous coordinate of $\boldsymbol{Z}$, say Z_1, to unity (*e.g.*, $\boldsymbol{\gamma}_{Z_1} = 1$). Then, fixing the remaining values of $\boldsymbol{X}$ and $\boldsymbol{Z}$ at specified values ($\boldsymbol{X} = x$, $\hat{\boldsymbol{Z}} = \hat{z}$, where $\hat{\boldsymbol{Z}}$ is $\boldsymbol{Z}$ removed of its first coordinate) and tracing Z_1 over its support, we identify the distribution of U_I over the support of Z_1, assumed to lie in an interval $[C_L, C_U)$ which may or may not be the support of U_I.

$\Pr(Y_0 \leq y_0 \mid \mu_I(\boldsymbol{X}, \boldsymbol{Z}) + U_I \leq 0, \boldsymbol{X}, \boldsymbol{Z})$, we can form

$$F(Y_0 \mid S = 0, \boldsymbol{X}, \boldsymbol{Z}) \Pr(S = 0 \mid \boldsymbol{X}, \boldsymbol{Z}) = \Pr(Y_0 \leq y_0, I^* \leq 0 \mid \boldsymbol{X}, \boldsymbol{Z}).$$

The left hand side of this expression is known (we observe Y_0 when $S = 0$ and we know the probability that $S = 0$ given $\boldsymbol{X}, \boldsymbol{Z}$). The right hand side can be written as

$$\Pr\left(Y_0 \leq y_0, \frac{U_I}{\sigma_I} \leq -\frac{\mu_I(\boldsymbol{X}, \boldsymbol{Z})}{\sigma_I} \mid \boldsymbol{X}, \boldsymbol{Z}\right).$$

We know $\frac{\mu_I(\boldsymbol{X}, \boldsymbol{Z})}{\sigma_I}$ and can vary it for each fixed $\boldsymbol{X}$. In particular if $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ gets small ($\mu_I(\boldsymbol{X}, \boldsymbol{Z}) \to -\infty$) we can recover the marginal distribution Y_0 from which we can recover $\mu_0(\boldsymbol{X})$.

Using (9a), we can express this probability as

$$\Pr\left(U_0 \leq y_0 - \mu_0(\boldsymbol{X}), \frac{U_I}{\sigma_I} \leq \frac{-\mu_I(\boldsymbol{X}, \boldsymbol{Z})}{\sigma_I} \mid \boldsymbol{X}, \boldsymbol{Z}\right).$$

Assuming U_I is absolutely continuous, we can thus identify

$$F_{U_I}\left(u_I \mid C_L \leq U_I < C_U\right) = \frac{F_{U_I}(u_I)}{F_{U_I}(C_U) - F_{U_I}(C_L)}.$$

Since $\Pr\left(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}\right) = F_{U_I}\left(\boldsymbol{X}\boldsymbol{\gamma}_{\boldsymbol{X}} + \boldsymbol{Z}\boldsymbol{\gamma}_{\boldsymbol{Z}}\right)$, if the supports of Z_1 and U_I match, we can invert for each $\boldsymbol{X}$, $\boldsymbol{Z}$

$$F_{U_I}^{-1}\left(\Pr\left(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}\right)\right) = \boldsymbol{X}\boldsymbol{\gamma}_X + \boldsymbol{Z}\boldsymbol{\gamma}_Z$$

and identify the coefficients $\boldsymbol{\gamma}_X$, $\boldsymbol{\gamma}_Z$ provided that $(\boldsymbol{X}, \boldsymbol{Z})$ is of full rank. However, if the support of U_I strictly contains that of Z_1, the same operation identifies

$$F_{U_I}^{-1}\left(\frac{\Pr\left(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}\right)}{F_{U_I}(C_U) - F_{U_I}(C_L)}\right) = \boldsymbol{X}\boldsymbol{\gamma}_X + \boldsymbol{Z}\boldsymbol{\gamma}_Z$$

where $F_{U_I}(C_U) - F_{U_I}(C_L)$ is unknown.

Note that $\boldsymbol{X}$ and $\boldsymbol{Z}$ can be varied and y_0 is a number. Thus we can trace out the joint distribution of $\left(U_0, \frac{U_I}{\sigma_I}\right)$. Thus we can recover the joint distribution of

$$(Y_0, I^*) = \left(\mu_0(\boldsymbol{X}) + U_0, \frac{\mu_I(\boldsymbol{X}, \boldsymbol{Z}) + U_I}{\sigma_I}\right).$$

Notice the three key ingredients. (i) The independence of (U_0, U_I) and $(\boldsymbol{X}, \boldsymbol{Z})$. (ii) The assumption that we can set $\frac{\mu_I(\boldsymbol{X},\boldsymbol{Z})}{\sigma_I}$ to be very small (so we get the marginal distribution of Y_0 and hence $\mu_0(\boldsymbol{X})$). (iii) The assumption that $\frac{\mu_I(\boldsymbol{X},\boldsymbol{Z})}{\sigma_I}$ can be varied independently of $\mu_0(\boldsymbol{X})$. This enables us to trace out the joint distribution of $\left(U_0, \frac{U_I}{\sigma_I}\right)$.

Appendix 2 Combining data sets to estimate a life cycle model

A serious empirical problem plagues most life cycle analyses. It is a rare data set that includes the full life cycle earnings experiences of persons along with their test scores, measurements, schooling choices, and background variables. Many data sets like the National Longitudinal Survey of Youth (NLSY 79) have partial information up to some age. A few other data sets (*e.g.*, the Panel Survey on Income Dynamics or PSID) have full information on some life cycle variables but lack the detail of the richer data which provide information only on truncated life cycles. This section considers two issues: (i) What can be identified from the truncated life cycle data and (ii) What can be learned from combining the truncated data with a data set with fewer variables but with information on schooling and earnings on entire life cycles? Our factor model provides a natural framework for combining samples to

produce identification even when the model is not identified in each sample.

To fix ideas and motivate the empirical work, suppress the individual i subscripts and write

$$Y_{s,t} = \boldsymbol{X}\boldsymbol{\beta}_{s,t} + \theta_1\alpha_{s,t,1} + \theta_2\alpha_{s,t,2} + \theta_3\alpha_{s,t,3} + \varepsilon_{s,t}, \qquad t = 0, \ldots, 4, \quad s = 0, 1, \tag{17}$$

where $\alpha_{s,t,3} = 0$ for $t = 0, 1$. An individual picks $S = 1$ if

$$\sum_{t=0}^{4} \frac{1}{(1+r)^{t-1}} E\left(Y_{1,t} - Y_{0,t} \mid \mathcal{I}_0\right) - E\left(Cost \mid \mathcal{I}_0\right) > 0,$$

that is $S = 1$ if

$$\sum_{t=0}^{4} \frac{1}{(1+r)^{t-1}} \left[E\left(\boldsymbol{X} \mid \mathcal{I}_0\right)\left(\boldsymbol{\beta}_{1,t} - \boldsymbol{\beta}_{0,t}\right) + \sum_{j=1}^{3} E\left(\theta_j \mid \mathcal{I}_0\right)\left(\alpha_{1,t,j} - \alpha_{0,t,j}\right) + E\left(\varepsilon_{1,t} - \varepsilon_{0,t} \mid \mathcal{I}_0\right) \right]$$
$$- E\left(\boldsymbol{Z} \mid \mathcal{I}_0\right)\boldsymbol{\gamma} - \sum_{j=1}^{3} E\left(\theta_j \mid \mathcal{I}_0\right)\left(\alpha_{C,j}\right) - E\left(\varepsilon_C \mid \mathcal{I}_0\right) \geq 0,$$

where $\boldsymbol{Z}$ may include elements in common with $\boldsymbol{X}$. It will prove convenient to write the choice equation in reduced form letting $\boldsymbol{Q}$ combine $\boldsymbol{X}$ and $\boldsymbol{Z}$:

$$I = \boldsymbol{Q}\boldsymbol{\gamma}_I + \theta_1\alpha_{I,1} + \theta_2\alpha_{I,2} + \theta_3\alpha_{I,3} + \varepsilon_I, \tag{18}$$

where ε_I is the composite of the errors from the choice equation. Finally, the external

measurements are written as

$$M_k = \boldsymbol{X}_M \boldsymbol{\beta}_{M,k} + \theta_1 \alpha_{M,k,1} + \varepsilon_{M,k}, \qquad k = 1, \ldots, K,$$

where K is the number of measurements (test scores in our application). For the case in which we have access to full life cycle data, the contribution to the likelihood of an individual who chooses $S = s$, is given by

$$\int_{\underline{\Theta}} \prod_{t=0}^{4} \prod_{s=0}^{1} \left\{ f\left(Y_{s,t} | \boldsymbol{\theta}, \boldsymbol{X}\right) Pr\left(S = s | \boldsymbol{Z}, \boldsymbol{\theta}\right) \right\}^{1(S=s)} \prod_{k=1}^{K} f\left(M_k | \boldsymbol{\theta}, \boldsymbol{X}_M\right) dF\left(\boldsymbol{\theta}\right), \tag{19}$$

where it is assumed that the $(\boldsymbol{X}, \boldsymbol{Z})$ are independent of $\boldsymbol{\theta}$ and the ε and $\underline{\Theta}$ is the support of $\boldsymbol{\theta}$. Identification follows from the analysis of Carneiro *et al.* (2003).

Now, suppose that we only have access to a sample A in which we only observe some of the variables at early stages of the life cycle. In particular, assume that sample A does not include observations on $\{Y_{s,t}\}_{t=2}^{4}$ as is the case with the NLSY, which contains no information on earnings after age 43.

The contribution to the likelihood of an individual who chooses, for example, $S = 1$ is

$$\begin{aligned} & \int_{\underline{\Theta}} \left[\prod_{t=0}^{1} f\left(Y_{1,t} | \boldsymbol{\theta}, \boldsymbol{X}\right) \right] \left[Pr\left(I \geq 0 | \boldsymbol{Z}, \boldsymbol{\theta}\right) \right] \prod_{k=1}^{K} f\left(M_k | \boldsymbol{\theta}, \boldsymbol{X}_M\right) \left\{ \prod_{t=2}^{4} \int f\left(Y_{1,t} | \boldsymbol{\theta}, \boldsymbol{X}\right) dF\left(Y_{1,t}\right) \right\} dF\left(\boldsymbol{\theta}\right) \\ = & \int_{\underline{\Theta}} \left[\prod_{t=0}^{1} f\left(Y_{1,t} | \boldsymbol{\theta}, \boldsymbol{X}\right) \right] Pr\left(I \geq 0 | \boldsymbol{Z}, \boldsymbol{\theta}\right) \prod_{k=1}^{K} f\left(M_k | \boldsymbol{\theta}, \boldsymbol{X}_M\right) dF\left(\boldsymbol{\theta}\right).^{36} \end{aligned} \tag{20}$$

We integrate out earnings for the periods in which we do not observe them. Using the first

[36] If she had chosen $S = 0$ then we would write $Pr\left(I < 0 | \boldsymbol{Z}, \theta\right)$ instead.

two periods of data, we can identify a model in which we have K external measurements, 2 time periods for earnings, and a reduced form schooling equation that combines parameters of earnings with cost parameters. For $K \geq 3$, from the measurements conditional on $\boldsymbol{X}$ we can form

$$Cov\left(M_k, M_{k'} \mid \boldsymbol{X}_M\right) = \alpha_{M,k}\alpha_{M,k'}\sigma^2_{\theta_1}.\ ^{37}$$

Taking ratios of these covariances, we can identify the factor loadings up to one normalization.[38] We can identify the distributions of the error terms for measurements. From these, we can identify the distributions of $\theta_1, \{\varepsilon_k\}_{k=1}^K$ nonparametrically by using Kotlarski's theorem. Then, under the support assumptions in Carneiro *et al.* (2003), and noting that we have identified $\sigma^2_{\theta_1}$, we can form

$$Cov\left(M_k, Y_{s,t} \mid \boldsymbol{X}_M, \boldsymbol{X}\right) = \alpha_{M,k}\alpha_{s,t,1}\sigma^2_{\theta_1}, \quad s = 0, 1,\ t = 0, 1,$$

and identify the loadings on the first factor for each s for $t = 0, 1$.[39] From the covariances of $\frac{I}{\sigma_I}$ with either M or $Y_{s,t}$, we can identify the factor loadings associated with (18) up to scale σ_I.

Once we identify all of the parameters related to θ_1 we can, for each schooling level s

[37] Or if $K \geq 2$ and we use either the choice equation or one of the earnings equations. See Carneiro *et al.* (2003).

[38] The means of the functions (and so the β_k) are trivially identified from $E\left(\theta_1\right) = E\left(\theta_2\right) = E\left(\varepsilon_{s,t}\right) = 0$.

[39] As before, given the support conditions the means are identified from the mean zero assumptions on the error term.

(remember that $\alpha_{s,t,3} = 0$ for $t = 0, 1$), form

$$\begin{aligned} Cov\left(Y_{s,0}, Y_{s,1} \mid \boldsymbol{X}\right) - \alpha_{s,0,1}\alpha_{s,1,1}\sigma^2_{\theta_1} &= \alpha_{s,0,2}\alpha_{s,1,2}\sigma^2_{\theta_2} \\ Cov\left(Y_{s,t}, I \mid \boldsymbol{X}, \boldsymbol{Z}\right) - \alpha_{s,t,1}\alpha_{I,1}\sigma^2_{\theta_1} &= \alpha_{s,t,2}\alpha_{I,2}\sigma^2_{\theta_2}, \quad t = 0, 1, \end{aligned}$$

where the left hand side is known and the loadings on factor 1 are identified up to scale from earnings and choice. Recall that the third factor does not enter the earnings equations for $t = 0, 1$. We can then solve for the loadings on θ_2 in the earnings and choice equations. Proceeding as before, we can recover the distributions of θ_2, and $\left\{\left\{\varepsilon_{s,t}\right\}_{s=0}^{1}\right\}_{t=0}^{1}$ provided we have at least one exclusion (one continuous element of $\boldsymbol{Z}$ not in $\boldsymbol{X}$). Notice that, since we are not able to identify any of the parameters of earnings for $t > 1$, we cannot identify all of the structural parameters in the choice equation so we cannot separate the effect of costs from the effect of future earnings.

Now, suppose that we have access to a second independent sample B that is generated by the same process that generates sample A.[40] In this second sample, we do not observe $\{M_k\}_{k=1}^{K}$ but we do observe earnings and schooling choices (and $\boldsymbol{X}$ and $\boldsymbol{Z}$) for all time periods. For sample B, an individual with $S = 1$ has a contribution to the likelihood that would be given by integrating out the test scores from the likelihood (19):

$$\begin{aligned} & \int_{\underline{\Theta}} \left[\prod_{t=0}^{4} f\left(Y_{1,t}|\boldsymbol{\theta}, \boldsymbol{X}\right)\right] Pr\left(I \geq 0|\boldsymbol{Z}, \boldsymbol{\theta}\right) \left\{\prod_{k=1}^{K} \int f\left(M_k|\boldsymbol{\theta}, \boldsymbol{X}_M\right) dF\left(M_k\right)\right\} dF\left(\boldsymbol{\theta}\right) \\ = & \int_{\underline{\Theta}} \left[\prod_{t=0}^{4} f\left(Y_{1,t}|\boldsymbol{\theta}, \boldsymbol{X}\right)\right] Pr\left(I \geq 0|\boldsymbol{Z}, \boldsymbol{\theta}\right) dF\left(\boldsymbol{\theta}\right). \end{aligned} \tag{21}$$

[40] By this we mean that the parameters and distributions of the implied random variables of both samples are the same.

From this sample alone we cannot recover the loadings or the marginal distributions of $\theta_1, \theta_2, \{\varepsilon_{1,t}\}_{t=0}^4$, and ε_I, without additional assumptions.[41]

We combine both samples so that a person's contribution to likelihood is given by (20) if an individual comes from sample A and is given by (21) if he comes from sample B. In this case, we would be able to recover all of the elements of the model. To see why, notice that from sample A alone the only unidentified parameters are the coefficients and distributions for earnings in $t > 1$. In sample B we can form the left hand sides of

$$\begin{aligned} Cov\left(Y_{s,t}, Y_{s,0} \mid \boldsymbol{X}\right) &= \alpha_{s,t,1}\alpha_{s,0,1}\sigma^2_{\theta_1} + \alpha_{s,t,2}\alpha_{s,0,2}\sigma^2_{\theta_2} \\ Cov\left(Y_{s,t}, Y_{s,1} \mid \boldsymbol{X}\right) &= \alpha_{s,t,1}\alpha_{s,1,1}\sigma^2_{\theta_1} + \alpha_{s,t,2}\alpha_{s,1,2}\sigma^2_{\theta_2}, \qquad t = \{2,3,4\}, \end{aligned}$$

where all parameters except $\alpha_{s,t,1}$ and $\alpha_{s,t,2}$ for $t = 2, 3, 4$ are identified from data on sample A. These covariances form a system of two linear equations in two unknowns that, under a standard rank condition, we can solve for the unknowns $\alpha_{s,t,1}$ and $\alpha_{s,t,2}$ for $t > 1$ and $s = 0, 1$. A similar argument allows us to recover the parameters associated with θ_3 using the covariances of the outcomes $Y_{s,t}$ after period 1. Since we have identified all of the parameters of the earnings equations, we can solve for the structural parameters of the choice equation and separate costs from future earnings. More generally, we can obtain more efficient estimates for the overidentified parameters by pooling samples.

This procedure abstracts from cohort effects for the coefficients and factor loadings, and cohort effects for the distributions of $\boldsymbol{\theta}$. With additional structure (*e.g.*, additivity), we can

[41] It is clear we will never recover any of the parameters of the measurement equations in this sample. If we changed our normalizations on the rest of the system however, so that θ_2 does not enter the earnings equation at $t = 0, 1$ for example, and there was no third factor, we could recover all of the remaining parameters of the model.

identify such effects, but we acknowledge that general cohort effects can dramatically bias estimates based on pooling the data.

References

[1] **Aiyagari, S.R.** (1994). 'Uninsured idiosyncratic risk and aggregate saving.' *The Quarterly Journal of Economics,* **109**, 659–84.

[2] **Blundell, R., Pistaferri, L., and Preston, I.** (2004). 'Consumption inequality and partial insurance,' IFS Working Paper, October 2004 WP04/28, Institute for Fiscal Studies, London, England.

[3] **Blundell, R. and Preston, I.** (1998). 'Consumption inequality and income uncertainty.' *The Quarterly Journal of Economics,* **113**, 603–40.

[4] **Browning, M., Hansen, L., and Heckman, J.** (1999). 'Micro data and general equilibrium models,' in J. B. Taylor and M. Woodford, (eds) *Handbook of Macroeconomics,*North-Holland, Amsterdam, **1A**, 543-633.

[5] **Card, D.** (2001). 'Estimating the return to schooling: Progress on some persistent econometric problems.' *Econometrica,* **69**, 1127–60.

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

[7] **Carneiro, P., Hansen, K., and Heckman, J.** (2003). '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**, 361–422.

[8] **Carneiro, P., Heckman, J., and Vytlacil, E.** (2004). 'Estimating the return to education when it varies among individuals.' Unpublished manuscript, University of Chicago Department of Economics.

[9] **Chernozhukov, V. and Hansen, C**. (2005) 'An IV model of quantile treatment effects.' forthcoming, *Econometrica*.

[10] **Cosslett, S.R.** (1983). 'Distribution-free maximum likelihood estimator of the binary choice model.' *Econometrica,* **51**, 765–82.

[11] **Cunha, F., Heckman, J., and Navarro, S.** (2004). 'Separating heterogeneity from uncertainty an Aiyagari-Laitner economy,' Paper presented at the Goldwater Conference on Labor Markets in Arizona, March, 2004.

[12] **Cunha, F., Heckman, J., and Navarro, S.** (2005). 'Counterfactual analysis of inequality and social mobility,' in G. Fields *et al.* (eds), *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, Stanford University Press, Palo Alto.

[13] **Ferguson, T.S.** (1983). 'Bayesian density estimation by mixtures of normal distributions,' in M. Rizvi, J. Rustagi, and D. Siegmund (eds), *Recent Advances in Statistics,* Academic Press, New York, 287–302.

[14] **Flavin, M.** (1981). 'The adjustment of consumption to changing expectations about future income.' *Journal of Political Economy,* **89**, 974–1009.

[15] **Hartog, J. and Vijverberg, W.** (2002). 'Do wages really compensate for risk aversion and skewness affection?,' *IZA Working Paper DP 426*, IZA, Bonn.

[16] **Hause, J.** (1978). 'The fine structure of earnings and the on-the-job training hypothesis.' *Econometrica*, **48**, 1013–29.

[17] **Heckman, J.** (1990), 'Varieties of selection bias.' *American Economic Review*, **80**(2), 313–18.

[18] **Heckman, J.** (2001). 'Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture.' *Journal of Political Economy,* **109**, 673–748.

[19] **Heckman, J. and Honoré, B. (**1990). 'The empirical content of the Roy model.' *Econometrica,* **58**, 1121–49.

[20] **Heckman, J., Lochner, L., and Todd, P.** (2004). 'Earnings functions and rates of return: the Mincer equation and beyond.' unpublished manuscript, University of Chicago, 2004.

[21] **Heckman, J., Matzkin, R., Navarro, S., and Urzua, S.** (2004). 'Nonseparable factor analysis,' Unpublished manuscript, University of Chicago, July.

[22] **Heckman, J. and Navarro, S.** (2004). 'Using matching, instrumental variables and control functions to estimate economic choice models.' *The Review of Economics and Statistics,* **86**, 30–57.

[23] **Heckman, J. and Robb, R.** (1986, reprinted 2000). '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,* Springer-Verlag, New York, 63–107.

[24] **Heckman, J. and Scheinkman, J.** (1987). 'The importance of bundling in a Gorman-Lancaster model of earnings.' *Review of Economic Studies,* **54**, 243–55.

[25] **Heckman, J. and Smith, J.** (1998). 'Evaluating the welfare state,' in S. Strom (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial,* Econometric Society Monograph Series, Cambridge University Press, Cambridge.

[26] **Heckman, J., Smith, J., and Clements, N.** (1997). 'Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts.' *Review of Economic Studies,* **64**, 487–535.

[27] **Heckman, J. and Vytlacil, E.** (1999). 'Local instrumental variables and latent variable models for identifying and bounding treatment effects.' *Proceedings of the National Academy of the Sciences,* **96**, 4730–4.

[28] **Heckman, J. and Vytlacil, E.** (2005). 'Structural equations, treatment effects and econometric policy evaluation.' forthcoming, *Econometrica*.

[29] **Hicks, J.** (1946). *Value and Capital* (2nd ed.), Clarendon Press, Oxford.

[30] **Jencks, C.S., Smith, M., Acland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B., and Michelson, S.** (1972). *Inequality: A Reassessment of the Effect of Family and Schooling in America,* Basic Books, New York.

[31] **Jöreskog, K.** (1977). 'Structural equations models in the social sciences: Specification, estimation and testing,' in P.R. Krishnaih (ed.) *Applications of Statistics,* North-Holland, Amsterdam, 265–87.

[32] **Jöreskog, K. and Goldberger, A.S.** (1975). 'Estimation of a model with multiple indicators and multiple causes of a single latent variable.' *Journal of the American Statistical Association,* **70**, 631–39.

[33] **Kotlarski, I.** (1967). 'On characterizing the gamma and normal distribution.' *Pacific Journal of Mathematics,* **20**, 729–38.

[34] **Laitner, J**. (1992). 'Random earnings differences, lifetime liquidity constraints, and altruistic intergenerational transfers.' *Journal of Economic Theory,* **58**, 135–70.

[35] **Lillard, L. and Willis, R.** (1978). 'Dynamic aspects of earning mobility.' *Econometrica,* **46**, 985–1012.

[36] **MaCurdy, T.** (1982). 'The use of time series processes to model the error structure of earnings in a longitudinal data analysis.' *Journal of Econometrics,* **18**, 83–114.

[37] **Manski, C.** (1988). 'Identification of binary response models.' *Journal of the American Statistical Association,* **83**, 729–38.

[38] **Matzkin, R.** (1992). 'Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models.' *Econometrica,* **60**, 239–70.

[39] **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.

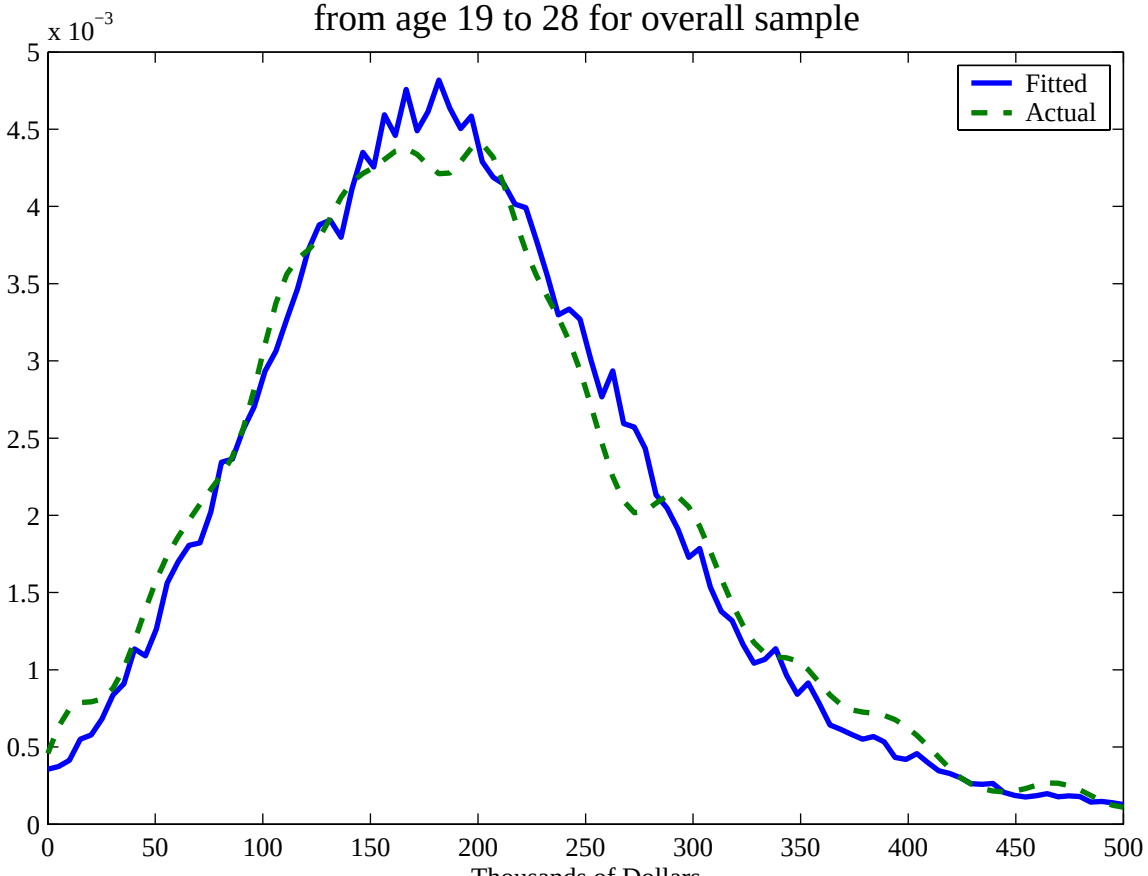
[40] **Pistaferri, L.** (2001), 'Superior information, income shocks, and the permanent income hypothesis.' *Review of Economics and Statistics,* **83**, 465–76.

[41] **Roy, A.** (1951). Some thoughts on the distribution of earnings.' *Oxford Economic Papers,* **3**, 135–46.

[42] **Shaikh, A. M. and Vytlacil, E.** (2004). 'Threshold crossing models and bounds on treatment effects: A nonparametric analysis' forthcoming, *Journal of Econometrics.*

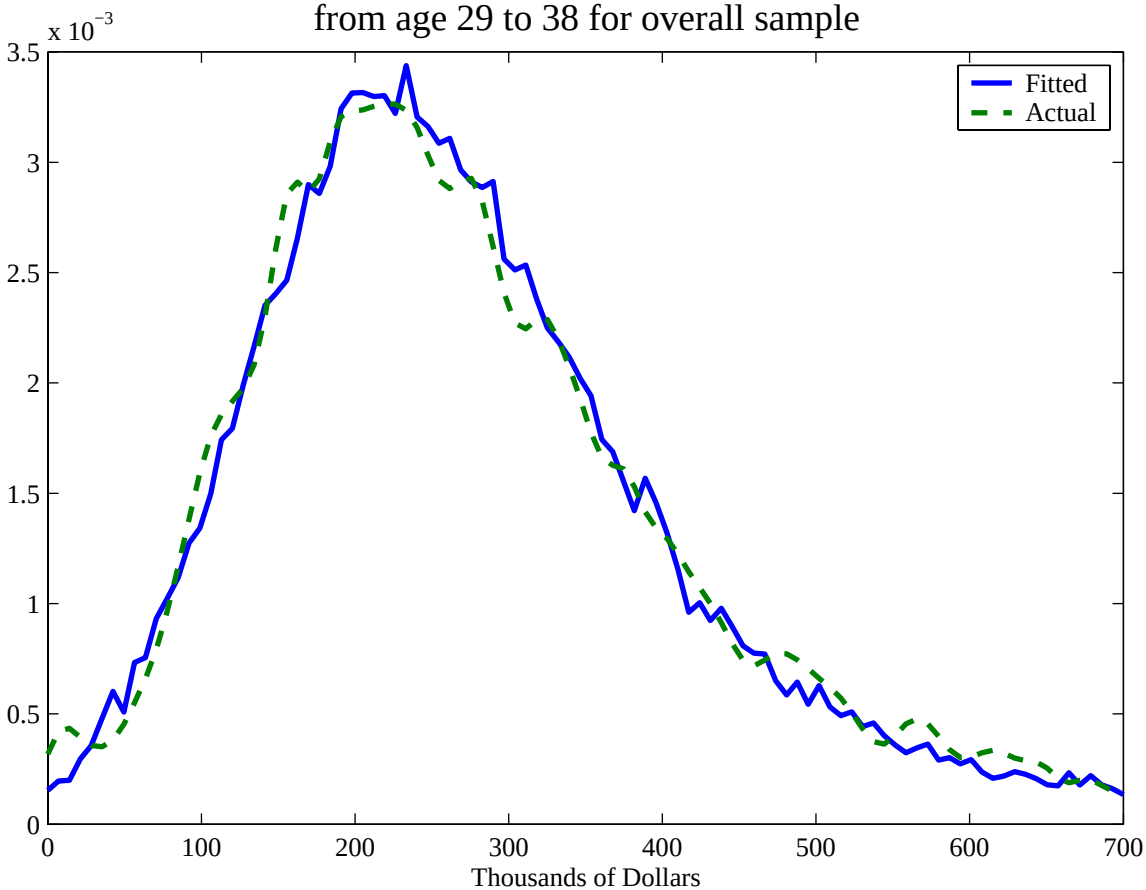
[43] **Sims, C.A.** (1972). 'Money, income, and causality.' *American Economic Review,* **62**, 540–52.

Figure 1.1
Densities of fitted and actual present value of earnings
from age 19 to 28 for overall sample



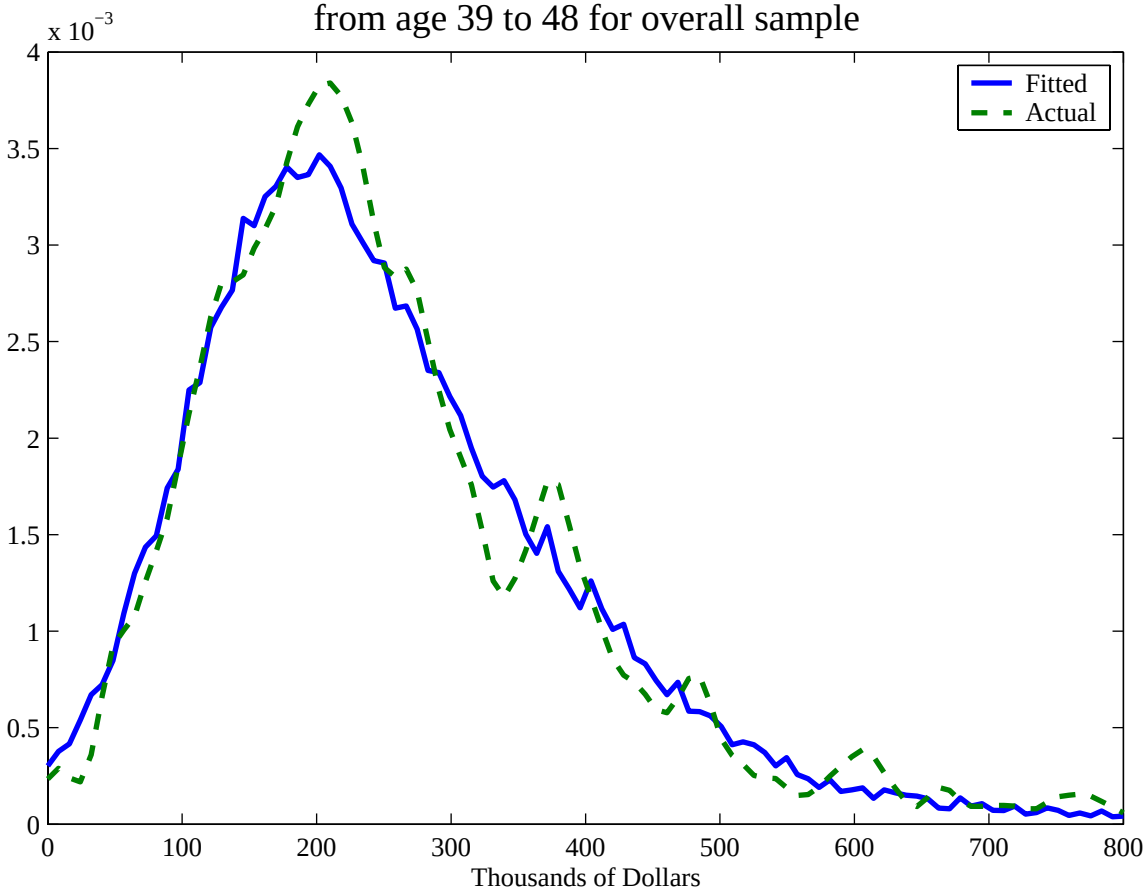
Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote choice of the high school sector, and S=1 denote choice of the college sector. Define observed earnings as $Y = SY_1 + (1 - S)Y_0$. Finally, let f(y) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 1.2
Densities of fitted and actual present value of earnings
from age 29 to 38 for overall sample



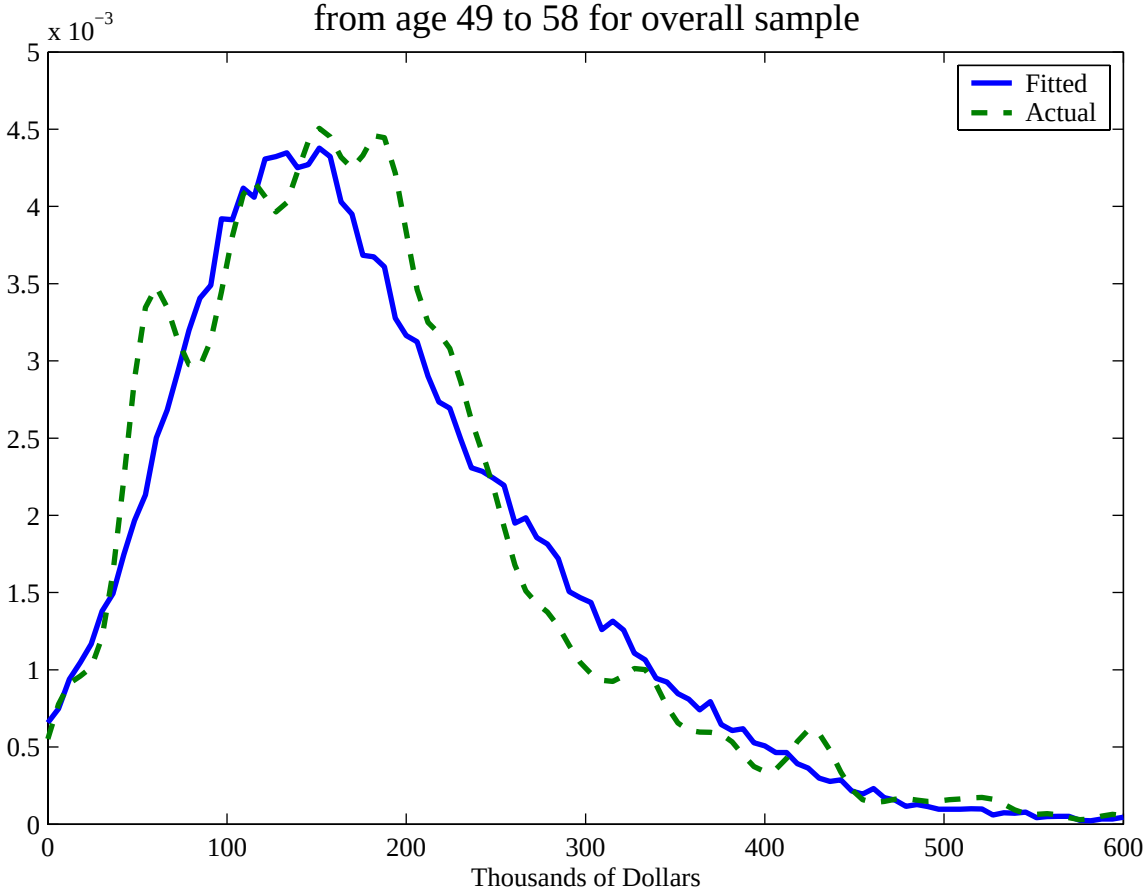
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote choice of the high school sector, and S=1 denote choice of the college sector. Define observed earnings as $Y = SY_1 + (1 - S)Y_0$. Finally, let $f(y)$ denote the density function of observed earnings. Here we plot the density functions f generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 1.3
Densities of fitted and actual present value of earnings
from age 39 to 48 for overall sample



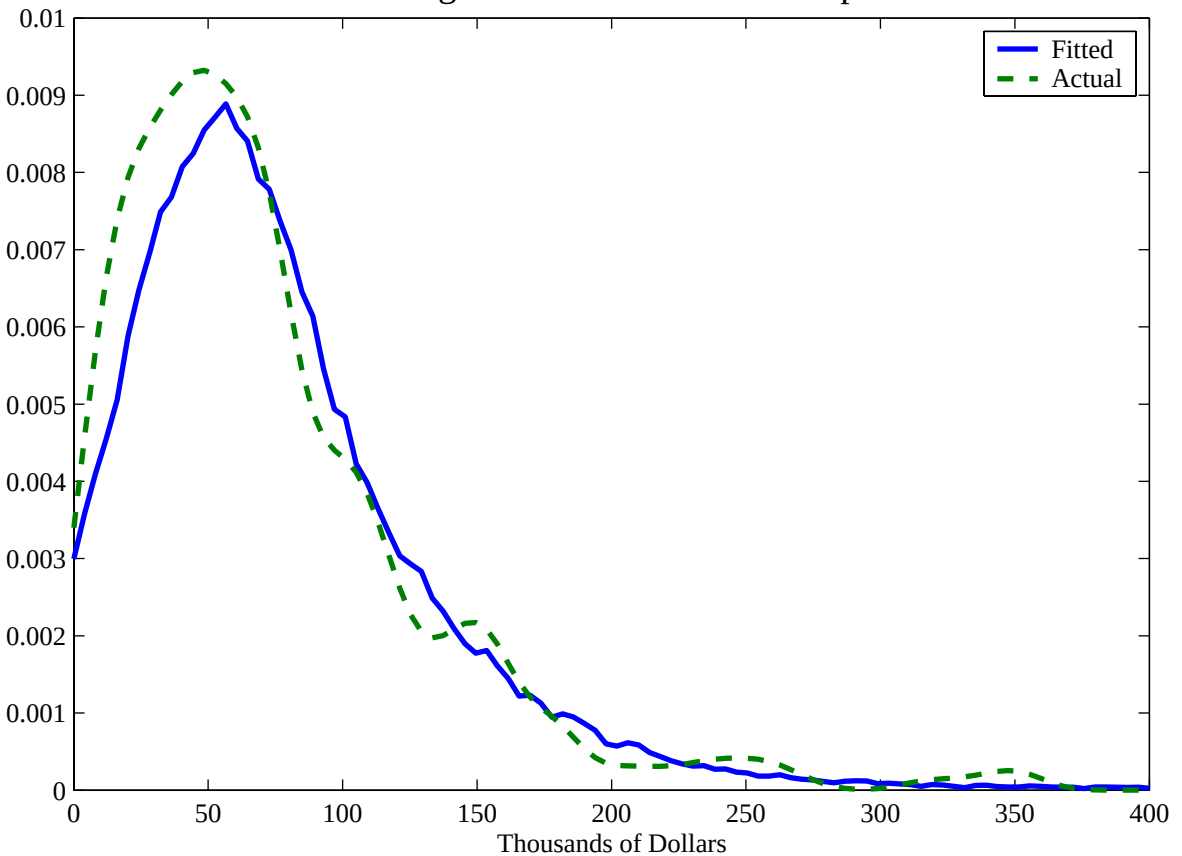
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote choice of the high school sector, and S=1 denote choice of the college sector. Define observed earnings as $Y = SY_1 + (1 - S)Y_0$. Finally, let $f(y)$ denote the density function of observed earnings. Here we plot the density functions f generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 1.4
Densities of fitted and actual present value of earnings
from age 49 to 58 for overall sample



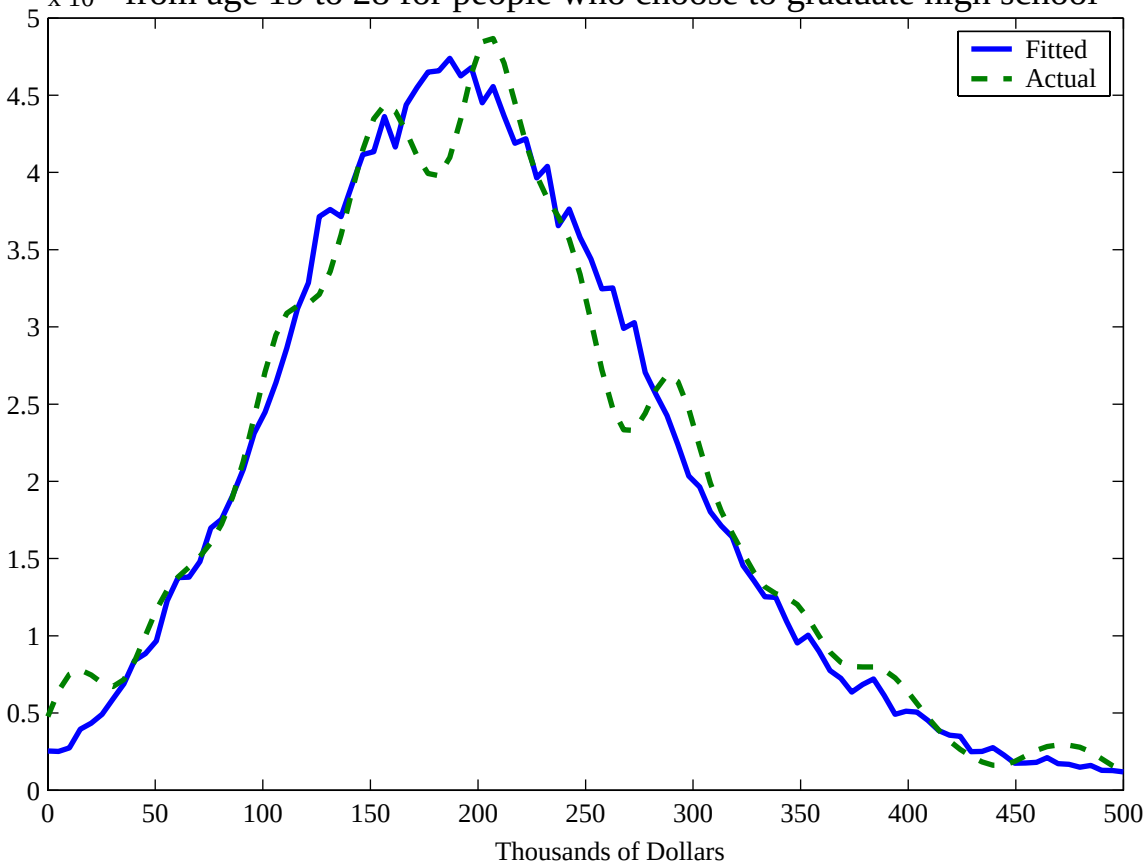
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote choice of the high school sector, and S=1 denote choice of the college sector. Define observed earnings as $Y = SY_1 + (1 - S)Y_0$. Finally, let $f(y)$ denote the density function of observed earnings. Here we plot the density functions f generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 1.5
Densities of fitted and actual present value of earnings
from age 59 to 65 for overall sample



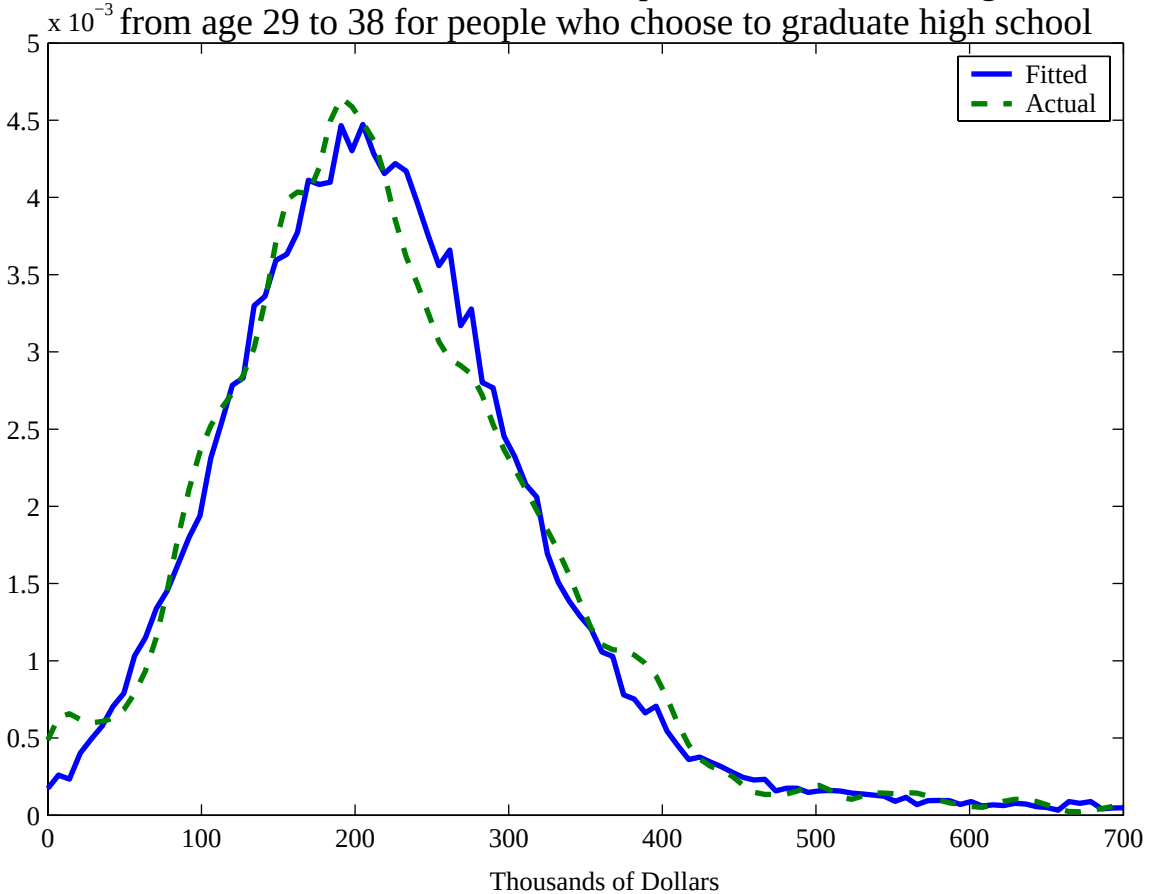
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote choice of the high school sector, and S=1 denote choice of the college sector. Define observed earnings as $Y = SY_1 + (1 - S)Y_0$. Finally, let $f(y)$ denote the density function of observed earnings. Here we plot the density functions f generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 2.1
Densities of fitted and actual present value of earnings
from age 19 to 28 for people who choose to graduate high school



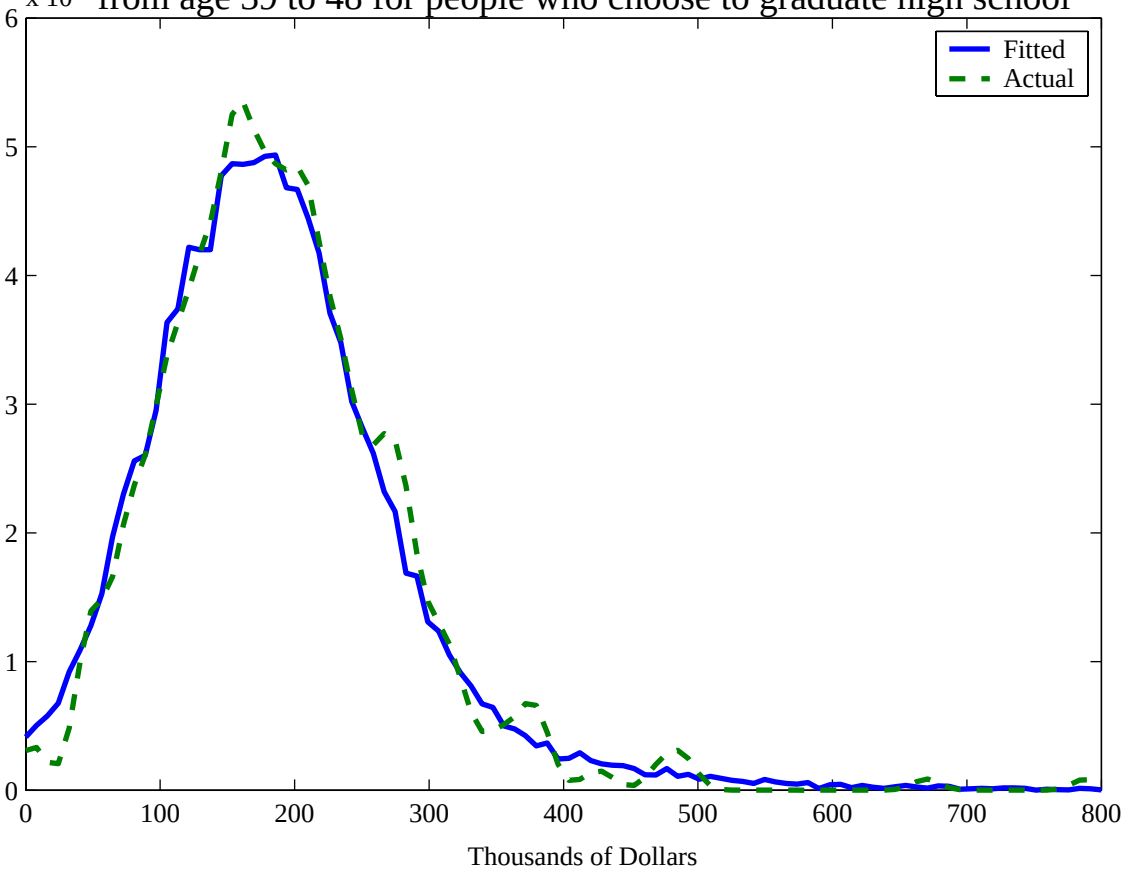
Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0| S=0)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 2.2
Densities of fitted and actual present value of earnings
from age 29 to 38 for people who choose to graduate high school



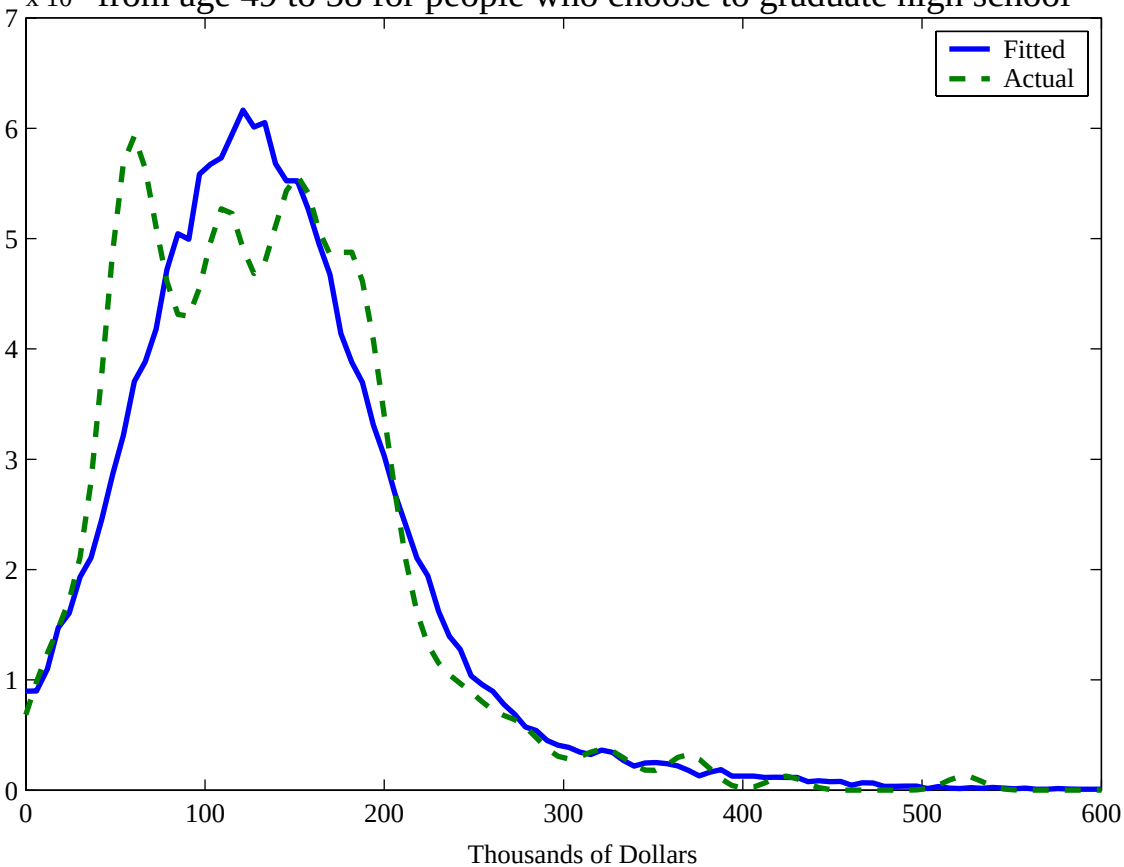
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0| S=0)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 2.3
Densities of fitted and actual present value of earnings
from age 39 to 48 for people who choose to graduate high school



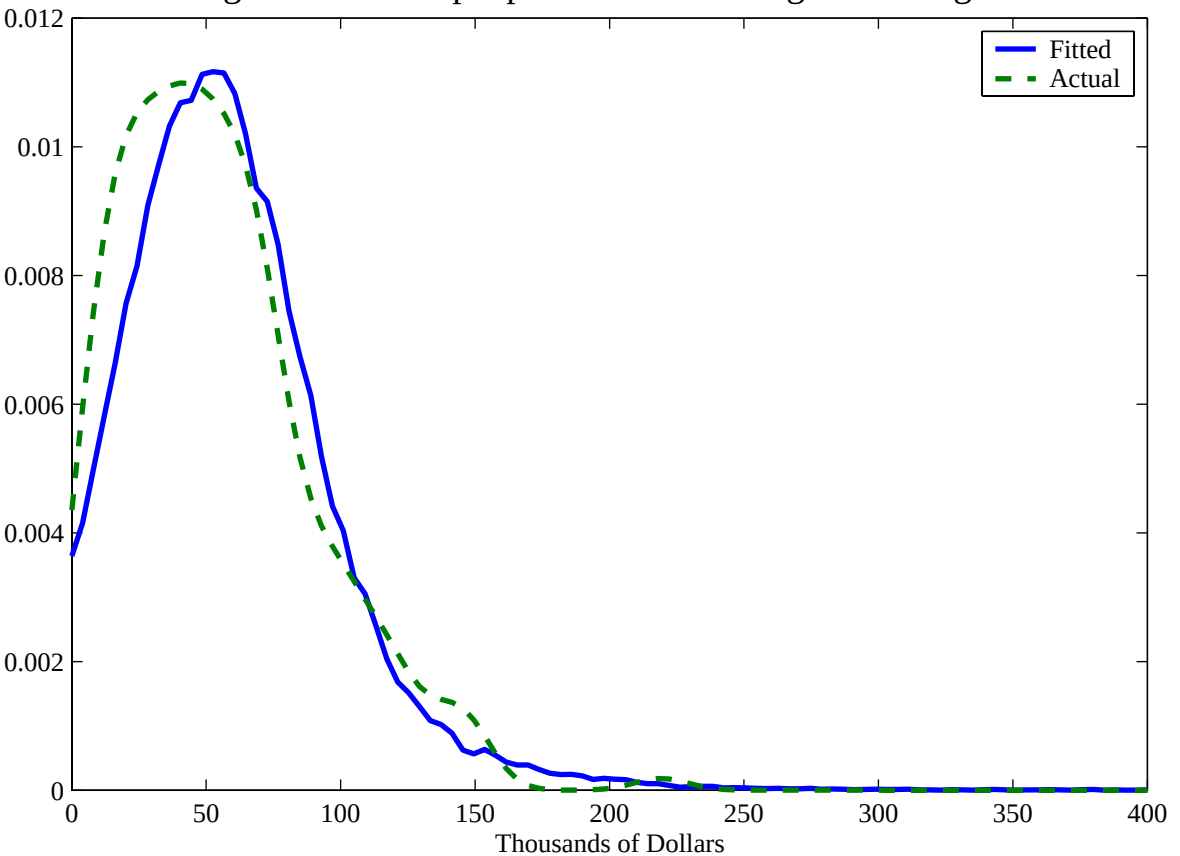
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|\ S=0)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 2.4
Densities of fitted and actual present value of earnings
from age 49 to 58 for people who choose to graduate high school



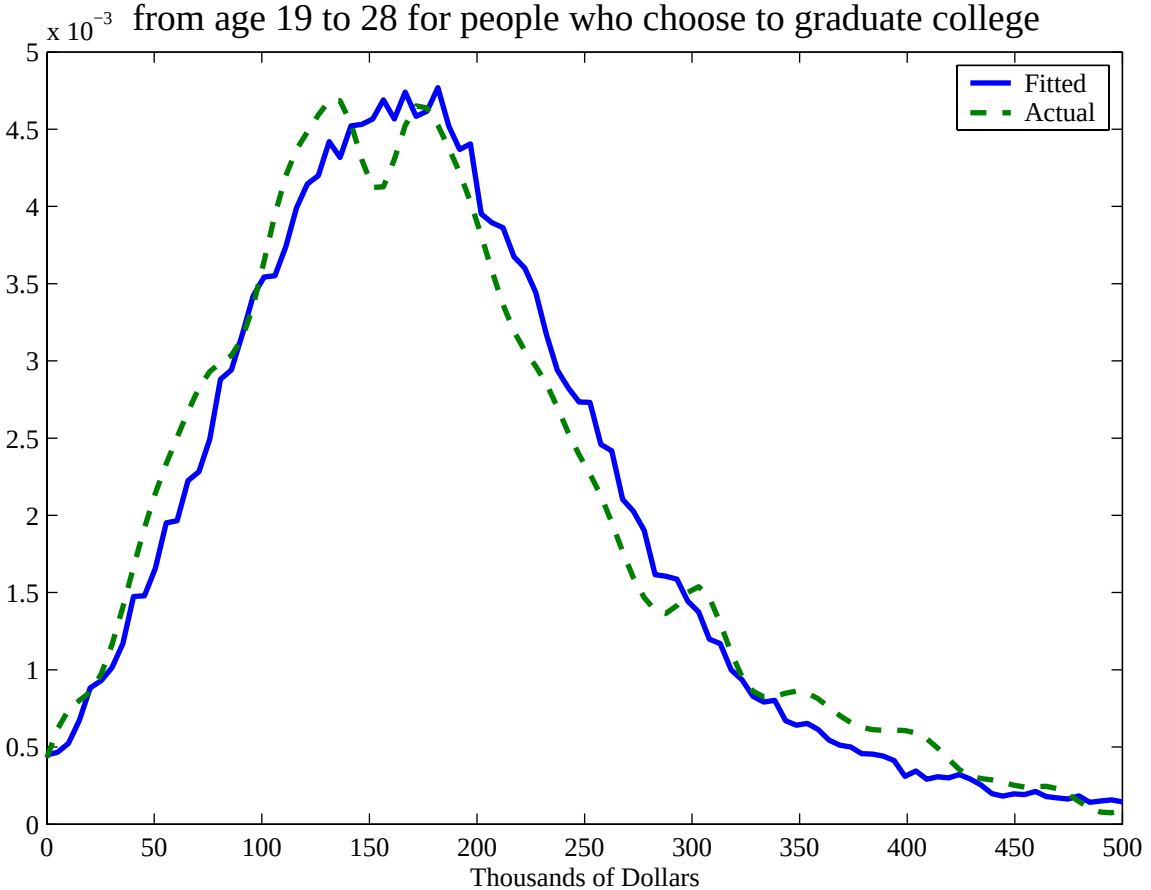
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|\ S=0)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 2.5
Densities of fitted and actual present value of earnings
from age 59 to 65 for people who choose to graduate high school



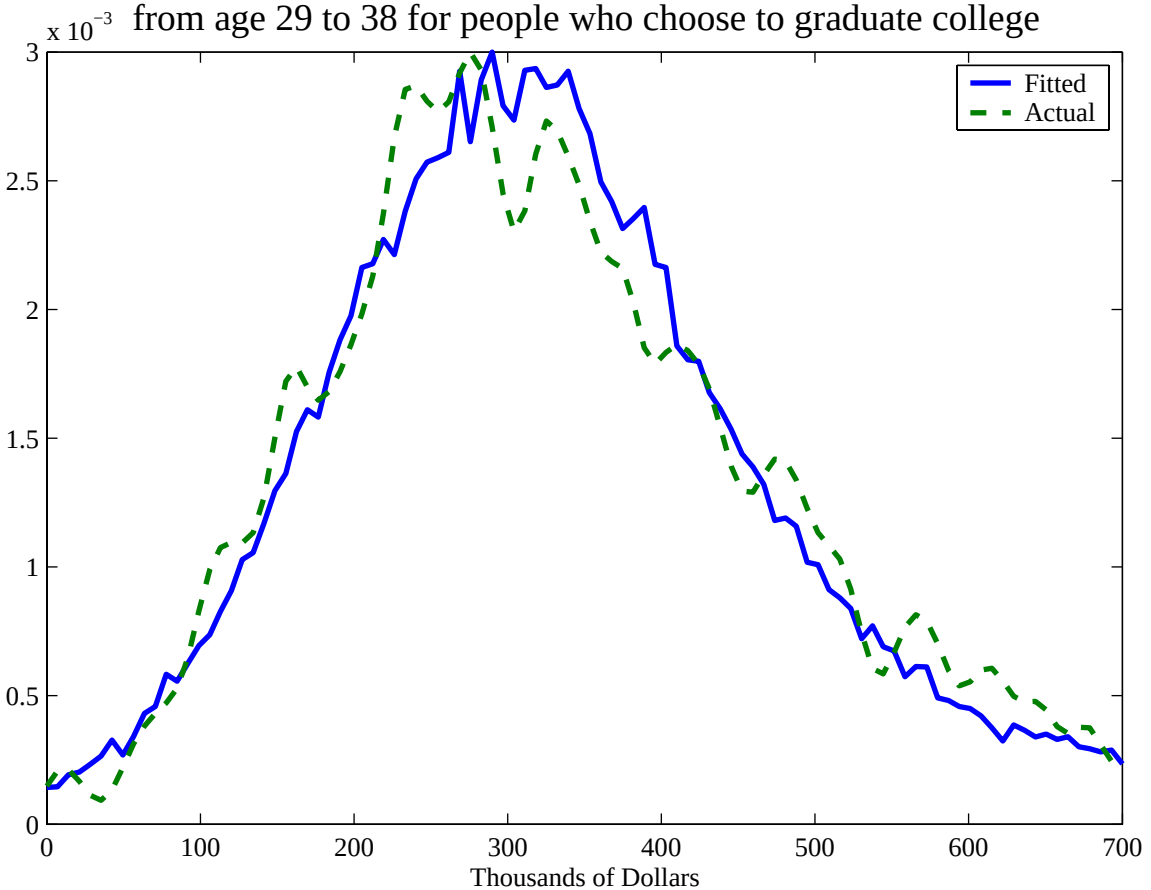
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 3.1
Densities of fitted and actual present value of earnings
from age 19 to 28 for people who choose to graduate college



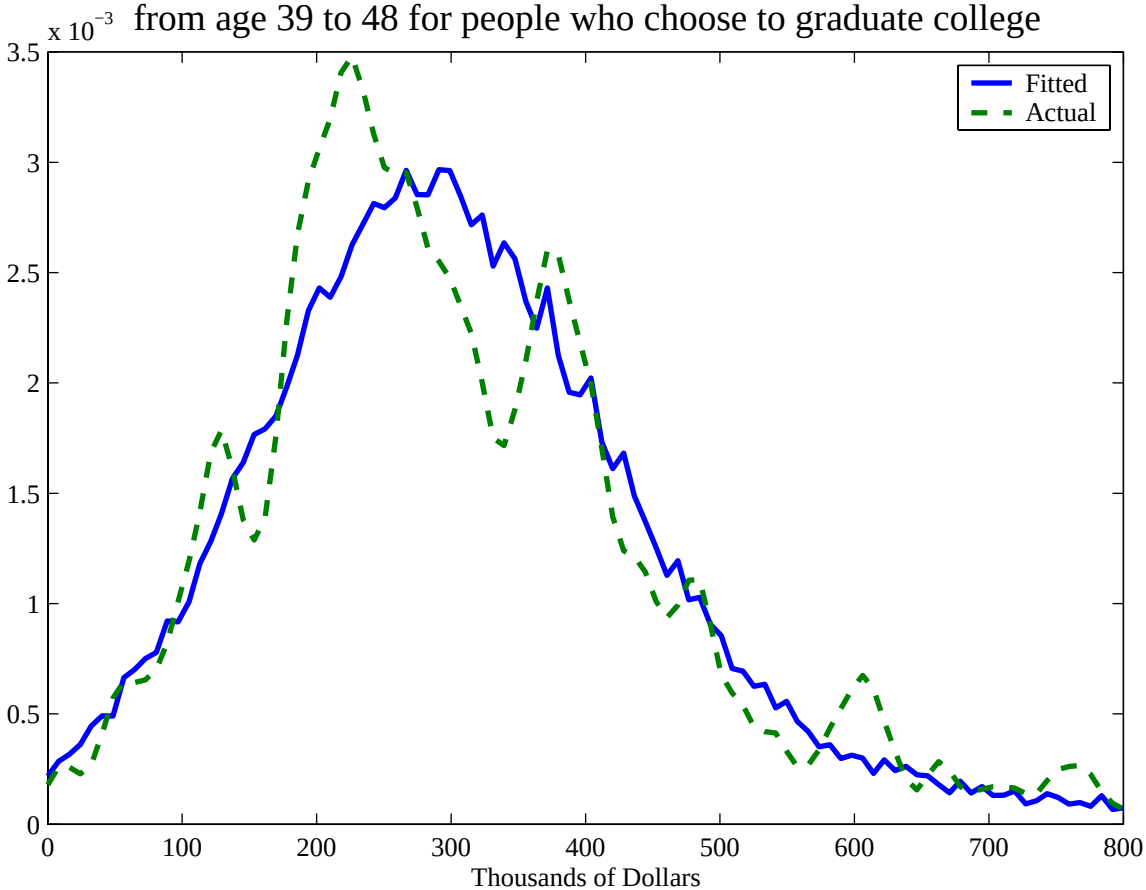
Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1| S=1)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 3.2
Densities of fitted and actual present value of earnings
from age 29 to 38 for people who choose to graduate college



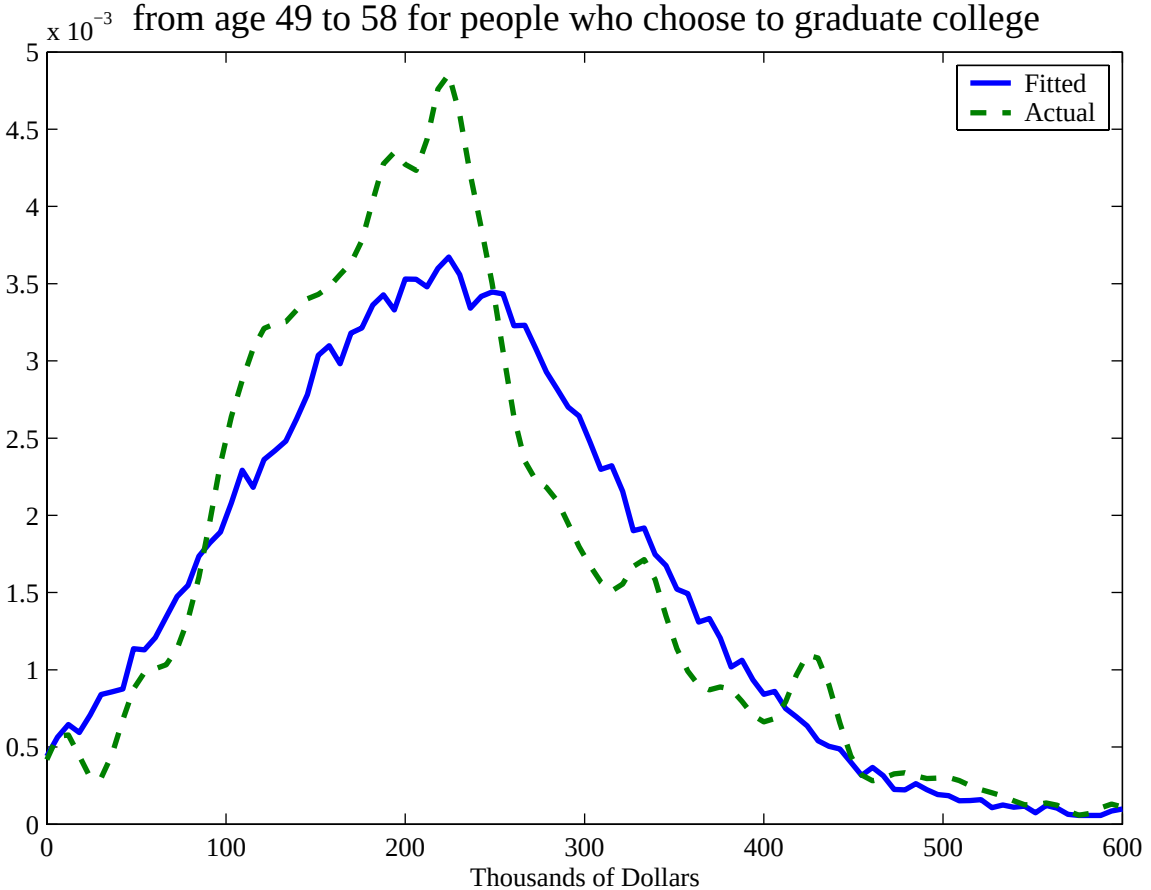
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1| S=1)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 3.3
Densities of fitted and actual present value of earnings
from age 39 to 48 for people who choose to graduate college



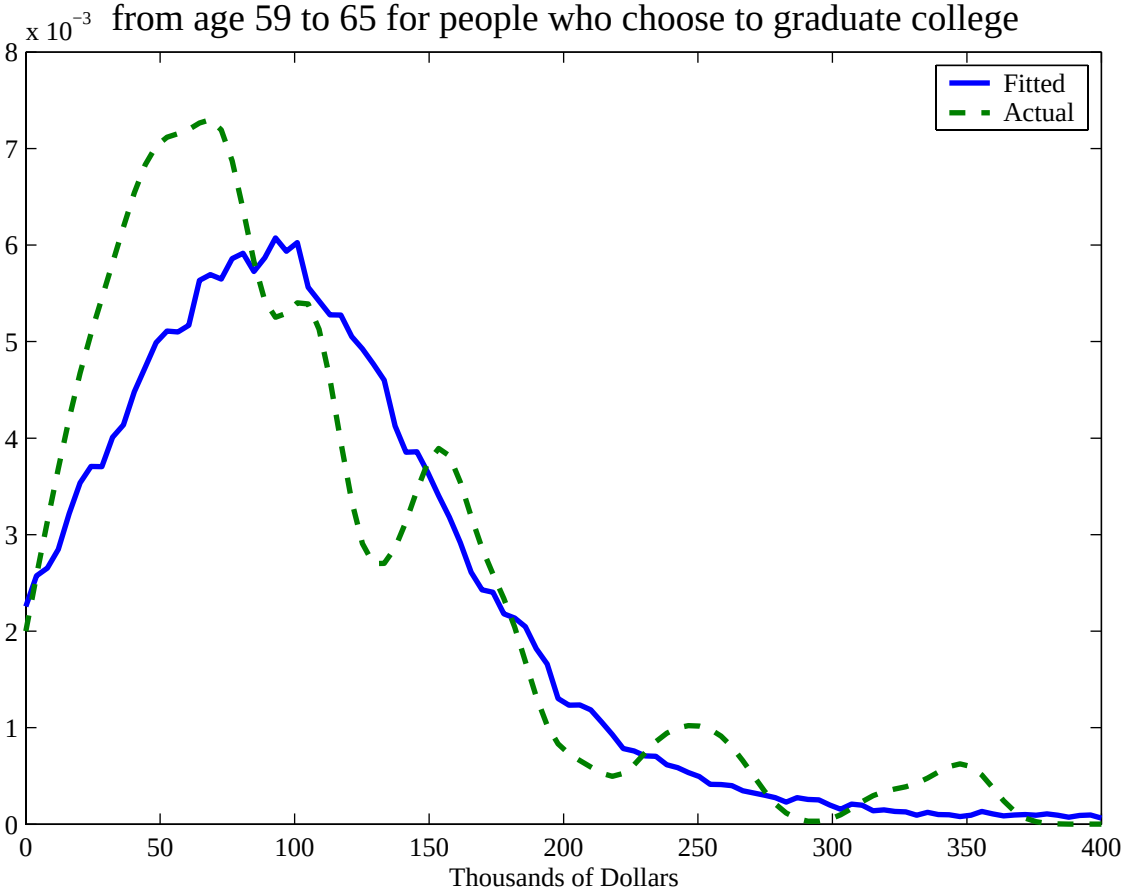
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1| S=1)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 3.4
Densities of fitted and actual present value of earnings
from age 49 to 58 for people who choose to graduate college



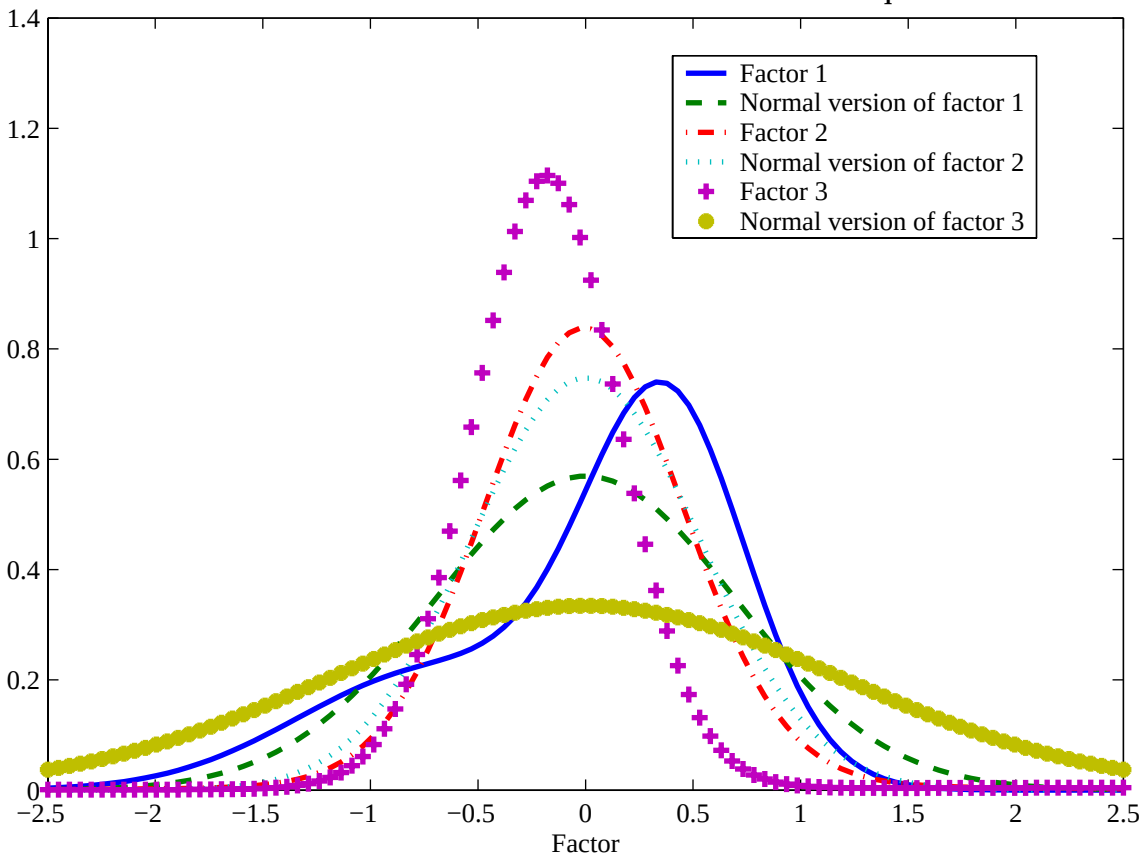
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1| S=1)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 3.5
Densities of fitted and actual present value of earnings
from age 59 to 65 for people who choose to graduate college



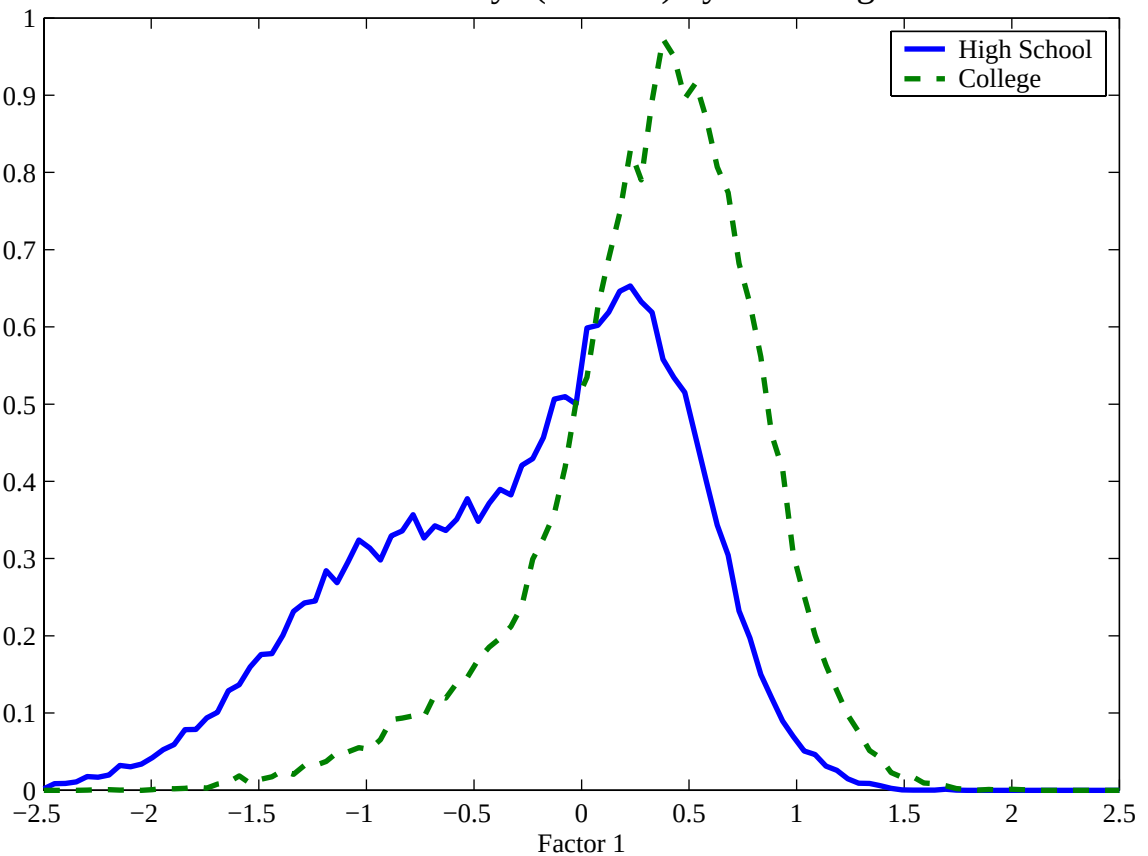
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1| S=1)$ generated from the data (the dashed line), and that predicted by the model (the solid line). We use kernel density estimation to smooth these functions.

Figure 4
Densities of estimated factors and their normal equivalents



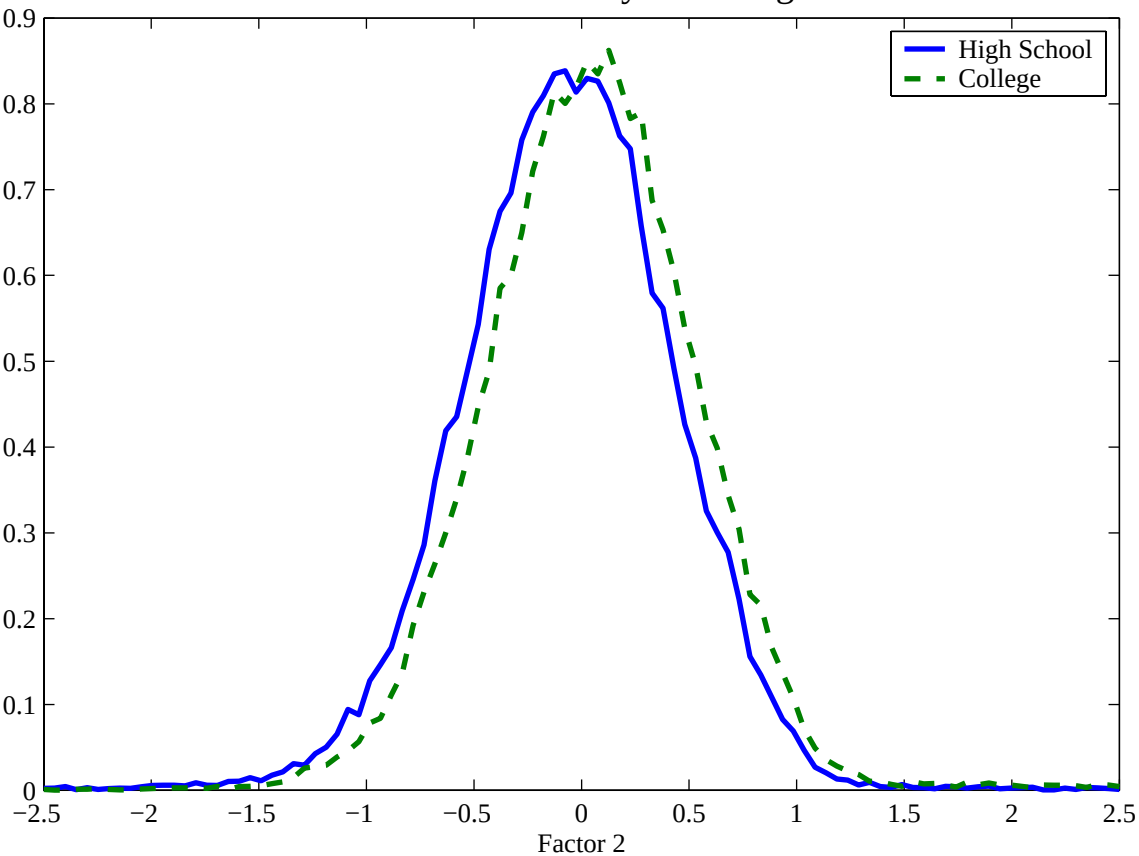
Let $f(\theta_1)$ denote the probability density function of factor θ_1. We allow $f(\theta_1)$ to be a mixture of normals. Assume $\mu_1=E(\theta_1)$ and $\sigma_1=Var(\theta_1)$. Let $\phi(\mu_1,\sigma_1)$ denote the density of a normal random variable with mean μ_1 and variance σ_1. The solid curve is the actual density of factor θ_1, $f(\theta_1)$, while the dashed curve is the density of a normal random variable with mean μ_1 and variance σ_1. We proceed similarly for factors 2 and 3 using the notation in the legend.

Figure 5.1
Densities of "ability" (factor 1) by schooling level



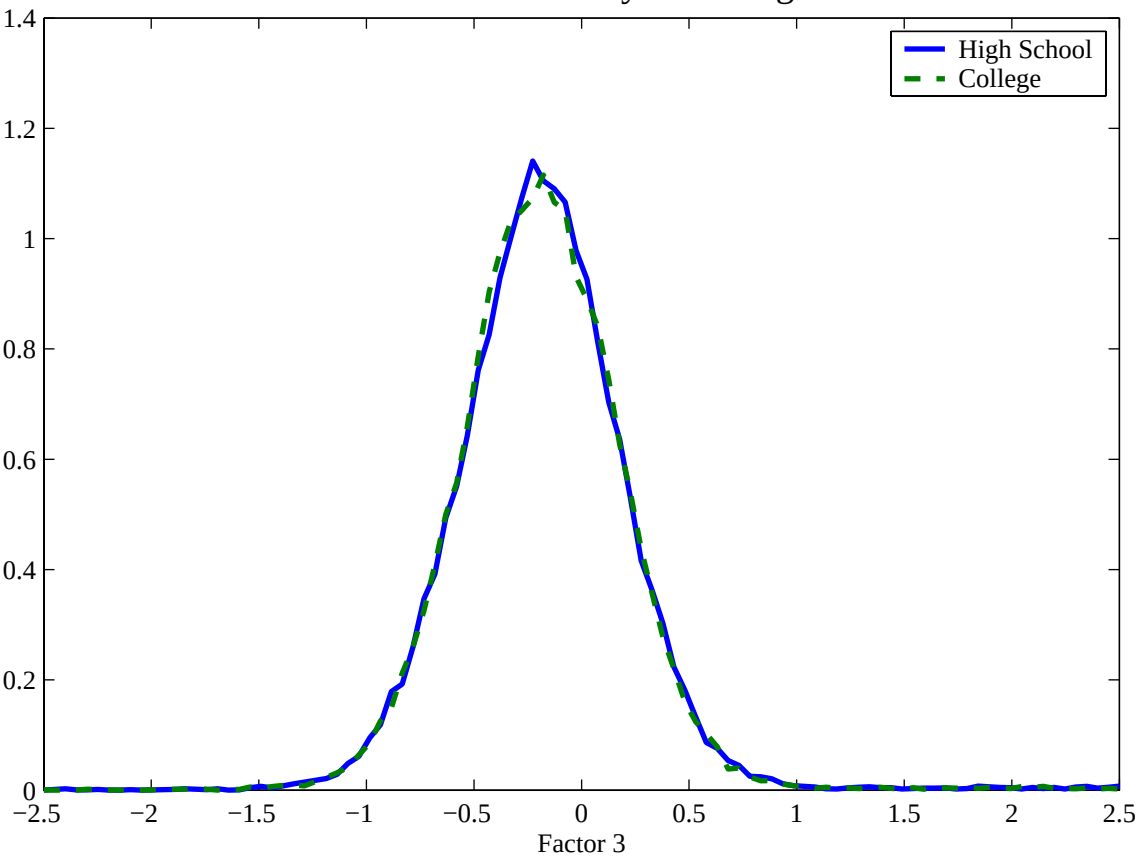
Let $f(\theta_1)$ denote the probability density function of factor θ_1. We allow $f(\theta_1)$ to be a mixture of normals. The solid line plots the density of factor 1 conditional on choosing the high school sector, that is, $f(\theta_1|\text{choice=high school})$. The dashed line plots the density of factor 1 conditional on choosing the college sector, that is, $f(\theta_1|\text{choice=college})$.

Figure 5.2
Densities of factor 2 by schooling level



Let $f(\theta_2)$ denote the probability density function of factor θ_2. We allow $f(\theta_2)$ to be a mixture of normals. The solid line plots the density of factor 2 conditional on choosing the high school sector, that is, $f(\theta_2|\text{choice=high school})$. The dashed line plots the density of factor 2 conditional on choosing the college sector, that is, $f(\theta_2|\text{choice=college})$.

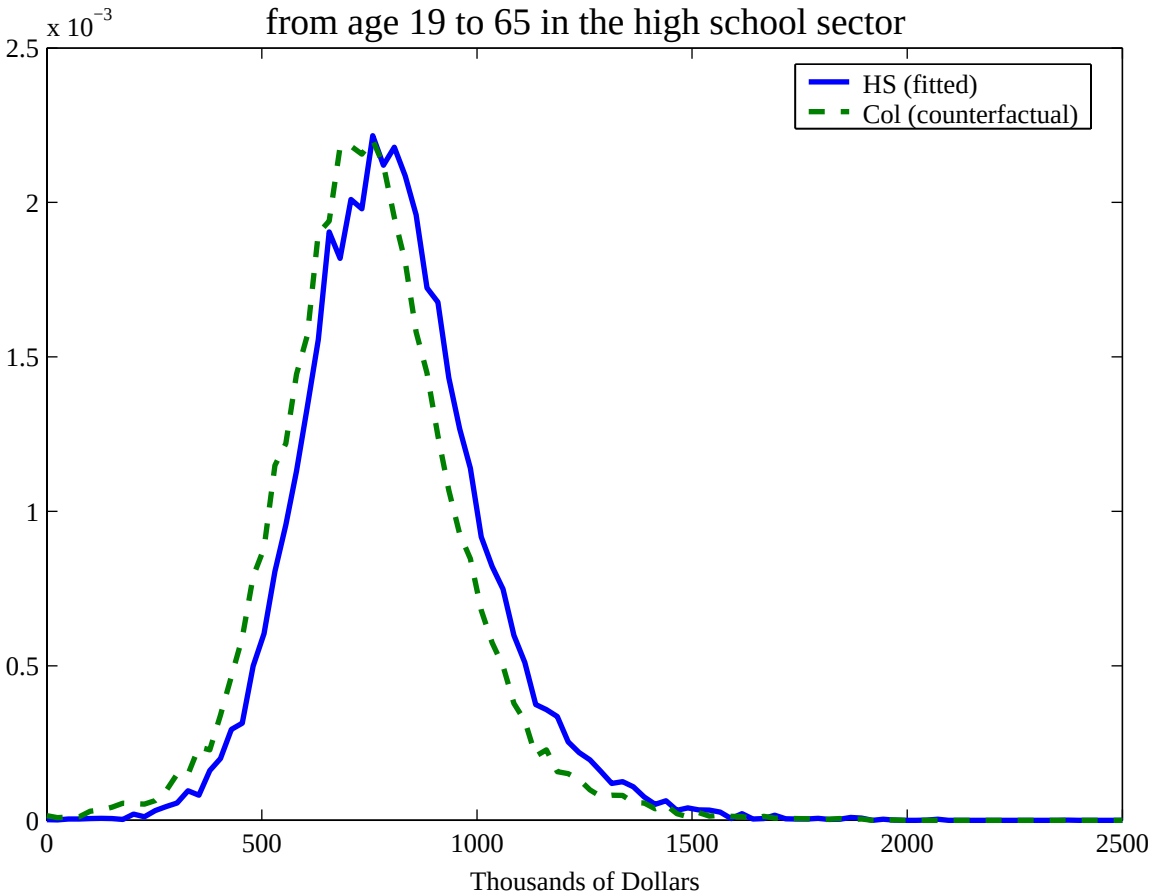
Figure 5.3
Densities of factor 3 by schooling level



Let $f(\theta_3)$ denote the probability density function of factor θ_3. We allow $f(\theta_3)$ to be a mixture of normals. The solid line plots the density of factor 3 conditional on choosing the high school sector, that is, $f(\theta_3|\text{choice=high school})$. The dashed line plots the density of factor 3 conditional on choosing the college sector, that is, $f(\theta_3|\text{choice=college})$.

Figure 6.1

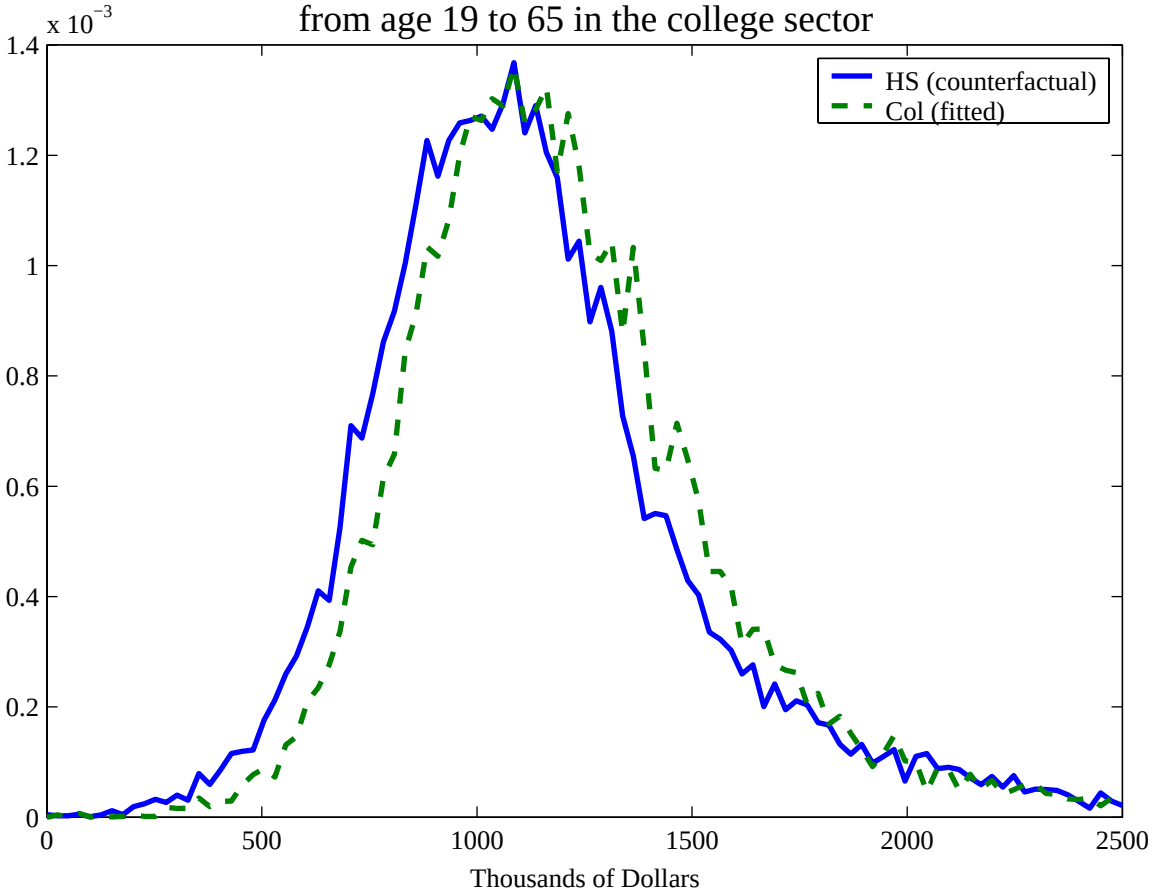
Densities of *ex post* present value of counterfactual and fitted earnings from age 19 to 65 in the high school sector



Let Y_0 denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(Y_0)$ denote its density function. The solid line plots the predicted Y_0 density conditional on choosing high school, that is, $f(Y_0| S=0)$, while the dashed line shows the counterfactual density function of Y_0 for those agents who are actually college graduates, that is, $f(Y_0| S=1)$. This assumes that the agent chooses schooling without knowing θ_3 and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$

Figure 6.2

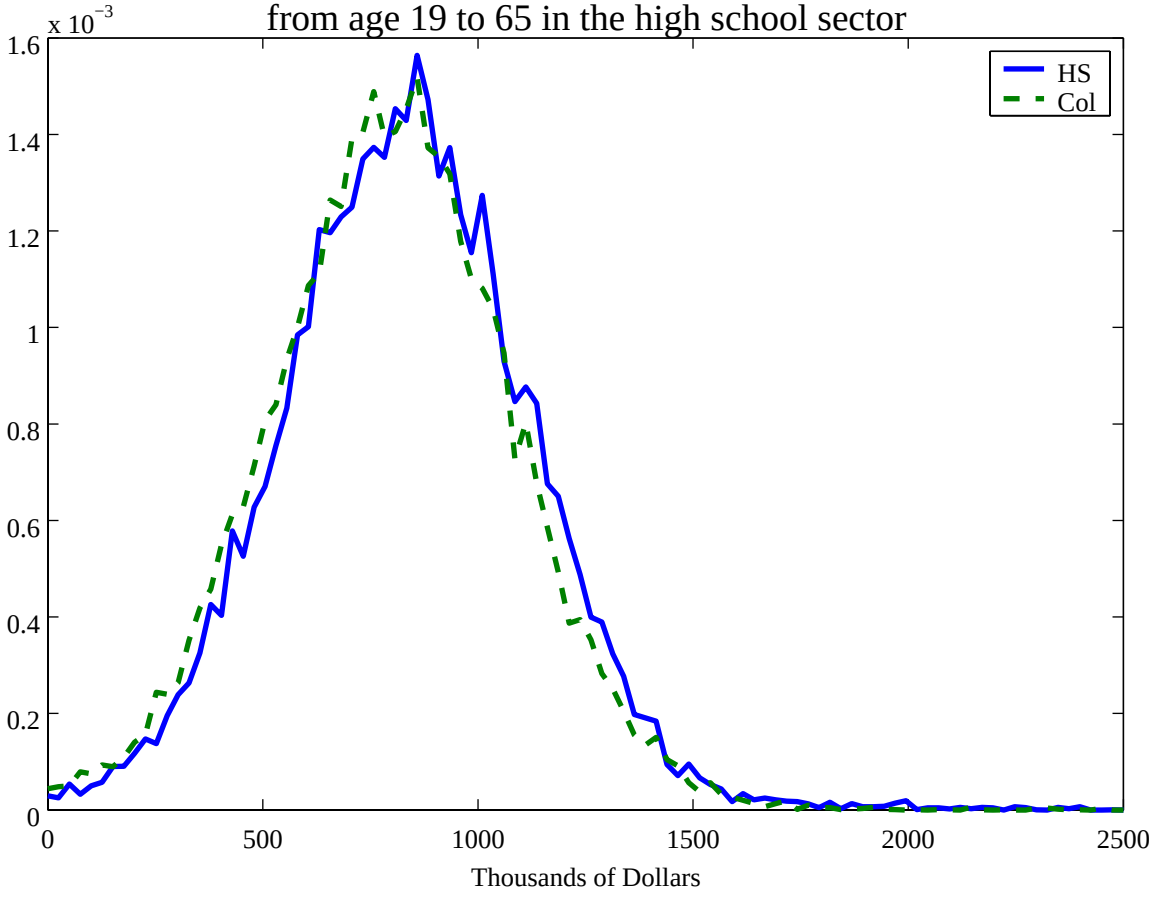
Densities of *ex post* present value of counterfactual and fitted earnings from age 19 to 65 in the college sector



Let Y_1 denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The dashed line plots the predicted Y_1 density conditional 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 who are actually high school graduates, that is, $f(Y_1|S=0)$. This assumes that the agent chooses schooling without knowing θ_3 and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$

Figure 6.3

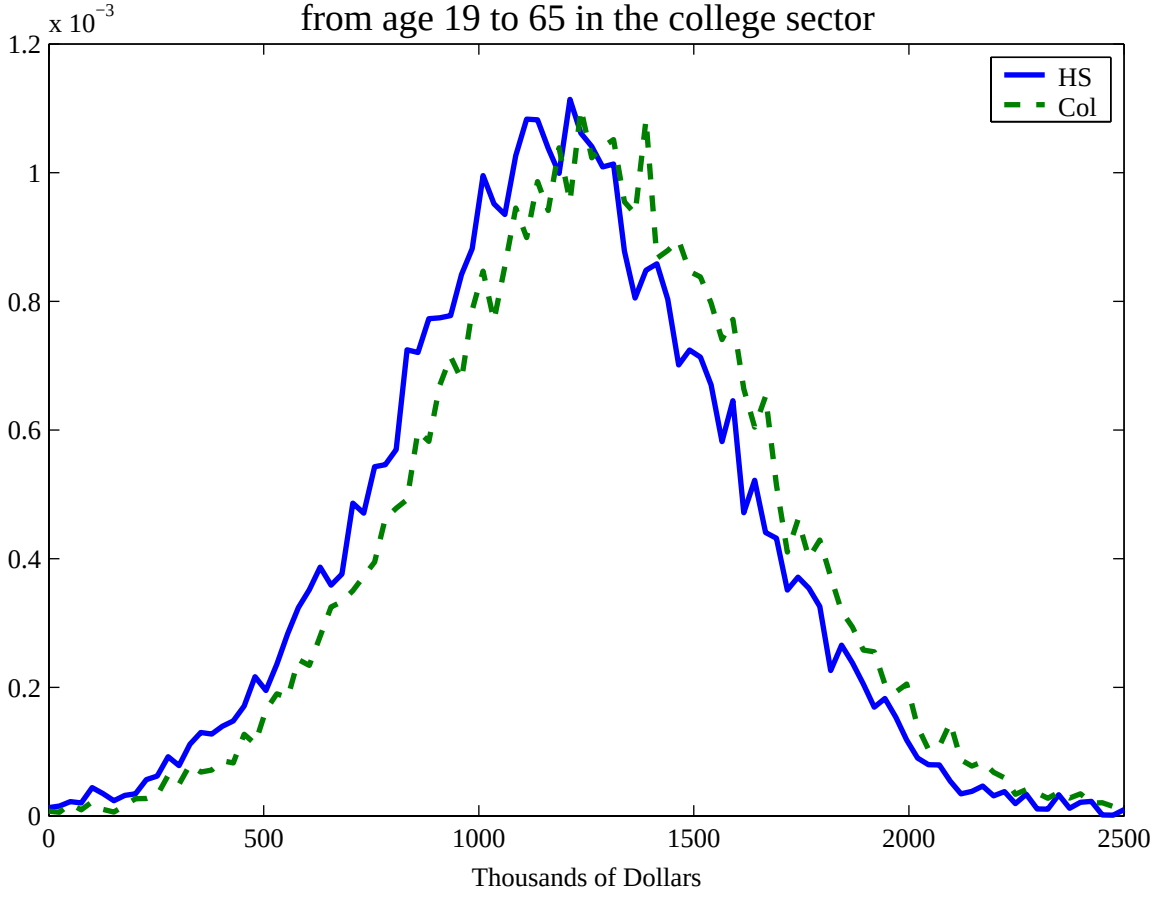
Densities of *ex ante* present value of counterfactual and fitted earnings from age 19 to 65 in the high school sector



Let $\varepsilon=(\varepsilon_{0,t'}\ \varepsilon_{1,t'}\ t=0,...T)$. Let $E_{\theta_3,\varepsilon}(Y_0)$ denote the ex ante present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(E_{\theta_3,\varepsilon}(Y_0)$ denote its density function. The solid curve plots the predicted Y_0 density conditional on choosing high school, that is, $f(E_{\theta_3,\varepsilon}(Y_0)|S=0)$, while the dashed line shows the counterfactual density function of $E_{\theta_3,\varepsilon}(Y_0)$ for those agents who are actually college graduates, that is, $f(E_{\theta_3,\varepsilon}(Y_0)|S=1)$. This is constructed assuming that the agent chooses schooling without knowing θ_3 and ε.

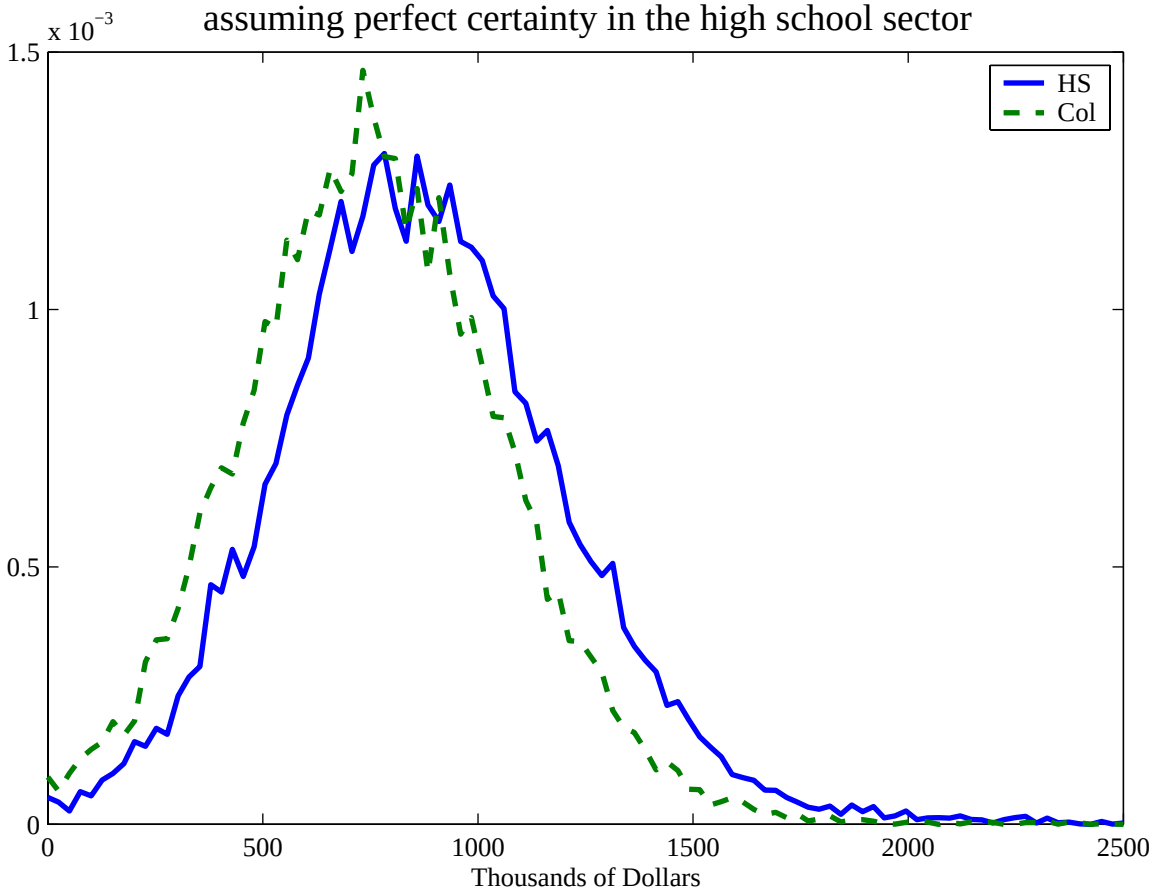
Figure 6.4

Densities of *ex ante* present value of counterfactual and fitted earnings from age 19 to 65 in the college sector



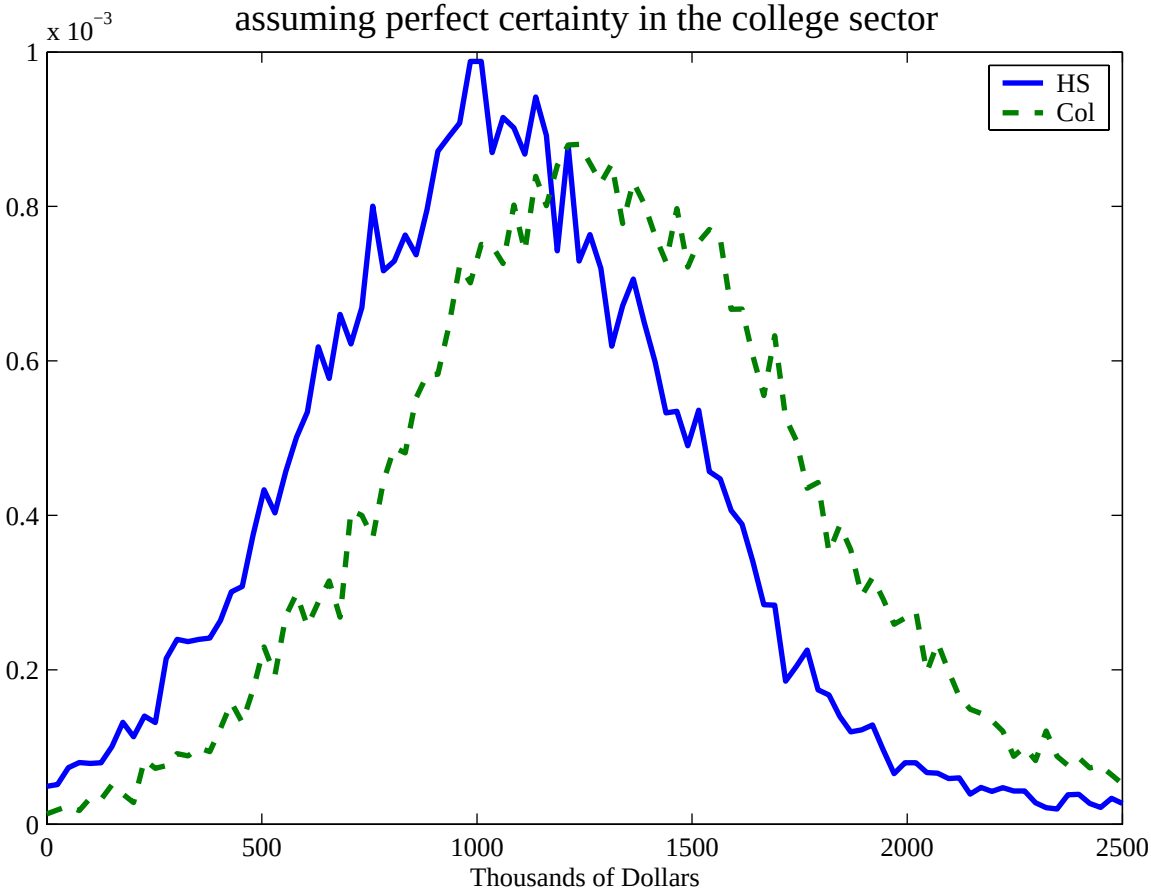
Let $\varepsilon=\{\varepsilon_{0,t},\varepsilon_{1,t},t=0,\ldots,T\}$. Let $E_{\theta_3,\varepsilon}(Y_1)$ denote the *ex ante* present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(E_{\theta_3,\varepsilon}(Y_1))$ denote its density function. The solid line plots the counterfactual Y_1 density conditional on choosing high school, that is, $f(E_{\theta_3,\varepsilon}(Y_1)|\, S=0)$, while the dashed line shows the predicted density function of $E_{\theta_3,\varepsilon}(Y_1)$ for those agents who are actually college graduates, that is, $f(E_{\theta_3,\varepsilon}(Y_1)|\, S=0)$. This is constructed assuming that the agent chooses schooling without knowing θ_3 and ε.

Figure 6.5
Densities of present value of counterfactual and fitted earnings from age 19 to 65 assuming perfect certainty in the high school sector



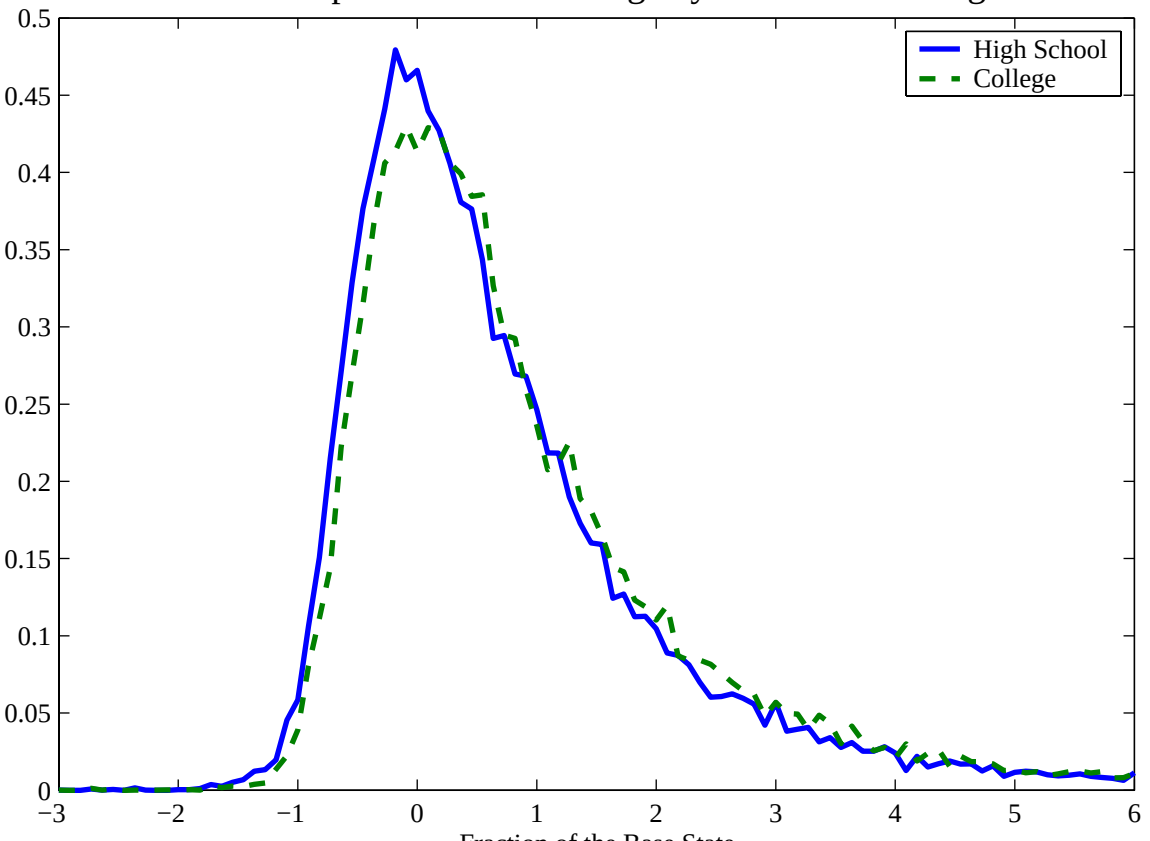
Let Y_0 denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(Y_0)$ denote its density function. The solid curve plots the predicted Y_0 density conditional on choosing high school, that is, $f(Y_0|S=0)$, while the dashed line shows the counterfactual density function of Y_0 for those agents who are actually college graduates, that is, $f(Y_0|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.

Figure 6.6
Densities of present value of counterfactual and fitted earnings from age 19 to 65 assuming perfect certainty in the college sector



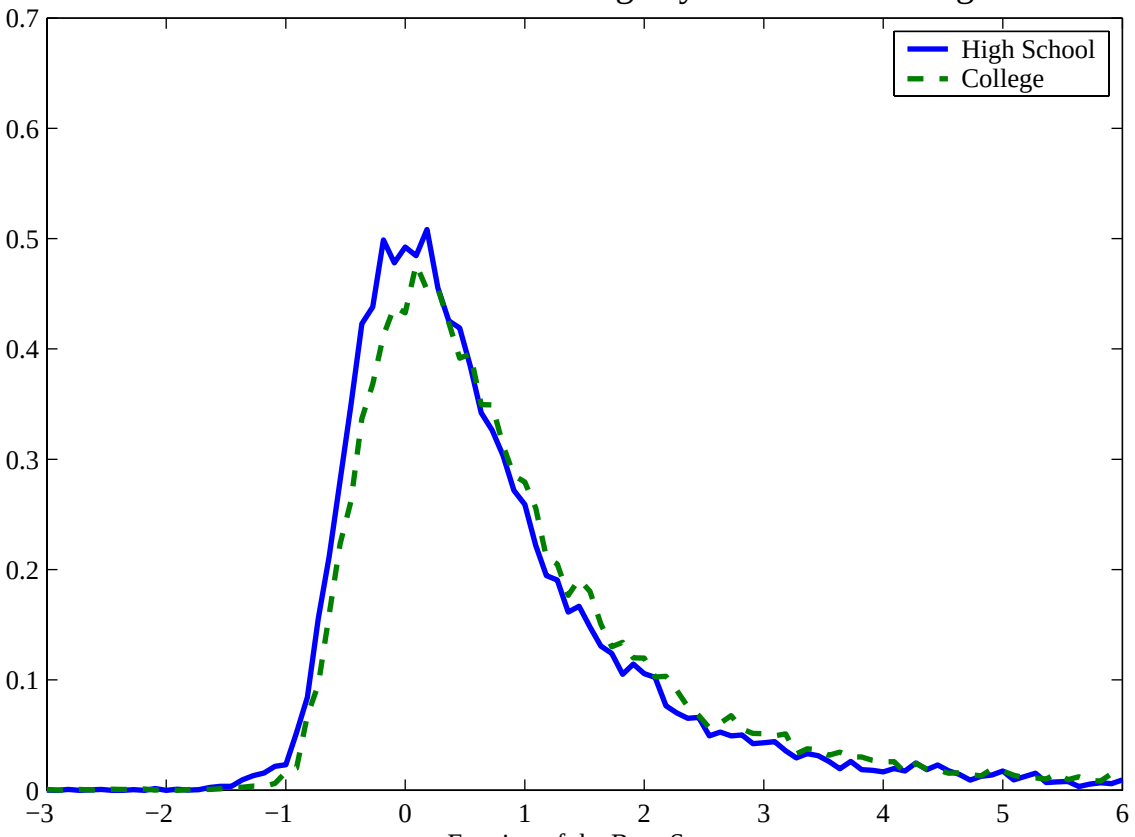
Let Y_1 denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The solid curve plots the counterfactual Y_1 density conditional on choosing high school, that is, $f(Y_1|S=0)$, while the dashed line shows the predicted density function of Y_1 for those agents who are actually college graduates, that is, $f(Y_1|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.

Figure 7.1
Densities of ex post returns to college by level of schooling chosen



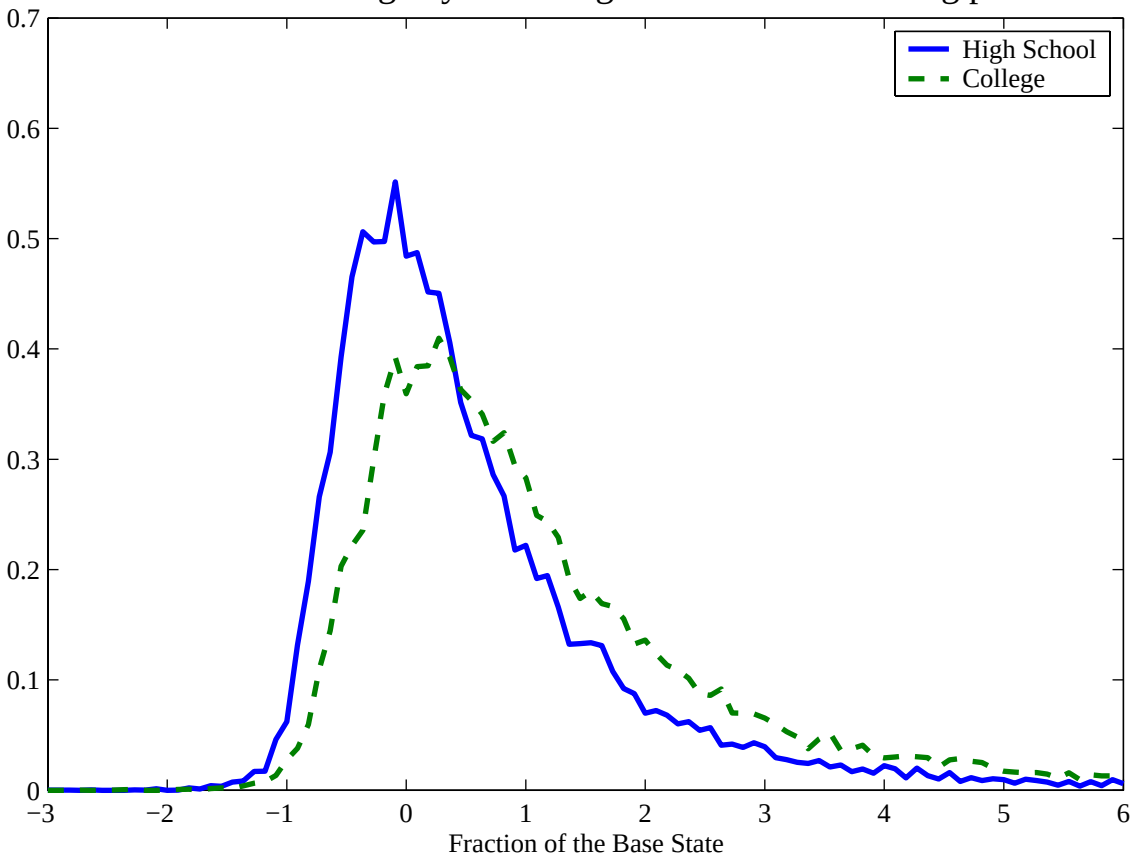
Let Y_0, Y_1 denote the present value of earnings in the high school and college sectors, respectively. Define ex post returns to college as the ratio $R=(Y_1-Y_0)/Y_0$. Let f(r) denote the density function of the random variable R. The solid line is the density of ex post returns to college for high 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). This assumes that the agent chooses schooling without knowing θ_3 and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$

Figure 7.2
Densities of ex ante returns to college by level of schooling chosen



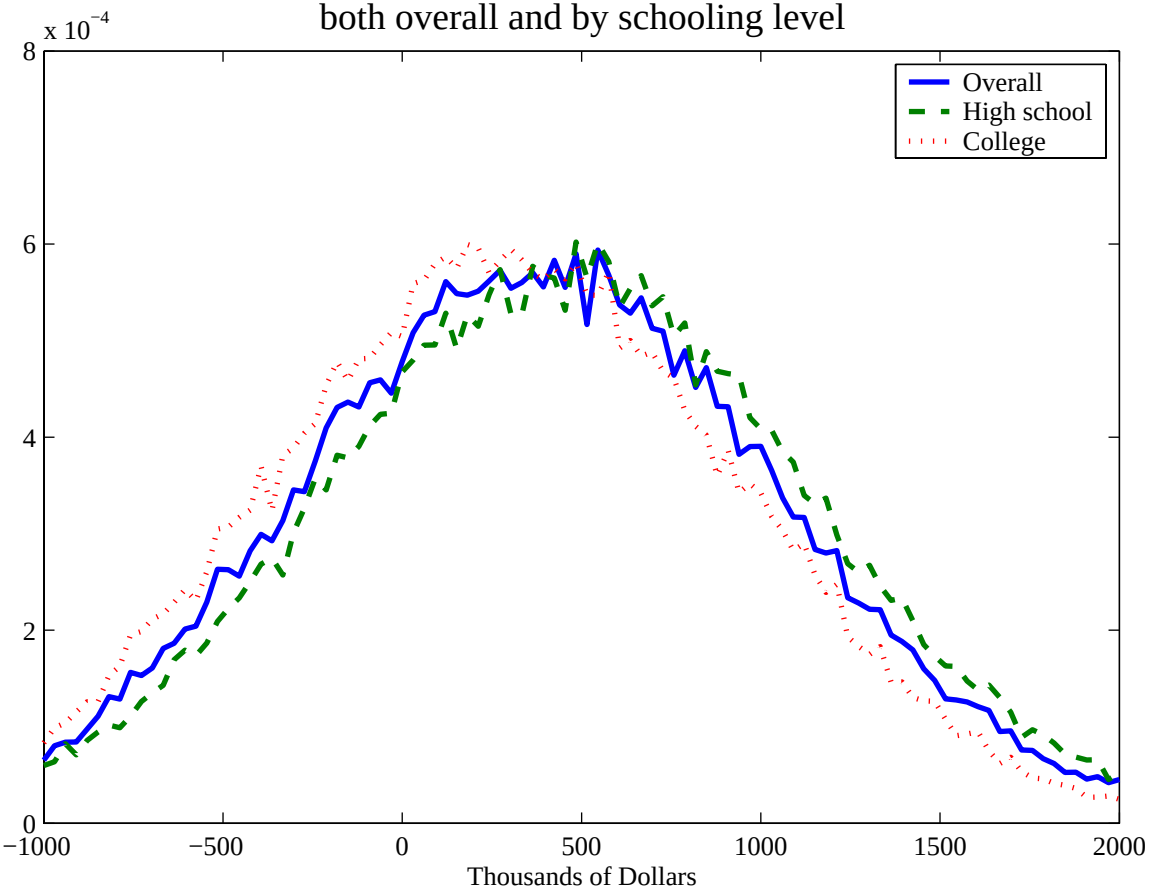
Let $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$. Let Y_0, Y_1 denote the present value of earnings in the high school and college sectors, respectively. Define ex ante returns to college as the ratio $E_{\theta_3,\varepsilon}(R)=E_{\theta_3,\varepsilon}((Y_1-Y_0)/Y_0)$. Let f(r) denote the density function of the random variable $E_{\theta_3,\varepsilon}(R)$. 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). This assumes that the agent chooses schooling without knowing θ_3 and ε.

Figure 7.3
Densities of returns to college by schooling level chosen assuming perfect certainty



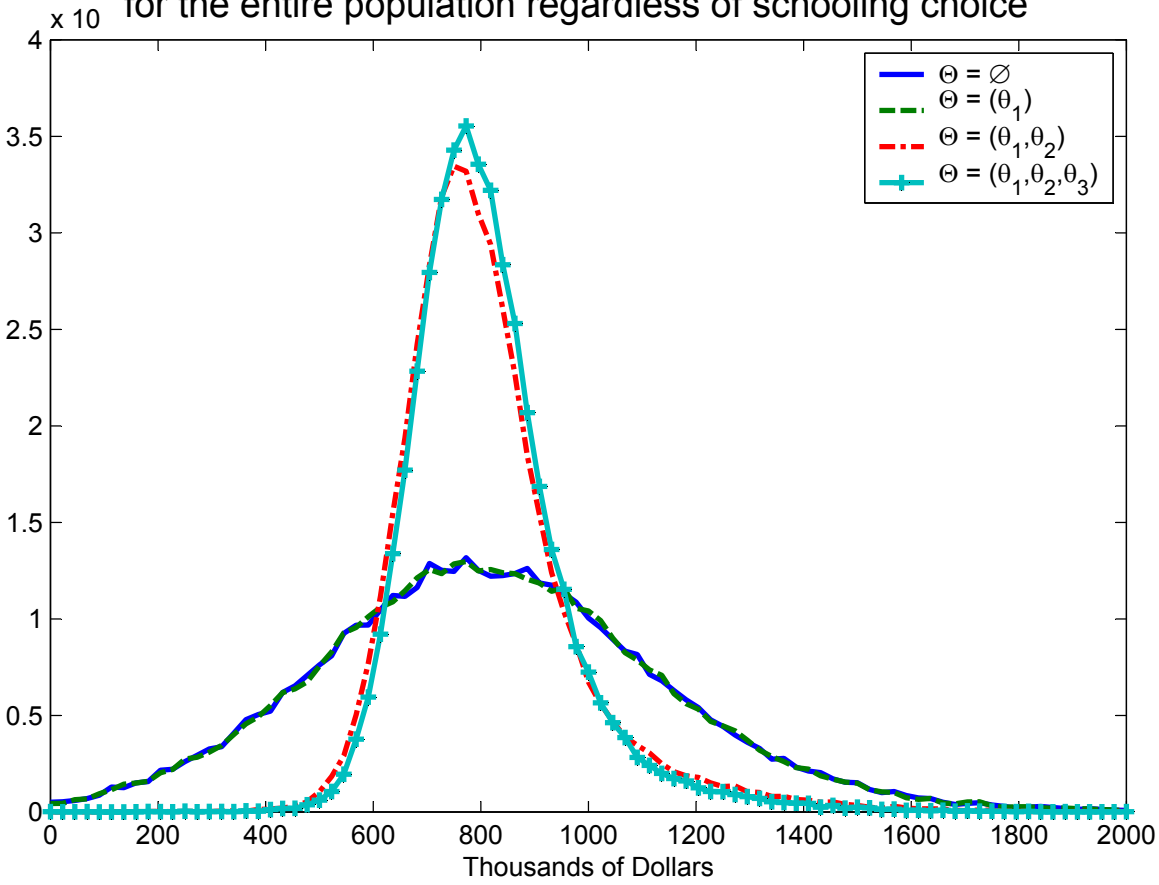
Let Y_0,Y_1 denote the present value of earnings in the high school and college sectors, respectively (discounted at a 3% interest rate). Define returns to college as the ratio $R=(Y_1-Y_0)/Y_0$. Let f(r) denote the density function of the random variable R. The solid line is the density of returns to college for high school graduates, that is f(r|S=0). The dashed line is the density of returns to college for college graduates, that is, f(r|S=1). This assumes that the agent chooses schooling with complete knowledge of future earnings.

Figure 8
Densities of monetary value of psychic cost
both overall and by schooling level



Let C denote the monetary value of psychic costs. Let f(c) denote the density function of psychic costs in monetary terms. The dashed line shows the density of psychic costs for high school graduates, that is f(c|S=0). The dotted line shows the density of psychic costs for college graduates, that is, f(c|S=1). The solid line is the unconditional density of the monetary value of psychic costs, f(c).

Figure 9.1
Densities of present value of high school earnings
under different information sets for the agent calculated
for the entire population regardless of schooling choice



Let Θ denote the agent's information set. Let Y_0 denote the present value of earnings in the high school sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of the present value of earnings in high school conditioned on the information set Θ. Then:

The solid line plots $f(y_0|\Theta)$ under no information, i.e. $\Theta=\varnothing$.

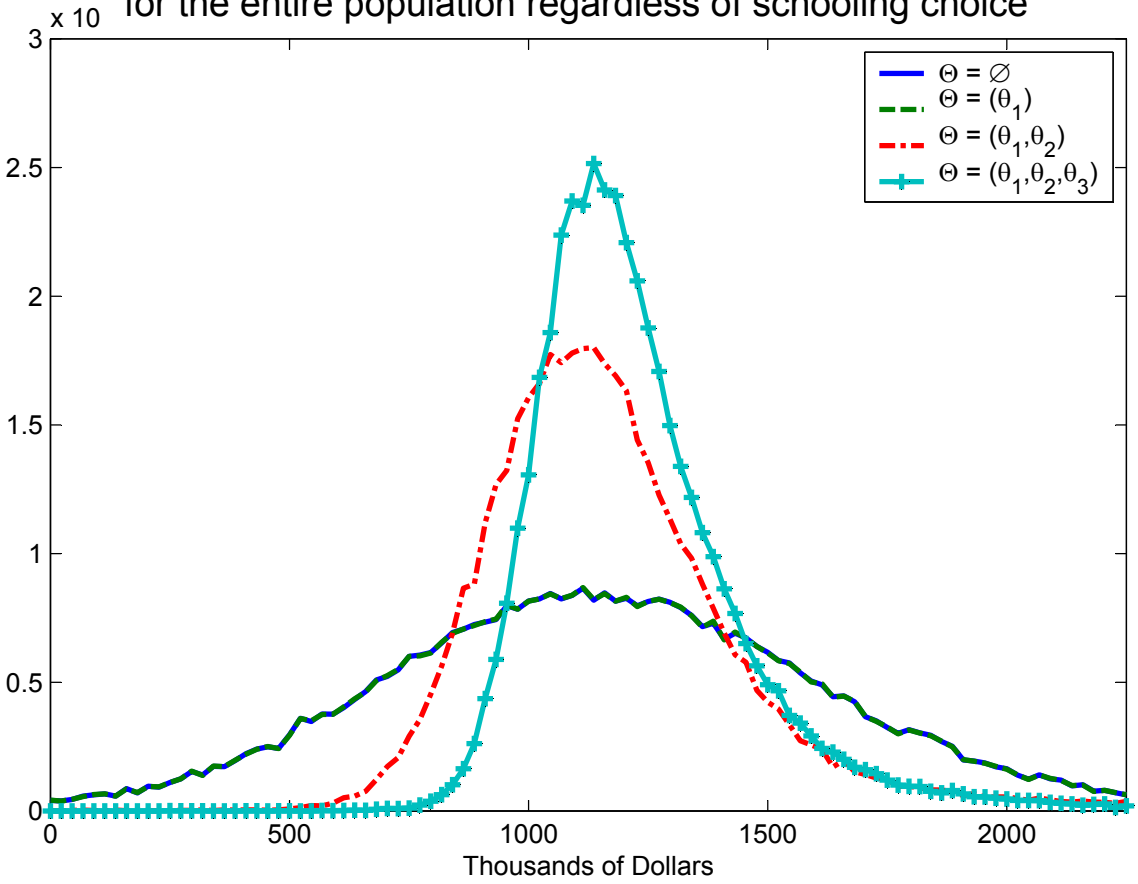
The dashed line plots $f(y_0|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.

The dashed-dotted line plots $f(y_0|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.

The crossed line plots $f(y_0|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.

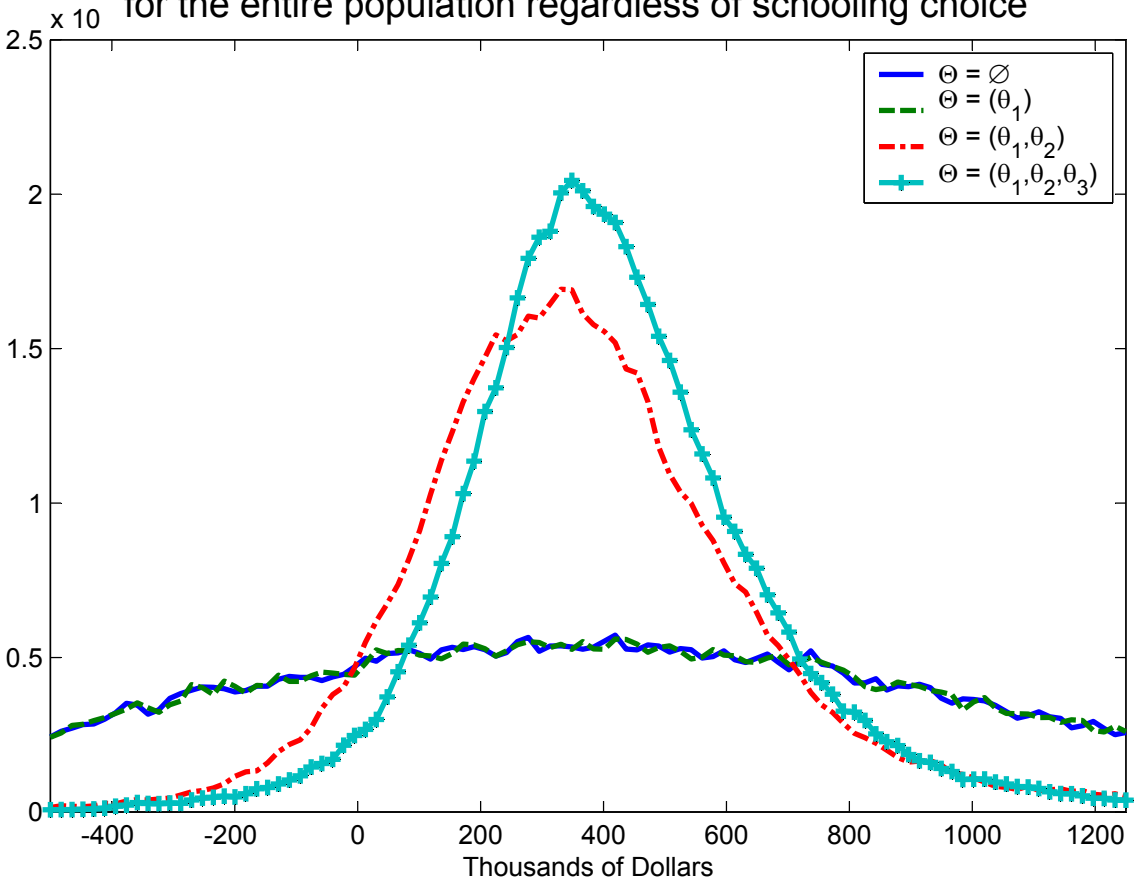
The X are put at the mean and are assumed to be known. The θ, when known, are set at their mean of zero.

Figure 9.2
Densities of present value of college earnings
under different information sets for the agent calculated
for the entire population regardless of schooling choice



Let Θ denote the agent's information set. Let Y_1 denote the present value of earnings in the college sector (discounted at a 3% interest rate). Let $f(y_1|\Theta)$ denote the density of the present value of earnings in high school conditioned on the information set Θ. Then:
The solid line plots $f(y_1|\Theta)$ under no information, i.e. $\Theta=\varnothing$.
The dashed line plots $f(y_1|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.
The dashed-dotted line plots $f(y_1|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.
The crossed line plots $f(y_1|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.
The X are put at the mean and are assumed to be known. The θ, when known, are set at their mean of zero.

Figure 9.3
Densities of returns college vs high school
under different information sets for the agent calculated
for the entire population regardless of schooling choice



Let Θ denote the agent's information set. Let Y_0, Y_1 denote the present value of earnings in the high school and college sectors, respectively (discounted at a 3% interest rate). Let $D=Y_0-Y_1$ be the difference of the present value of earnings in the college and high school sector. $f(d|\Theta)$ denote the density of the difference of present value of earnings conditioned on the information set Θ. Then:

The solid line plots $f(d|\Theta)$ under no information, i.e. $\Theta=\varnothing$.

The dashed line plots $f(d|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.

The dashed-dotted line plots $f(d|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.

The crossed line plots $f(d|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.

The X are put at the mean and are assumed to be known The θ when known are set at their mean of zero

Table 1

Estimated Effects of *Ex Post* Returns to Schooling on Schooling Choice using OLS and IV To Estimate The *Ex Post* Returns To Schooling

Log Earnings Regression*		
	OLS	
Variable	**Coefficient**	**Std. Error**
School (High School vs. College)	0.2735	0.0344
School*ASVAB	0.0279	0.0063
	Instrumental Variables†	
School	0.2573	0.0451
School*ASVAB	0.0153	0.0083

Schooling Choice Probit Equation‡		
	Using OLS Results	
Variable	**Coefficient**	**Std. Error**
$b_{School} + b_{School*ASVAB}$*ASVAB	12.6244	0.7284
Marginal Effect	4.8333	0.2654
	Using IV Coefficients	
$b_{School} + b_{School*ASVAB}$*ASVAB	22.9150	1.3221
Marginal Effect	8.7731	0.4817

*Includes controls for Mincer experience (age - years of schooling - 6), experience squared, cohort dummies, and ASVAB scores.

†We use parental education, family income, broken home, number of siblings, distance to college, local tuition, cohort dummies, South at age 14 and urban at age 14 to instrument for schooling and schooling interacted with ASVAB scores.

‡We use the predicted return to school to test whether future earnings affect current schooling choices. We include controls for family background, cohort dummies, distance to college, and local tuition.

Table 2.1

Descriptive Statistics from the Pooled NLSY/1979 and PSID (white males)

	Full Sample					High School Sample					College Sample				
Variable Name	Obs	Mean	Std. Dev	Min	Max	Obs	Mean	Std. Dev	Min	Max	Obs	Mean	Std. Dev	Min	Max
Asvab AR*	1362	0.72	0.95	-1.78	1.96	747	0.26	0.89	-1.78	1.96	615	1.27	0.70	-1.36	1.96
Asvab PC*	1362	0.42	0.80	-2.68	1.36	747	0.07	0.86	-2.68	1.36	615	0.84	0.44	-1.06	1.36
Asvab WK*	1362	0.52	0.72	-2.29	1.34	747	0.20	0.76	-2.29	1.34	615	0.92	0.41	-1.36	1.34
Asvab MK*	1362	0.62	1.03	-1.62	2.11	747	0.00	0.81	-1.62	2.11	615	1.38	0.73	-1.46	2.11
Asvab CS*	1362	0.21	0.85	-2.52	2.49	747	-0.08	0.79	-2.52	2.08	615	0.56	0.77	-2.52	2.49
Urban at age 14	3695	0.79	0.40	0.00	1.00	1953	0.75	0.44	0.00	1.00	1742	0.85	0.36	0.00	1.00
Parents Divorced	3695	0.15	0.36	0.00	1.00	1953	0.18	0.38	0.00	1.00	1742	0.13	0.34	0.00	1.00
Number of Siblings	3695	2.86	1.96	0.00	17.00	1953	3.19	2.08	0.00	14.00	1742	2.49	1.74	0.00	17.00
Father's Education	3695	4.31	1.94	1.00	8.00	1953	3.56	1.51	1.00	8.00	1742	5.15	2.03	1.00	8.00
Mother's Education	3695	4.21	1.55	1.00	8.00	1953	3.68	1.26	1.00	8.00	1742	4.79	1.63	1.00	8.00
Born between 1906 and 1915	3695	0.01	0.10	0.00	1.00	1953	0.01	0.12	0.00	1.00	1742	0.00	0.06	0.00	1.00
Born between 1916 and 1925	3695	0.04	0.19	0.00	1.00	1953	0.04	0.21	0.00	1.00	1742	0.03	0.18	0.00	1.00
Born between 1926 and 1935	3695	0.07	0.25	0.00	1.00	1953	0.07	0.26	0.00	1.00	1742	0.06	0.24	0.00	1.00
Born between 1936 and 1945	3695	0.09	0.29	0.00	1.00	1953	0.07	0.26	0.00	1.00	1742	0.11	0.31	0.00	1.00
Born between 1946 and 1955	3695	0.20	0.40	0.00	1.00	1953	0.17	0.37	0.00	1.00	1742	0.24	0.43	0.00	1.00
Born between 1956 and 1965	3695	0.55	0.50	0.00	1.00	1953	0.56	0.50	0.00	1.00	1742	0.53	0.50	0.00	1.00
Born between 1966 and 1975	3695	0.04	0.21	0.00	1.00	1953	0.07	0.25	0.00	1.00	1742	0.02	0.14	0.00	1.00
Education	3695	1.47	0.50	1.00	2.00	1953	1.00	0.00	1.00	1.00	1742	2.00	0.00	2.00	2.00
Age in 1980	3695	26.87	12.32	5.00	68.00	1953	26.53	13.10	5.00	68.00	1742	27.25	11.39	9.00	68.00
Grade Completed 1980	1362	12.06	1.66	8.00	18.00	747	11.44	0.92	8.00	12.00	615	12.80	2.03	9.00	18.00
Enrolled in 1980	1362	0.57	0.50	0.00	1.00	747	0.33	0.47	0.00	1.00	615	0.86	0.35	0.00	1.00
PV of Earnings†	7152	2.38	1.64	0.00	18.59	3708	1.95	1.14	0.00	11.52	3444	2.83	1.95	0.00	18.59
Tuition at age 17	3695	1.80	0.72	0.00	5.55	1953	1.82	0.74	0.00	5.55	1742	1.76	0.70	0.00	5.55

*Note:

AR=Arithmetic Reasoning

PC=Paragraph Composition

WK= Word Knowledge

MK=Math Knowledge

CS=Coding Speed

†In thousands of Dollars

Table 2.2
List of Variables Included and Excluded in Each System

Variable Name	Cost Function (Z)	Test System (X_M)	Earnings (X)
Urban at age 14	Yes	Yes	No
Parents Divorced	Yes	Yes	No
Number of Siblings	Yes	Yes	No
Father's Education	Yes	Yes	No
Mother's Education	Yes	Yes	No
Born between 1906 and 1915	Yes	No	Yes
Born between 1916 and 1925	Yes	No	Yes
Born between 1926 and 1935	Yes	No	Yes
Born between 1936 and 1945	Yes	No	Yes
Born between 1946 and 1955	Yes	No	Yes
Born between 1956 and 1965	Yes	No	Yes
Born between 1966 and 1975	Yes	No	Yes
Age in 1980	No	Yes	No
Grade Completed 1980	No	Yes	No
Enrolled in 1980	No	Yes	No
Tuition at age 17	Yes	No	No

Table 2.3
Estimated Coefficients in Schooling Choice Equation

Coefficients	Mean	Standard Deviation
Constant	-2.2504	0.3587
Mother's Education	0.2250	0.0274
Father's Education	0.3386	0.0246
Parents Divorced	-0.1976	0.0845
Number of Siblings	-0.1012	0.0163
Urban Residence at age 14	0.1998	0.0755
Dummy birth 1916-1925	0.6076	0.3582
Dummy birth 1926-1935	0.5553	0.3471
Dummy birth 1936-1945	0.7050	0.3417
Dummy birth 1946-1955	0.4160	0.3355
Dummy birth 1956-1965	-0.2064	0.3346
Dummy birth 1966-1975	-1.4159	0.3703
Tuition at 4-year college	-0.0953	0.0447
Loading Factor 1	1.3523	0.1315
Loading Factor 2	0.4785	0.1335
Loading Factor 3	-0.0624	0.1274

Table 2.4

Estimated Coefficients for High School Earnings Equation

	Period Zero		Period One		Period Two		Period Three		Period 4	
Coefficients	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Dummy birth 1916-1925	-	-	-	-	-	-	-	-	-0.1054	0.0829
Dummy birth 1926-1935	-	-	-	-	-	-	-0.0225	0.0974	-0.1443	0.0809
Dummy birth 1936-1945	-	-	-	-	-0.1105	0.1034	-0.0201	0.0989	0.0616	0.1276
Dummy birth 1946-1955	-	-	-0.1779	0.0987	-0.2636	0.0917	0.1657	0.1973	-	-
Dummy birth 1956-1965	-0.7107	0.0637	-0.2936	0.0883	-0.0757	0.1385	-	-	-	-
Dummy birth 1966-1975	-0.6730	0.0960	-0.2360	0.2267	-	-	-	-	-	-
Constant	2.6276	0.0658	2.4021	0.0935	1.8880	0.0870	1.2819	0.0870	0.6147	0.0746
Loading Factor 1	0.1636	0.0433	0.1059	0.0485	0.0164	0.0949	0.0466	0.1122	-0.0077	0.0775
Loading Factor 2	-1.2138	0.0903	-1.6282	0.1142	-1.4415	0.1172	-1.1225	0.1056	-0.3924	0.0763
Loading Factor 3	0.0000	0.0000	0.0000	0.0000	0.2428	0.1684	0.2791	0.1510	0.1327	0.1013

Estimated Coefficients for College Earnings Equation

	Period Zero		Period One		Period Two		Period Three		Period 4	
Coefficients	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Dummy birth 1916-1925	-	-	-	-	-	-	-	-	-0.2976	0.3218
Dummy birth 1926-1935	-	-	-	-	-	-	-0.0881	0.1846	-0.3743	0.3147
Dummy birth 1936-1945	-	-	-	-	-0.0059	0.1710	0.0384	0.1696	-0.2256	0.3457
Dummy birth 1946-1955	-	-	-0.1944	0.1262	-0.0512	0.1568	0.2122	0.2238	-	-
Dummy birth 1956-1965	-0.7375	0.0686	-0.2340	0.1182	-0.1081	0.2910	-	-	-	-
Dummy birth 1966-1975	-0.3459	0.1736	1.3144	0.7365	-	-	-	-	-	-
Constant	2.2802	0.0670	3.5270	0.1191	3.1859	0.1720	2.4843	0.1914	1.3632	0.3367
Loading Factor 1	0.2225	0.0853	0.3137	0.1296	-0.2870	0.2415	-0.2676	0.2656	-0.0144	0.2300
Loading Factor 2	1.0000	0.0000	2.3887	0.1573	2.3194	0.1715	1.7102	0.1806	0.7481	0.1231
Loading Factor 3	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	1.5354	0.1627	0.8876	0.1665

Table 2.5
Estimated Coefficients of Test Equations

	Arithmetic Reasoning		Paragraph Composition		Word Knowledge		Math Knowledge		Coding Speed	
Coefficients	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Constant	-1.1198	0.2256	-1.0262	0.1719	-0.5180	0.2032	-1.4751	0.2265	-1.2706	0.2281
Mother's Education	0.0735	0.0177	0.0529	0.0136	0.0614	0.0158	0.0469	0.0178	0.0561	0.0175
Father's Education	0.0494	0.0136	0.0593	0.0105	0.0461	0.0121	0.0168	0.0139	0.0870	0.0135
Family Income in 1979	0.0008	0.0015	0.0009	0.0012	0.0000	0.0014	0.0038	0.0016	0.0021	0.0015
Parents Divorced	-0.0584	0.0564	-0.0514	0.0440	-0.0947	0.0508	0.0458	0.0569	-0.0138	0.0560
Number of Siblings	-0.0193	0.0111	-0.0397	0.0086	-0.0143	0.0099	-0.0273	0.0115	-0.0313	0.0110
South Residence at age 14	-0.1278	0.0463	-0.0906	0.0358	-0.0064	0.0423	-0.1418	0.0475	-0.1365	0.0464
Urban Residence at age 14	0.0640	0.0461	-0.0243	0.0361	0.0117	0.0422	0.0258	0.0468	0.0529	0.0466
Enrolled at School at Test Date	0.0646	0.0528	-0.0036	0.0403	-0.0515	0.0471	0.0074	0.0527	0.3122	0.0529
Age at Test Date	0.0096	0.0164	0.0237	0.0128	-0.0170	0.0148	0.0048	0.0165	-0.0510	0.0166
Highest Grade Completed at Test Date	0.0911	0.0198	0.0604	0.0155	0.0721	0.0179	0.1082	0.0201	0.1732	0.0198
Loading Factor 1	1.0000	0.0000	0.6801	0.0321	0.8069	0.0377	0.5648	0.0319	0.9562	0.0293
Loading Factor 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Loading Factor 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 3a
Goodness of Fit Tests: Predicted Earnings Densities vs. Actual Densities
The Three-Factor Model

		High School	College	Overall
Period 1	χ^2 Statistic	91.9681	74.2503	204.3823
	Critical Value*	107.5217	82.5287	178.4854
Period 2	χ^2 Statistic	86.6649	107.6417	207.6152
	Critical Value*	116.5110	116.5110	218.8205
Period 3	χ^2 Statistic	26.2658	45.5301	106.5721
	Critical Value*	43.7730	55.7585	91.6702
Period 4	χ^2 Statistic	35.3846	29.7218	55.5758
	Critical Value*	31.4104	30.1435	55.7585
Period 5	χ^2 Statistic	23.2193	14.9131	41.8657
	Critical Value*	23.6848	16.9190	35.1725

* 95% Confidence, equiprobable bins with approx. 15 people per bin

Table 3b
Goodness of Fit Tests: Predicted Earnings Densities vs Actual Earnings Densities
The Two-Factor Model

		High School	College	Overall
Period 1	χ^2 Statistic	109.5702	132.3027	267.4894
	Critical Value*	107.5217	82.5287	178.4854
Period 2	χ^2 Statistic	104.1649	150.5556	247.6732
	Critical Value*	116.5110	116.5110	218.8205
Period 3	χ^2 Statistic	40.7028	61.7322	114.1692
	Critical Value*	43.7730	55.7585	91.6702
Period 4	χ^2 Statistic	39.7253	47.5559	64.2503
	Critical Value*	31.4104	30.1435	55.7585
Period 5	χ^2 Statistic	18.3217	26.5855	40.4078
	Critical Value*	23.6848	16.9190	35.1725

* 95% Confidence, equiprobable bins with approx. 15 people per bin

Table 4.1

Ex-Post Conditional Distributions (College Earnings Conditional on High School Earnings)

$Pr(d_i<Yc\leq d_i+1 \mid d_j<Yh\leq d_j+1)$ where d_i is the ith decile of the College Lifetime Ex-Post Earnings Distribution and d_j is the jth decile of the High School Ex-Post Lifetime Earnings Distribution

$Corrrelation(Y_C,Y_H) = -0.3899$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0035	0.0109	0.0240	0.0326	0.0524	0.7538	0.1137	0.1557	0.2511	0.2808
2	0.0098	0.0244	0.0419	0.0631	0.0894	0.1122	0.1391	0.1747	0.2048	0.1407
3	0.0160	0.0466	0.0741	0.0877	0.1041	0.1213	0.1441	0.1549	0.1581	0.0931
4	0.0236	0.0603	0.0911	0.1062	0.1220	0.1298	0.1348	0.1372	0.1266	0.0683
5	0.0439	0.0848	0.1108	0.1227	0.1303	0.1309	0.1211	0.1139	0.0928	0.0489
6	0.0627	0.1074	0.1214	0.1304	0.1330	0.1218	0.1168	0.0954	0.0695	0.0415
7	0.0963	0.1256	0.1340	0.1334	0.1200	0.1200	0.0937	0.0784	0.0554	0.0433
8	0.1378	0.1659	0.1529	0.1396	0.1114	0.0925	0.0740	0.0561	0.0296	0.0402
9	0.1939	0.1970	0.1498	0.1180	0.1002	0.0771	0.0534	0.0362	0.0200	0.0543
10	0.3354	0.1983	0.1167	0.0812	0.0515	0.0351	0.0266	0.0152	0.0130	0.1271

Table 4.2

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

$Pr(d_i<Yc\leq d_i+1 \mid d_j<Yh\leq d_j+1)$ where d_i is the ith decile of the College Lifetime Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution

Individual expects out θ_3 and $\varepsilon_{s,t}$ for t=0, … , 4, which are unknown by the agent at the time of the schooling choice.

$Corrrelation(Y_C,Y_H) = -0.6993$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0002	0.0079	0.0108	0.0226	0.0421	0.0594	0.0909	0.1447	0.2236	0.3978
2	0.0044	0.0180	0.0286	0.0530	0.0720	0.1010	0.1362	0.1686	0.2114	0.2068
3	0.0106	0.0362	0.0578	0.0786	0.1062	0.1152	0.1498	0.1618	0.1692	0.1146
4	0.0200	0.0546	0.0786	0.1024	0.1204	0.1266	0.1376	0.1406	0.1290	0.0902
5	0.0390	0.0740	0.1004	0.1130	0.1291	0.1387	0.1295	0.1206	0.1010	0.0546
6	0.0454	0.1017	0.1253	0.1353	0.1333	0.1323	0.1189	0.1011	0.0754	0.0314
7	0.0873	0.1299	0.1437	0.1451	0.1299	0.1199	0.0965	0.0777	0.0519	0.0180
8	0.1336	0.1603	0.1613	0.1431	0.1160	0.0974	0.0793	0.0589	0.0389	0.0112
9	0.2063	0.2016	0.1651	0.1293	0.1056	0.0840	0.0540	0.0317	0.0155	0.0068
10	0.4123	0.2318	0.1393	0.0868	0.0556	0.0365	0.0210	0.0110	0.0049	0.0006

Table 5.1
Average present value of earnings* for high school graduates
Fitted and Counterfactual†

White males from NLSY79

	High School (Fitted)	College (counterfactual)
Average Present Value of Earnings	605.92	969.34
Std. Err.	*13.719*	*67.164*

Average returns§ to college for high school graduates

Average returns	1.17
Std. Err.	*0.1350*

*Thousands of dollars. Discounted using a 3% interest rate.

†The counterfactual is constructed using the estimated college outcome equation applied to the population of persons selecting high school

§As a fraction of the base state, i.e. (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Table 5.2
Average present value of earnings[*] for college graduates
Fitted and Counterfactual[†]

White males from NLSY79

	High School (Counterfactual)	College (fitted)
Average Present Value of Earnings	536.43	1007.64
Std. Err.	*26.187*	*35.113*

Average returns[§] to college for college graduates

Average returns	1.33
Std. Err.	*0.0958*

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

[†]The counterfactual is constructed using the estimated high school outcome equation applied to the population of persons selecting college

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

Table 5.3

Average present value of earnings[*] for population of persons indifferent between high school and college

Conditional on education level

White males from NLSY79

	High School	College
Average Present Value of Earnings	571.33	975.16
Std. Err.	*37.066*	*70.557*

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

	High School vs Some College
Average returns	1.26
Std. Err.	*0.3691*

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

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

Table 5.4

Average ex-post, ex-ante and perfect certainty returns*

White males from NLSY79

For people who choose high school			
	ex-post†	ex-ante‡	perfect certainty§
Average	1.1594	1.1594	0.9337
Std. Err.	0.1362	0.1362	0.1154
For people who choose college			
	ex-post†	ex-ante‡	perfect certainty§
Average	1.3398	1.3398	1.6121
Std. Err.	0.1083	0.1083	0.1082
For people indifferent between high school and college			
	ex-post†	ex-ante‡	perfect certainty§
Average	1.2585	1.2585	1.2418
Std. Err.	0.3868	0.3868	0.1067

* Let Y_0, Y_1 denote the present value of earnings in high school and college, respectively. The return to college R is defined as

$$R = \left(\frac{Y_1 - Y_0}{Y_0} \right)$$

† Let I denote the schooling choice index. Let Θ_0 denote the information set of the agent at the time of the schooling choice. Let R denote the return to college. The ex-post mean return to college for a high-school graduate is $E\left(R \mid E_0\left(I\right) < 0\right)$, where $E_0\left(I\right) = E\left(I \mid \Theta_0\right)$. Similarly, the ex-post mean return to college for a college graduate is $E\left(R \mid E_0\left(I\right) \geq 0\right)$. The ex-post mean return to an agent just indifferent between college and high-school is $E\left(R \mid E_0\left(I\right) = 0\right)$.

‡ Let I denote the schooling index. Let Θ_0 denote the information set of the agent at the time of the schooling choice. Let R denote the return to college. The ex-ante mean return to college for a high-school graduate is $E\left(E_0\left(R\right) \mid E_0\left(I\right) < 0\right)$. Similarly, the ex-ante mean return to college for a college graduate is $E\left(E_0\left(R\right) \mid E_0\left(I\right) \geq 0\right)$. The ex-ante mean return to an agent just indifferent between college and high-school is $E\left(E_0\left(R\right) \mid E_0\left(I\right) = 0\right)$. By a property of means, the mean ex-ante and the mean ex-post returns must be equal for the same conditioning set, i.e. $E\left(E_0\left(R\right) \mid E_0\left(I\right) \geq 0\right) = E\left(R \mid E_0\left(I\right) \geq 0\right)$.

§ Let I denote the schooling index. Let R denote the return to college. The return to college under perfect certainty for a high-school graduate is $E\left(R \mid I < 0\right)$. Note that now the agent makes his schooling choice under perfect certainty (that is why we condition on I). Similarly, the return to college under perfect certainty for a college graduate is $E\left(R \mid I \geq 0\right)$. The return to college under perfect certainty for an agent just indifferent between college and high-school is $E\left(R \mid I = 0\right)$.

Table 6.1
Agent's Forecast Variance of Present Value of Earnings* Under Different Information Sets
(fraction of the variance explained by Θ)†
The Calculation is for the Entire Population Regardless of Schooling Choice.

	$\mathrm{Var}(Y_c)$	$\mathrm{Var}\ (Y_h)$	$\mathrm{Var}(Y_c\text{-}Y_h)$
For lifetime:‡			
Variance when $\Theta = \varnothing$	156402.14	73827.89	267796.38
$\Theta = \{\theta_1\}$	0.95%	0.27%	0.44%
$\Theta = \{\theta_1, \theta_2\}$	29.10%	29.43%	47.42%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	68.03%	32.27%	62.65%

*We use an interest rate of 3% to calculate the present value of earnings.

†The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0095)*156402.14

‡Variance of the unpredictable component of earnings between age 19 and 65 as predicted at age 19.

Table 6.2
Agent's Forecast Variance of Period Zero Earnings*
Under Different Information Sets
(fraction of the variance explained by Θ)†
The Calculation is for the Entire Population Regardless of Schooling Choice.

	$\mathrm{Var}(Y_c)$	$\mathrm{Var}\ (Y_h)$	$\mathrm{Var}(Y_c\text{-}Y_h)$
For lifetime:‡			
Variance when $\Theta = \varnothing$	13086.24	14303.35	33910.17
$\Theta = \{\theta_1\}$	1.90%	0.91%	0.05%
$\Theta = \{\theta_1, \theta_2\}$	23.58%	30.08%	41.02%

*We use an interest rate of 3% to calculate the present value of earnings.

†The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0190)*13086.24

‡Variance of the unpredictable component of earnings between age 19 and 28 as predicted at age 19.

Table 6.3
Agent's Forecast Variance of Period One Earnings*
Under Different Information Sets
(fraction of the variance explained by Θ)†
The Calculation is for the Entire Population Regardless of Schooling Choice.

	$\text{Var}(\text{Y}_c)$	$\text{Var}\ (\text{Y}_h)$	$\text{Var}(\text{Y}_c\text{-Y}_h)$
For lifetime:‡			
Variance when $\Theta = \varnothing$	26618.64	17545.90	65804.89
$\Theta = \{\theta_1\}$	1.90%	0.31%	0.34%
$\Theta = \{\theta_1, \theta_2\}$	62.43%	43.00%	69.60%

*We use an interest rate of 3% to calculate the present value of earnings.

†So we would say that the variance of the unpredictable component of period 1 college earnings $\times = \{\theta_1\}$ is (1-0.0190)*26618.64

‡Variance of the unpredictable component of earnings between age 29 and 38 as predicted at age 19.

Table 6.4
Agent's Forecast Variance of Period Two Earnings* Under Different Information Sets
(fraction of the variance explained by Θ)†
The Calculation is for the Entire Population Regardless of Schooling Choice.

	$\mathrm{Var}(\mathrm{Y}_c)$	$\mathrm{Var}\ (\mathrm{Y}_h)$	$\mathrm{Var}(\mathrm{Y}_c\text{-}\mathrm{Y}_h)$
For lifetime:‡			
Variance when $\Theta = \varnothing$	40406.20	16716.50	68918.36
$\Theta = \{\theta_1\}$	0.95%	0.00%	0.63%
$\Theta = \{\theta_1, \theta_2\}$	38.66%	35.02%	58.63%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	75.25%	40.17%	70.98%

*We use an interest rate of 3% to calculate the present value of earnings.

†The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0095)*40406.20

‡Variance of the unpredictable component of earnings between age 39 and 48 as predicted at age 19.

Table 6.5

Agent's Forecast Variance of Period Three Earnings* Under Different Information Sets

(fraction of the variance explained by Θ)†

The Calculation is for the Entire Population Regardless of Schooling Choice.

	$\text{Var}(Y_c)$	$\text{Var}\ (Y_h)$	$\text{Var}(Y_c\text{-}Y_h)$
For lifetime:‡			
Variance when $\Theta = \varnothing$	53194.23	14605.29	66926.12
$\Theta = \{\theta_1\}$	0.65%	0.08%	0.73%
$\Theta = \{\theta_1, \theta_2\}$	16.18%	24.55%	34.65%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	81.20%	31.53%	70.11%

*We use an interest rate of 3% to calculate the present value of earnings.

†The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0065)*53194.23

‡Variance of the unpredictable component of earnings between age 49 and 58 as predicted at age 19.

Table 6.6
Agent's Forecast Variance of Period Four of Earnings* Under Different Information Sets
(fraction of the variance explained by Θ)†
The Calculation is for the Entire Population Regardless of Schooling Choice.

	$\mathrm{Var}(Y_c)$	$\mathrm{Var}\ (Y_h)$	$\mathrm{Var}(Y_c\text{-}Y_h)$
For lifetime:‡			
Variance when $\Theta = \varnothing$	23096.81	10656.83	32236.82
$\Theta = \{\theta_1\}$	0.00%	0.00%	0.00%
$\Theta = \{\theta_1, \theta_2\}$	6.84%	4.10%	11.41%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	56.70%	6.16%	37.95%

*We use an interest rate of 3% to calculate the present value of earnings.

†The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.00)*23096.81

‡Variance of the unpredictable component of earnings between age 59 and 65 as predicted at age 19.