Using Dynamic Models to Measure Uncertainty, Learning and Psychic Costs to Understand Schooling Choices and Schooling Outcomes

James Heckman and Salvador Navarro

ICE Workshop
July 22, 2005

This work is based on

Flavio Cunha, James Heckman and Salvador Navarro
"Seperating Uncertainty from Heterogeneity in Life Cycle Earnings"
Hicks Lecture, Published in April, 2005, *Oxford Economic Papers*

Flavio Cunha, James Heckman and Salvador Navarro
"The Evolution of Uncertainty in the U.S. Economy"
Presentation: World Congress of the Econometric Society, August, 2005

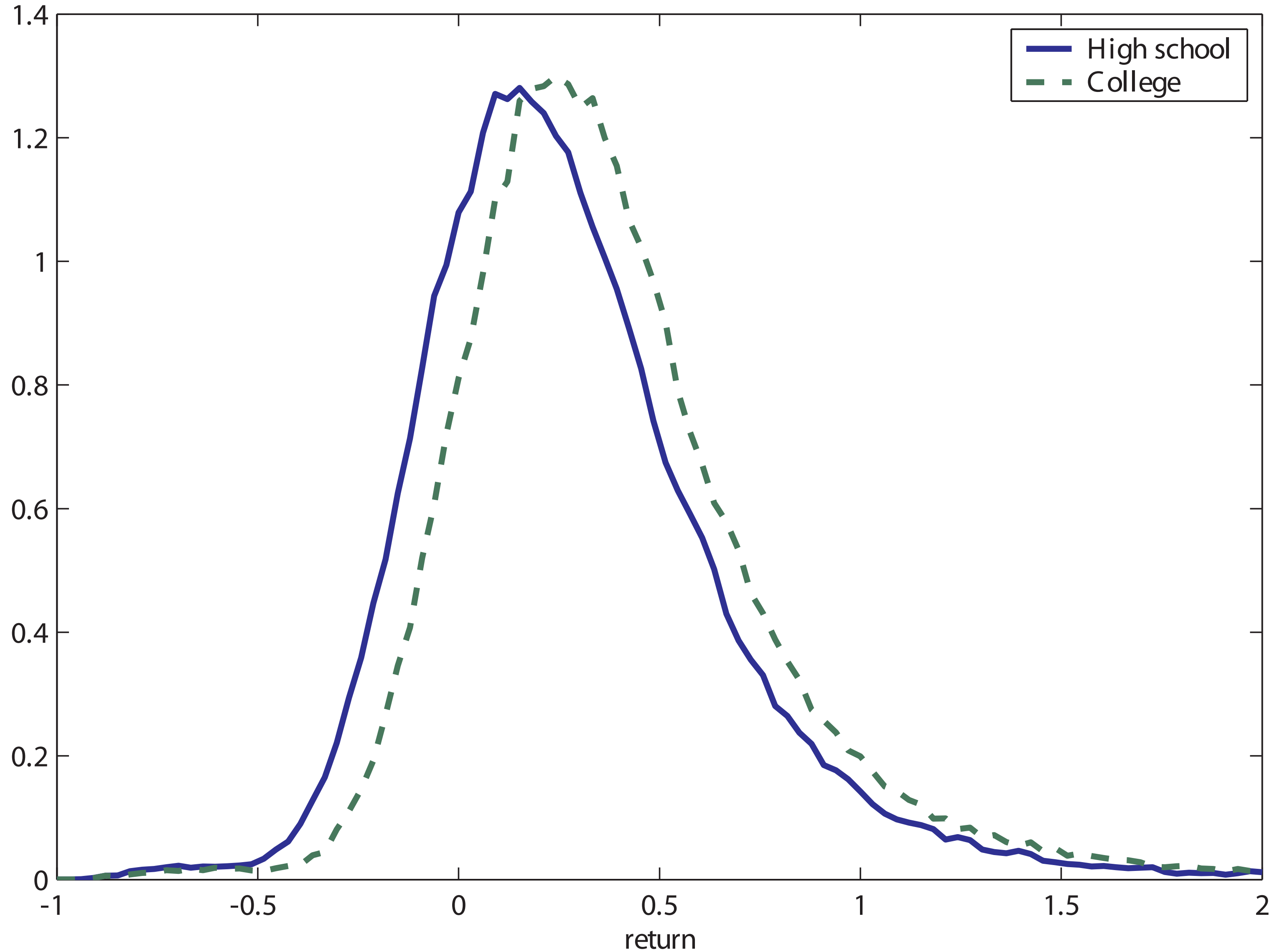
Salvador Navarro
"Understanding Schooling: Using Observed Choices to Infer Agent's Information in a Dynamic Model of Schooling Choice when Consumption Allocation is Subject to Borrowing Constraints"
Ph.D Thesis, University of Chicago, 2005

- Different analyses of schooling choice share the same conclusion: **there is money to be made by going to college and people do not seem to be taking advantage of this opportunity**.

- Define **counterfactual** return (Ignores psychic and monetary costs as is traditional):

$$Return = \frac{\text{PVEarnings College} - \text{PVEarnings High school}}{\text{PVEarnings High school}}$$

Figure 1
Density of realized monetary returns conditional on choice



Present value of earnings from age 19 to 65 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in the 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 the ex-post return as $R=(Y_1-Y_0)/Y_0$. $f(r|S=0)$ denotes the density function of returns for people who choose high school (solid line) and $f(r|s=1)$ denotes the density of ex-post returns for college graduates (dashed line).

- From Figure 1: Big proportion of the people who did not attend college would have obtained positive returns from attending.

- Other methodologies lead to the same basic conclusion. Judd (2000): when compared against assets with similar risk properties, measured returns to schooling are "too high."

- Focus of the analysis has been on the inability of individuals to obtain funds to pay for tuition. Evidence (Cameron and Heckman (2001), Keane and Wolpin (2001), Carneiro and Heckman (2002) and Cameron and Taber (2004)) points towards a small effect of tuition per se once you condition on ability

Not only are high returns NOT being exploited, but the enrollment response to what many people interpret as a rising return to schooling has been very sluggish.

Figure 1
Schooling Participation Rates by Year of Birth: Data from CPS 2000
A. Whites

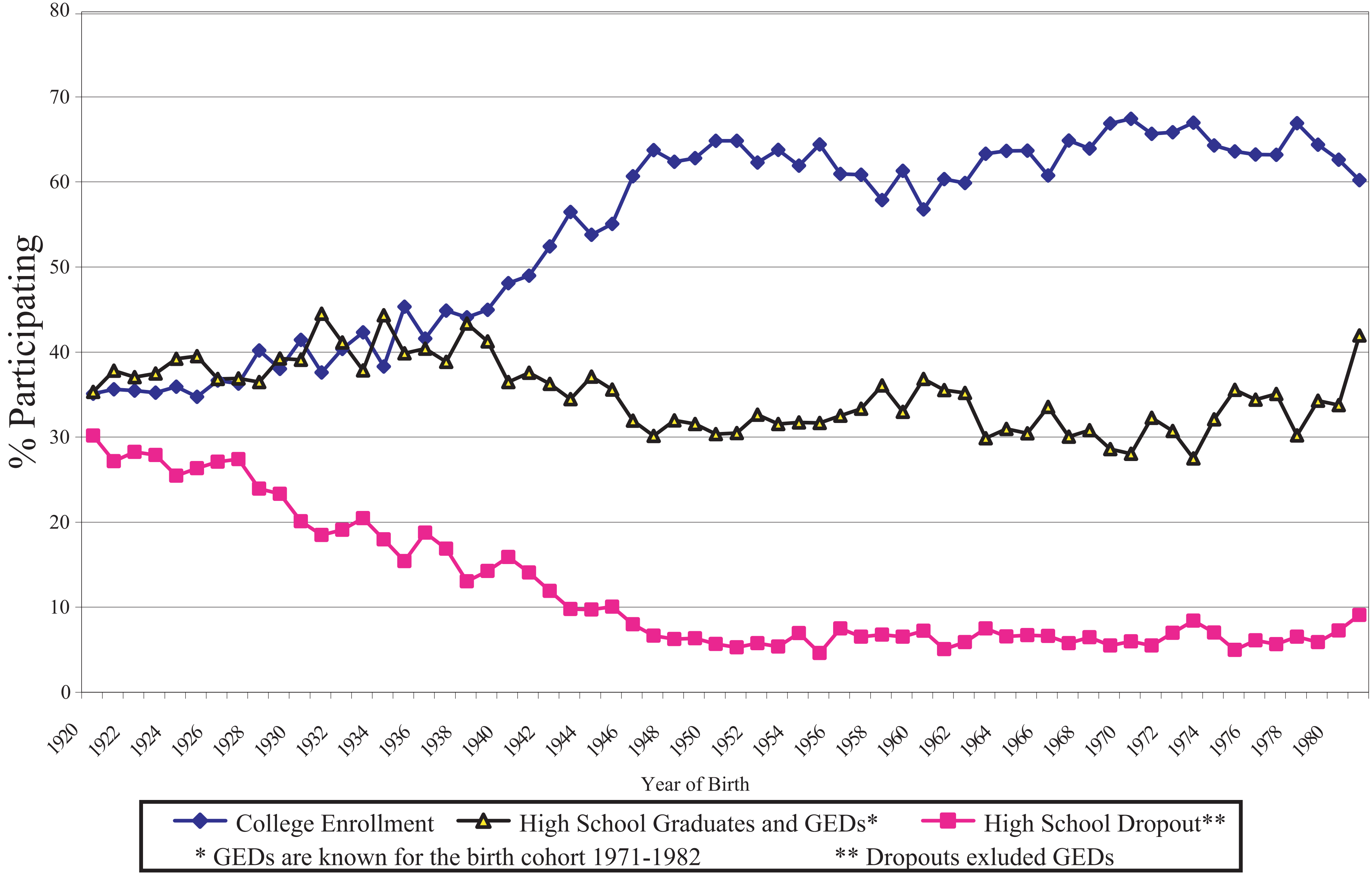


Figure 2
A. College Participation Rates by Year of Birth
Data from CPS 2000

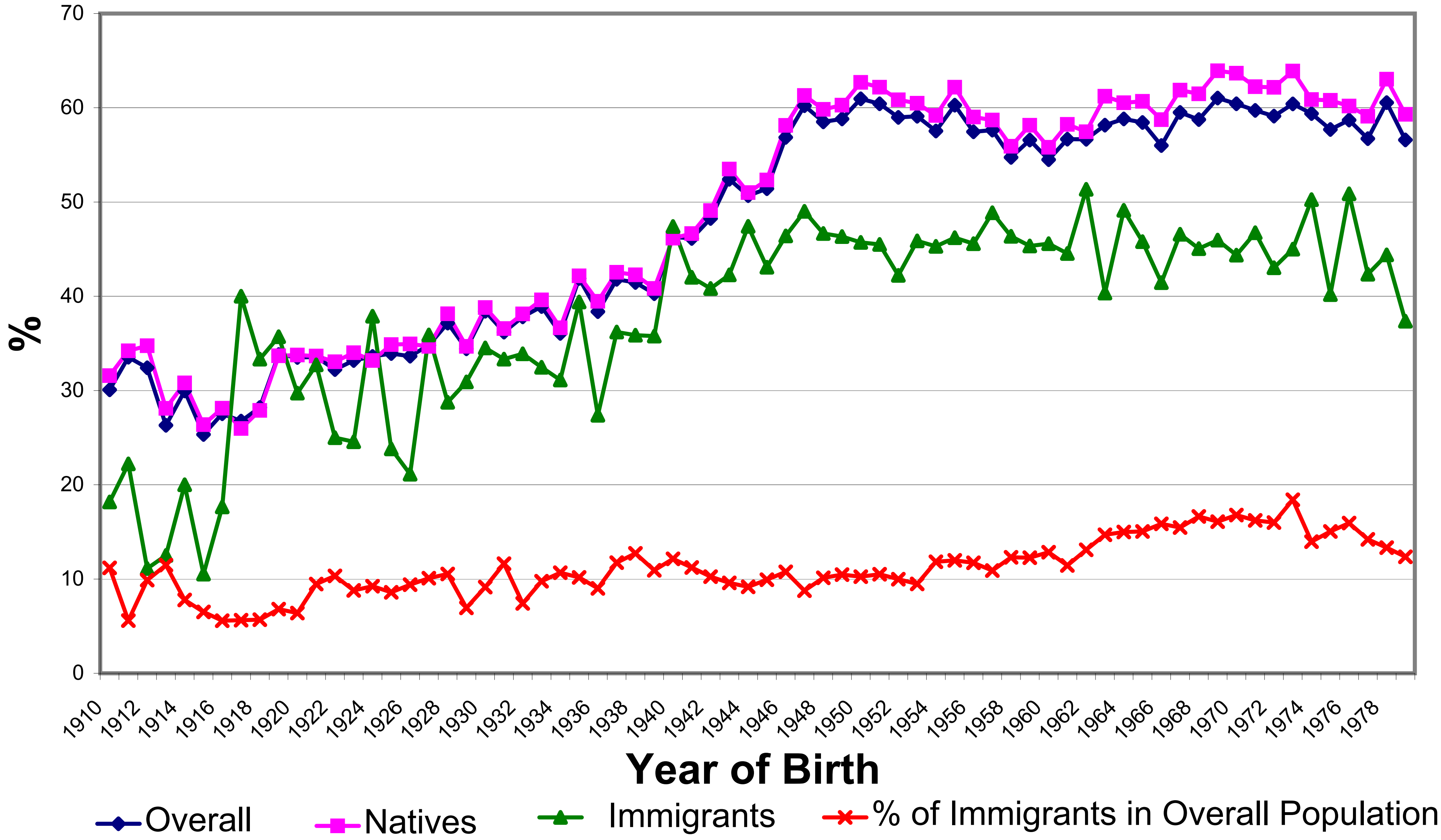
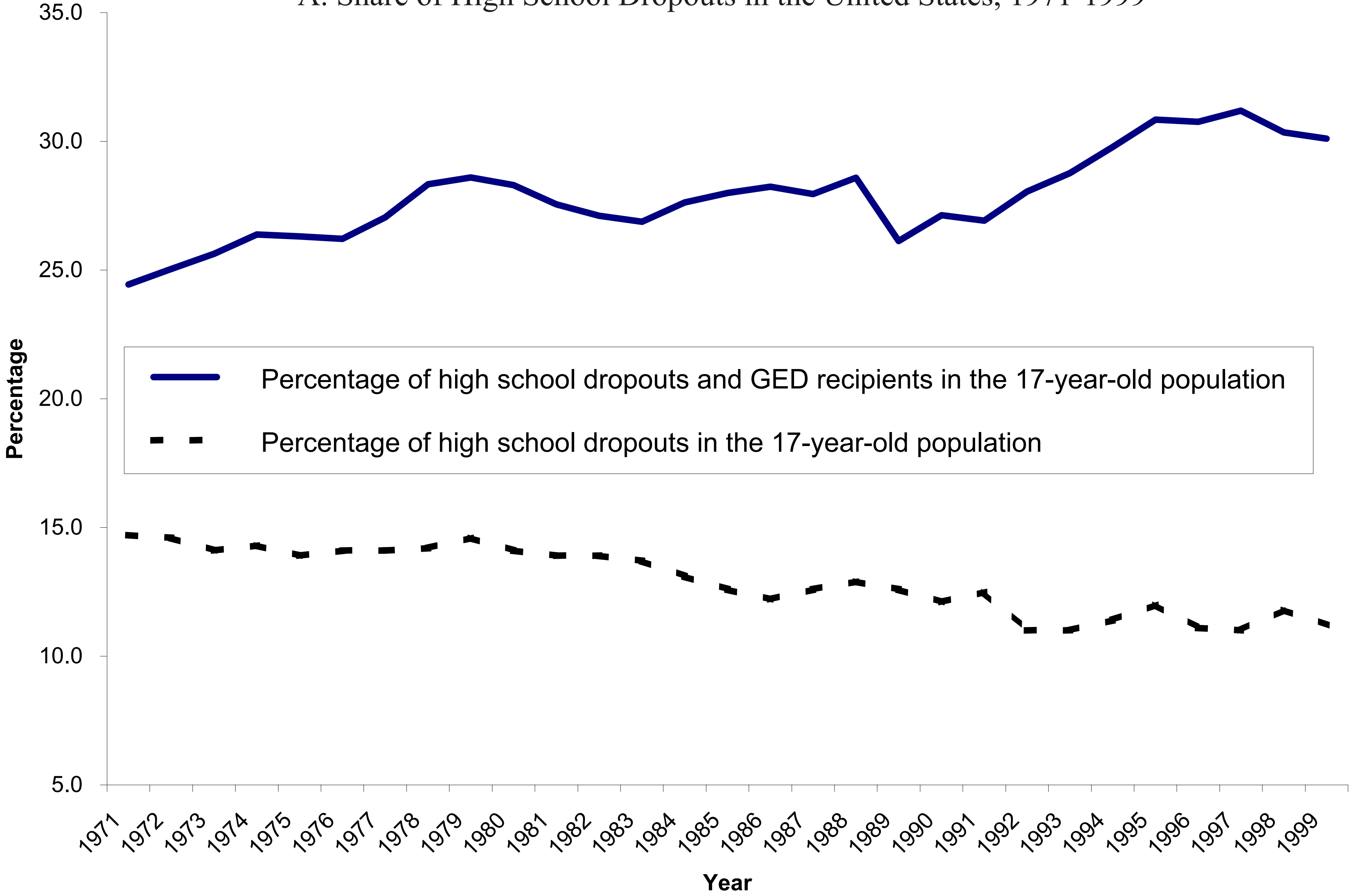


Figure 3
Educational Statistics by Category Over Time

A. Share of High School Dropouts in the United States, 1971-1999



- A feature of figure 1 that has received less attention. A big fraction of people who attend college who get a negative return.

- Credit constraints may partly explain why people with positive returns do not attend college, they cannot explain why people would attend college when returns are negative.

- Our work introduces the **uncertainty** and interaction with credit constraints and preferences.

- Estimate a semiparametrically identified structural model of schooling choice and consumption under borrowing constraints in a partial equilibrium setting.

- What is in the agent's information set? How to measure uncertainty? (Variablity $\neq$ uncertainty). Cannot use variances to impute true uncertainty; Evidence of a lot of predictable heterogeneity.

- Our work develops and implements a methodology to extract information unknown to the econometrician but forecastable by the agent and information unknown to both (fundamental *uncertainty*) is proposed and implemented.

- Key idea: individual choices embody all the information the agent has at the time the decision is being made. Dependence between future outcomes and choices -both consumption and schooling- reveals how much of future outcomes agents know.

- By searching over possible specifications of information sets infer what constitutes uncertainty for the agent.

Conventions About Measuring Uncertainty

- Standard practice in the literature on earnings dynamics writes

$$Y_t = X_t\beta + U_t$$

where Y_t is earnings, (log earnings, wages or log wages) at age t, the X_t are observables at t and the U_t are unobservables from the point of view of the observing economist.

- Statistical decomposition for error process

$$U_t = \delta + \nu_t$$

where δ is a "permanent" component or fixed effect and

$$B(L)\,\nu_t = A(L)\,\eta_t$$

η_t are independent increments.

- Choice of this process is arbitrary, but conventional. An alternative and statistically equivalent process is a factor analytic representation

$$U_t = \theta' \alpha_t + \varepsilon_t$$

where ε_t is independently distributed and θ (the factors) are invariant parameters.

- θ has the natural interpretation of a stand in for missing variables, e.g., ability, motivation, human capital stocks, etc. that affect outcomes differently in different time periods.

- Such decompositions by themselves cannot reveal the information known to, or more precisely, *acted on* by agents.

- Widely assumed that increasing variance$\Longrightarrow$increasing uncertainty. Cunha, Heckman and Navarro show otherwise.

- Standard approaches in labor economics, macroeconomics and finance (see, e.g., Gourinchas and Parker (2002), Carrol (1992), Hubbard, Skinner and Zeldes (1994)), assume that the $\{U_{t+j}\}_{j=1}^{T-t}$ or $\{\nu_{t+j}\}_{j=1}^{T-t}$ are unknown by the agent at date t and constitute a source of fundamental uncertainty, and use residual processes to measure uncertainty. But this is just an arbitrary statistical convention.

- In this paper, in this notation, we distinguish components of $\{U_{t+j}\}_{j=1}^{T-t}$ that are forecastable and *acted on* at date t by agents using the information set $\mathcal{I}_t$ and call them "heterogeneity." The unforecastable components we call "uncertainty".

- In the factor setting for example the $\{\varepsilon_{t+j}\}_{j=1}^{T-t}$ may be unknown as well as some components of θ at date t.

- The goal of this paper is to identify the information set that agents act on (*heterogeneity*) and to separate it from uncertainty in the evolution of life cycle earnings. The welfare implications of heterogeneity and uncertainty are different under different market structures.

- Using data on schooling choices, we separate the components *acted on* by agents in making their schooling decisions from other components of $\{U_{t+j}\}_{j=1}^{T-t}$. We find that a substantial fraction of the variance of lifetime returns to attending college is forecastable and acted on by age 17.

- As a consequence, welfare costs of uncertainty are substantially lower than what would be estimated under the assumption that $\{U_{t+j}\}_{j=1}^{T-t}$ is uncertain to the agent. We estimate that on average, people are willing to pay 5% of their lifetime earnings to eliminate uncertainty when compensated for the change in mean earnings.

Contributions of This Work

- First, we formulate, estimate and identify distinct life cycle earnings processes for persons of different schooling levels and we account for the endogeneity of schooling. The schooling choice equation is a major source of identifying information in separating heterogeneity from uncertainty, but it is not the only source.

- We measure components of the *ex post* realizations that are forecastable and unforecastable as of a given date.

- We examine the fit of the model to data on earnings and schooling under different assumed market structures including (a) complete autarky (first done in Carneiro, Hansen and Heckman, *IER*, 2003), (b) a version of the Aiyagari-Bewley-Schechtman consumption smoothing model with borrowing constraints (Cunha, Heckman and Navarro, 2004) and (c) full insurance (Cunha, Heckman and Navarro, 2003).

- We also look at time series properties of income processes by schooling levels in the multiperiod setting. In this talk, we consider simple and easily exposited two, three and five period models.

Basic Frameworks

- Three stage problem. First, fix the schooling level $s \in \{0, 1\}$ with outcome $Y_{s,t}$, $t = 1, \ldots, T$ (In our other work, we analyze multiple schooling levels with an associated stream of income over time $Y_{s,t}$, $t = 1, \ldots, T$, $s = 0, \ldots, \bar{S}$.)

- Maximize utility by picking a consumption path given schooling level s. For each s solve

$$V_s\left(Y_{s,t}, a_{s,t-1}\right) = \max_{a_{s,t}} \left\{ u\left(c_{s,t}\right) + \frac{1}{1+\rho} \int V_s\left(Y_{s,t+1}, a_{s,t}\right) dF\left(Y_{s,t+1} | \mathcal{I}_t\right) \right\}$$

$$s.t. \quad c_{s,t} = Y_{s,t} + (1+r)\, a_{s,t-1} - a_{s,t}; \quad a_{s,t} \in \mathcal{A}_{s,t}$$

where $c_{s,t}$ is consumption and $a_{s,t}$ is assets. $\mathcal{I}_t$ is information set at time t.

- We let $a_{s,t}$ be restricted to a set $\mathcal{A}_{s,t}$. Different assumptions about the structure of the market imply different restrictions on the elements of $\mathcal{A}_{s,t}$ that are feasible.

- For complete contingent claims markets, we need a more general structure.

Budget Set for Complete Markets

$q_t(\omega)$ Price of an AD security that pays 1 unit of consumption if and only if ω is realized at period t and nothing otherwise.

$a_t(\omega)$ Amount of such AD securities bought by the agents.

ω' Refers to next period.

$$c_{s,t}(\omega) + \int q_t(\omega')a_t(\omega')\, d\omega' = Y_{s,t-1}(\omega) + a_{t-1}(\omega)$$

- For each s, the agent gets the gross utility stream associated with income path s

$$V_s\left(Y_{s,t}, a_{s-1}\right). \quad Y_s \text{ indexes path}$$

- Pick the schooling level with maximum expected utility net of non-pecuniary costs. Choose college if

$$E_{\mathcal{I}_0}\left(V_1\left(Y_{1,1}, a_1\right) - V_0\left(Y_{0,1}, a_0\right) - Cost\right) > 0$$

  where the expectation is taken with respect to the information available to the agent at time 0 ($\mathcal{I}_0$, beginning of adult life when schooling decisions are being made). Work with separable restrictions for convenience only.

- We follow conventions in the literature and assume (when needed) that the utility function is CRRA

$$u\left(c\right) = \frac{c^{1-\phi} - 1}{1 - \phi}$$

  where ϕ is the coefficient of risk aversion (i.e., $\frac{1}{\phi}$ is the eis). In particular, for the Aiyagari model, we get $\phi = 2.00$ ($eis = 0.5$), well within the range of numbers reported in Browning, Hansen and Heckman (1999).

Market settings

Complete markets

- Only idiosyncratic risks. Rich set of contingent commodities can be traded. Known constant interest rate = discount rate (for expositional convenience).

- Optimal consumption path for schooling level s:

$$c_{s,t} = kY_s$$

where $Y_s = E_{\mathcal{I}_0}\left(\sum_{t=1}^{T}\left(\frac{1}{1+r}\right)^{t-1} Y_{s,t}\right)$ is permanent income

- Choose college if

$$E_{\mathcal{I}_0}\left(\left[\frac{Y_1^{1-\phi}-1}{1-\phi}-\frac{Y_0^{1-\phi}-1}{1-\phi}\right] K\sum_{t=1}^{T}\left(\frac{1}{1+\rho}\right)^{t-1} - Cost\right) > 0$$

K is a constant related to k.

- Unless we have good cost measures, we cannot identify ϕ since the ordering of $\frac{Y_1^{1-\phi}}{1-\phi}$ versus $\frac{Y_0^{1-\phi}}{1-\phi}$ does not change for all ϕ. That is, all that matters is expected present value of income.

- Perfect foresight with no borrowing or lending restrictions produces a similar economy.

No credit markets (Carneiro, Hansen and Heckman (*IER*, 2003))

- No possibility of transferring income over time. Known constant interest rate = discount rate .

- Optimal consumption path in schooling level s:

$$c_{s,t} = Y_{s,t}.$$

- Choose college if

$$E_{\mathcal{I}_0}\left(\sum_{t=1}^{T}\left(\frac{1}{1+\rho}\right)^{t-1}\left[\frac{Y_{1,t}^{1-\phi}-1}{1-\phi}-\frac{Y_{0,t}^{1-\phi}-1}{1-\phi}\right]-Cost\right)>0$$

Aiyagari-Bewley-Schechtman model (no contingent assets)

- Only idiosyncratic risks.
- Assume agents can borrow and lend as much as they want (transfer income over time) but cannot transfer income across states of nature (no contingent claims).
- Then, since there are no other assets, agents cannot die in debt:

$$a_{s,T} \geq 0$$

  and this imposes a natural borrowing limit.

- Example: take a two period model. Assume $a_0 = 0$ and that there are no bequest motives.

$$\max_{a_1} \left\{ u\left(Y_{s,1} - a_1\right) + \frac{1}{1+\rho} E_{\mathcal{I}_1}\left[u\left(Y_{s,2} + (1+r)\, a_1\right)\right] \right\}$$

- Let $Y_{\min}$ be the minimum possible income the agent might get. Then

$$a_1 \in \left[-\frac{Y_{\min}}{1+r}, Y_1 \right].$$

- Optimal consumption path in schooling level s is some policy function $c_{s,t} = g_t\left(Y_t\right)$ that gives some utility to each schooling level at $t = 1$.

- Decision rule is then

$$S = 1\left(E_{\mathcal{I}_0}\left(V_1\left(Y_{1,1}, a_0; \phi\right) - V_0\left(Y_{0,1}, a_0; \phi\right) - Cost\right) > 0\right).$$

To demonstrate our approach consider a

Roy economy with two sectors

S denotes different sectors.

$S = 0$ denotes choice of the high school sector, and $S = 1$ denotes choice of the college sector.

C reflects the cost associated with choosing the college sector.

$$Y_1 = \sum_{t=0}^{T} \frac{Y_{1it}}{(1+r)^t}$$

$$Y_0 = \sum_{t=0}^{T} \frac{Y_{0it}}{(1+r)^t},$$

Y_1, Y_0 and C are *ex post* realizations of cost and returns.

$\mathcal{I}_0$ denotes the information set of the agent at time period $t = 0$.

$$S = \begin{cases} 1, & \text{if } E\left(Y_1 - Y_0 - C \mid \mathcal{I}_0\right) \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Essential Idea

Suppose, contrary to what is possible, analyst observes Y_0, Y_1 and C.

Ideal data set

Observe two different lifetimes

Construct $Y_1 - Y_0 - C$ from *ex post* lifetime data.

Information set $\mathcal{I}_0$ of the agent. We seek to construct $E(Y_1 - Y_0 - C \mid \mathcal{I}_0)$. Suppose we assume we know the right information set $\left(\widetilde{\mathcal{I}}_0 = \mathcal{I}_0\right)$.

We then obtain

$$V_{\widetilde{\mathcal{I}}_0} = (Y_1 - Y_0 - C) - E\left(Y_1 - Y_0 - C \mid \widetilde{\mathcal{I}}_0\right)$$

Our test is to determine if S depends on $V_{\widetilde{\mathcal{I}}_0}$.

Test for correct specification of $\mathcal{I}_0$: test if the coefficient on $V_{\widetilde{\mathcal{I}}_0}$ in a discrete choice equation for S is different from zero.

Search among candidate information sets $\widetilde{\mathcal{I}}_0$ to determine which ones satisfy the requirement that the generated $V_{\widetilde{\mathcal{I}}_0}$ does not predict S.

Procedure is in the form of a Sims (1972) version of a Wiener-Granger causality test.

It is also a test for misspecification of the information.

Findings From Our Work

1. Future earnings are highly predictable. However, expectations are not equal to realizations.

2. Under certainty (but with credit constraints) almost 19% of high school graduates would instead choose to be college graduates and almost 23% of college graduates would regret their choice under uncertainty and pick high school instead.

3. Uncertainty also interacts with individual preferences since individuals are risk averse. Estimate a coefficient of relative risk aversion of 2.15. Risk averse individuals "discount" monetary returns in a different way since they dislike uncertainty.

4. Agents may also have preferences over schooling beyond the consumption value of earnings which I capture via an additive "psychic" cost function (big role).

5. Making college tuition free for everyone increases college attendance from 44% to 49%. Upper bound since it includes effect of decrease in price of college.

- Individuals expected utility maximizers who live for T periods. Idiosyncratic risks are associated with earnings. Single riskless asset A that pays a return r.

- $t = 0$ select schooling s (=*h*igh school, *c*ollege). Then, given the (expected) income sequence associated with s, select optimal intertemporal consumption allocation rule.

- $Y_{i,s,t}$: individual i earnings in schooling s. $\mathcal{I}_{i,t}$: information available to the agent at time t. $A_{i,s,t}$: assets transfered to the next period. $u\left(\zeta_{i,s,t}\right)$: individual utility. ρ: discount factor.

- Conditional on s choose consumption to maximize his utility:

$$V_{s,t}\left(\mathcal{I}_{i,t}\right) = max_{A_{i,s,t}} u\left(\zeta_{i,s,t}\right) + \frac{1}{1+\rho} E\left(V_{s,t+1}\left(\mathcal{I}_{i,t+1}\right) \mid \mathcal{I}_{i,t}\right) \tag{1}$$

$$s.t.\ \zeta_{i,s,t} = Y_{i,s,t} + (1+r) A_{i,s,t-1} - A_{i,s,t}$$

$$A_{i,s,0} = 0,\ \ A_{i,s,T} \geq 0.$$

- $A_{i,s,T} \geq 0$: imposes a borrowing constraint at any period.

- Given agents information ($\mathcal{I}_{i,t}$) solution to the consumption allocation problem (1) is

$$\begin{aligned} \zeta^*_{i,s,t} &= g_{s,t}\left(\mathcal{I}_{i,t}\right) \\ V^*_{i,s,t} &= G_{s,t}\left(\mathcal{I}_{i,t}\right) \end{aligned}$$

- Solution requires knowledge of the agent's information at time t ($\mathcal{I}_{i,t}$).

- Get the value associated with each s at time $1\left(V^*_{i,s,1}\right)$ and select schooling. Go to college if

$$E\left(V^*_{i,c,1}\left(\mathcal{I}_{i,1}\right) - V^*_{i,h,1}\left(\mathcal{I}_{i,1}\right) - Cost_i \mid \mathcal{I}_{i,0}\right) > 0. \quad (2)$$

- Individual chooses the schooling level that gives him the maximum expected utility net of psychic costs.

- Earnings for individual i at time t at schooling level s:

$$lnY_{i,s,t} = \mu_{s,t}\left(X_{i,s,t}\right) + U_{i,s,t} \tag{3}$$

  $X_{i,s,t}$: variables that the econometrician observes. $U_{i,s,t}$: variables he cannot observe.

- Agent knows X at all times. $U_{i,s,t}$ is revealed to him at period t.

- He may also know all or part of each $(U_{i,s,\tau}, \tau = t+1, ..., T)$ at time t. Uncertainty is thus associated with $\{U_{i,s,\tau}\}_{\tau=t+1}^{T}$.

- $U_{i,s,t}$ may also include measurement error in earnings (upper bound on uncertainty)

- I also let:

$$Cost_i = \phi\left(Z_i\right) + W_i$$

  Z_i are observable and W_i are unobservables.

- Assume utility is CRRA (common)

$$u(\zeta) = \frac{\zeta^{1-\psi} - 1}{1-\psi}$$

and consumption is measured with error:

$$\widehat{\zeta}_{i,t} = \zeta_{i,t} \eta_t \left(K_{i,t}, \xi_{i,t}\right) \tag{4}$$

where $\zeta_{i,t}$ is "true" consumption, $\widehat{\zeta}_{i,t}$ is measured consumption and $\eta_t \left(K_{i,t}, \xi_{i,t}\right) = exp(K_{i,t} \upsilon_t + \xi_{i,t})$ is multiplicative measurement error.

- The econometrician must know $\mathcal{I}_{i,t}$ to solve the model. To find a procedure that allows the analyst to infer the components of the agent's information set I further assume

$$U_{i,s,t} = \theta_i \alpha_{s,t} + \varepsilon_{i,s,t} \tag{5}$$

$$W_i = \theta_i \lambda + \omega_i$$

- θ_i is a **vector** of mean zero mutually independent "factors". Uniquenesses: $\varepsilon_{i,s,t}$ and ω_i also mean zero. Uniquenesses, factors and measurement error on consumption, $\xi_{i,t}$, all mutually independent of each other for all schooling levels s and time periods t.

- Natural starting point: elements of the vector θ_i represent missing variables like ability or skills and α prices.

- Only a statistical decomposition. By themselves, not informative about what is known to the agent at t.

- Interpret elements of θ_i as permanent shocks that activate and influence earnings at different points in time.

- Assume: earnings in the first τ_1 affected only by the first element of θ_i. Next τ_2 periods affected by the first two elements of θ_i, and so on.

- These elements of θ_i revealed to the agent through their effect on earnings. Let $\overline{\theta}_i(t)$ denote those elements of θ_i that affect earnings after t.

- Agent might forecast elements of θ_i that affect future earnings but do not affect past and currently observed outcomes.

- **(I-1)** The information revelation process of the agent is such that he either knows $\theta_{i,l}$ $(l = 1, ..., L)$ or he does not. Revelation of information when it happens, is instantaneous.

- **(I-2)** At period t, the agent observes his outcomes for the period and knows $\varepsilon_{i,s,t}$ and any elements of θ_i that affect outcomes in that period (or in any previous periods). That is, if $\theta_{i,l}$ affects outcomes at $\tau < t$, then it is known by the agent at time t.

- **(I-3)** Agents have rational expectations so that the expectations they take and the mathematical expectation operator with respect to the actual distributions in the model coincide.

- Agent has knowledge of the parameters of the model $(e.g.,\ \rho, \psi, \mu(X),$ $\phi(Z))$ as well as X_i, X_i^M, K_i, Z_i and ω_i. The econometrician never observes θ_i. By assumption, $\{\varepsilon_{i,s,\tau}\}_{\tau=t+1}^{T}$ is not part of the agent's information set $\mathcal{I}_{i,t}$.

- Cast the problem of determining agent information as a testing problem.
- First test: fit the model assuming different information sets for the agent, test which specification fits the data better penalizing for parameters.

An Example with Linear Consumption

- Assumptions in the example: extreme oversimplifications not required.
- Individual already chose schooling. Assume θ_i two elements. Suppose:

$$lnY_{i,s,t} = \mu_{s,t}\left(X_{i,s,t}\right) + \theta_{i,1}\alpha_{s,t,1} + \varepsilon_{i,s,t}$$

$$lnY_{i,s,t+1} = \mu_{s,t+1}\left(X_{i,s,t+1}\right) + \theta_{i,1}\alpha_{s,t+1,1} + \theta_{i,2}\alpha_{s,t+1,2} + \varepsilon_{i,s,t+1}.$$

- Suppose $\theta_{i,1}$ and $\theta_{i,2}$ known by the agent at time t:

$$max_{\zeta_{i,s,t}} u\left(\zeta_{i,s,t}\right) + \frac{1}{1+\rho}\int V_{s,t+1}\left(\mathcal{I}_{i,t+1}\right) dF\left(\varepsilon_{i,s,t+1}\right) \quad (6)$$

agent expects unobserved $\varepsilon_{i,s,t+1}$ out.

- If information set does not contain $\theta_{i,2}$ agent expects it out

$$max_{\zeta_{i,s,t}} u\left(\zeta_{i,s,t}\right) + \frac{1}{1+\rho}\int\int V_{s,t+1}\left(\mathcal{I}_{i,t+1}\right) dF\left(\varepsilon_{i,s,t+1}\right) dF\left(\theta_{i,2}\right) \quad (7)$$

- First case consumption is a function of $\theta_{i,1}$ and $\theta_{i,2}$. Further simplify:

$$ln\widehat{\zeta}_{i,s,t} = \Lambda_{s,t}\left(X_{i,s,t}\right) + \theta_{i,1}\delta_1 + \theta_{i,2}\delta_2 + \eta_{i,t}. \tag{8}$$

$$cov\left(ln\widehat{\zeta}_{i,s,t}, lnY_{i,s,t+1} \mid X\right) = \delta_1\alpha_{s,t+1,1}\sigma^2_{\theta_1} + \delta_2\alpha_{s,t+1,2}\sigma^2_{\theta_2}. \tag{9}$$

- Second case consumption does not depend on $\theta_{i,2}$:

$$ln\widehat{\zeta}_{i,s,t} = \Lambda^*_{s,t}\left(X_{i,s,t}\right) + \theta_{i,1}\delta^*_1 + \eta^*_{i,t}, \tag{10}$$

$$Cov\left(ln\widehat{\zeta}_{i,s,t}, lnY_{i,s,t+1} \mid X\right) = \delta^*_1\alpha_{s,t+1,1}\sigma^2_{\theta_1}. \tag{11}$$

- Assume model is fit under both information sets and agent knows both $\theta_{i,1}, \theta_{i,2}$. The model with both $\theta_{i,1}$ and $\theta_{i,2}$ entering the information set at t would be preferred.

- If the true information set only contains $\theta_{i,1}$ both models will fit the data equally well. The estimate of δ_2 in equation (8) will be zero. Right model would be selected again.

- Alternative test is to see if the coefficients on these factors are zero? Not quite if model not linear (same logic though)

- Use the intuition from linear example. If econometrician estimates the model using the right information set $\mathcal{I}_{i,t}$

$$ln\widehat{\zeta}_{i,s,t} = lng_{s,t}\left(\mathcal{I}_{i,t}\right) + ln\eta_t\left(K_{i,t}, \xi_{i,t}\right),$$

Policy function $g_{s,t}$ equals "true" consumption.

- This equation forms part of the contribution of individual i who selects schooling level s to the sample likelihood:

$$\int_{\Theta} \begin{array}{c} \prod_{t=1}^{T} f\left(lnY_{i,s,t} \mid \theta_i, X_i\right) \\ \Pr\left(S_i = s \mid \theta_i, Z_i, X_i\right) \\ \prod_{t=1}^{T-1} f\left(ln\widehat{\zeta}_{i,s,t} \mid \theta_i, X_i, Z_i, K_i\right) \end{array} dF\left(\theta\right).$$

- Suppose model is estimated using a candidate information set $\widetilde{\mathcal{I}}_{i,t}$. Likelihood misspecified model based on

$$ln\widehat{\zeta}_{i,t} = ln\widetilde{g}_{s,t}\left(\widetilde{I}_{i,t}\right) + ln\eta_t\left(K_{i,t}, \xi_{i,t}\right) \tag{12}$$

$ln\widetilde{g}_{s,t}$ under the assumption that $\overline{\theta}_i\left(t\right)$ is expected out by the agent.

- Instead of basing the likelihood on equation (12) use

$$ln\widehat{\zeta}_{i,t} = ln\widetilde{g}_{s,t}\left(\widetilde{I}_{i,t}\right) + \ln\eta_t\left(K_{i,t}, \xi_{i,t}\right) + \Phi_t\left(\overline{\theta}_i\left(t\right)\right)\Delta_t,$$

where $\Phi_t\left(\right)$ is some function of $\overline{\theta}_i\left(t\right)$, for example polynomials in the components of $\overline{\theta}_i\left(t\right)$.

- Consumption predicted by $\widetilde{g}_{s,t}$ will not depend on $\overline{\theta}_i\left(t\right)$. Actual consumption may contain elements of $\overline{\theta}_i\left(t\right)$.

- Test those elements of Δ equal zero.

- **The basic idea is that if element l of $\overline{\theta}_i\left(t\right)$ belongs in the information set of the agent at time t and the agent acts on it, it will affect the choices he makes at t.**

- Same idea to agent schooling choices. Instead of

$$E\left(G_{c,1}\left(\mathcal{I}_{i,1}\right) - G_{h,1}\left(\mathcal{I}_{i,1}\right) - Cost_i \mid \mathcal{I}_{i,0}\right) > 0,$$

estimating the model under $\widetilde{\mathcal{I}}_{i,0}$ and use

$$E\left(\widetilde{G}_{c,1}\left(\mathcal{I}_{i,1}\right) - \widetilde{G}_{h,1}\left(\mathcal{I}_{i,1}\right) - Cost_i \mid \widetilde{\mathcal{I}}_{i,0}\right) + \Phi_0\left(\overline{\theta}_i\left(0\right)\right)\Delta > 0$$

- Intuitive sketch of identification. Identification theory used to understand what in principle can be recovered nonparametrically.

- Use flexible parametric forms to obtain estimates of my high dimensional econometric model.

- The question of identifiability is a separate issue and should be judged independently of the choice of parametric forms for estimation purposes.

- Assume θ_i is a scalar (extends to multifactor case).

- Assume problem of selection is solved by using limit set arguments.

- Without loss of generality take the system of log earnings equations for high school:

$$lnY_{i,h,t} = \mu_{h,t}\left(X_{i,h,t}\right) + \theta_i \alpha_{h,t} + \varepsilon_{i,h,t}, \qquad t \geq 1$$

- Factor θ_i has no natural scale, *i.e.*, $\theta_i \alpha = \kappa \theta_i \frac{\alpha}{\kappa}$ for any constant κ. Normalize $\alpha_{h,1} = 1$.

- X independent of $\{U_{i,h,t}\}_{t=1}^{T}$. Form covariance matrix of high school log earnings from the data:

$$\frac{cov\left(lnY_{i,h,t}, lnY_{i,h,t'} \mid X\right)}{cov\left(lnY_{i,h,1}, lnY_{i,h,t'} \mid X\right)} = \frac{\alpha_{h,t}\alpha_{h,t'}\sigma_{\theta}^{2}}{\alpha_{h,t'}\sigma_{\theta}^{2}} = \alpha_{h,t}, \qquad t \neq t'.$$

$$\sigma_{\theta}^{2} = \frac{cov\left(lnY_{i,h,1}, lnY_{i,h,t} \mid X\right)}{\alpha_{h,t}}, \qquad t = 1, ...T.$$

$$var\left(lnY_{i,h,t} \mid X\right) - \alpha_{h,t}^{2}\sigma_{\theta}^{2} = \sigma_{\varepsilon_{h,t}}^{2}.$$

- If θ_i and $\{\varepsilon_{i,h,t}\}_{t=1}^{T}$ are normally distributed: done. Normality not required. The distributions of both θ_i and $\{\varepsilon_{i,h,t}\}_{t=1}^{T}$ can be nonparametrically identified.

- We can never form covariances of earnings across schooling levels since earnings are not observed on both schooling levels for anyone.

- If the distribution of θ_i is nonsymmetric, no problem. If θ_i has a symmetric distribution suppose access to:

$$M_{i,j} = \mu_{j}^{M}\left(X_{i,j}^{M}\right) + \theta_i\alpha_{j}^{M} + \varepsilon_{i,j}^{M}, \qquad j = 1, ...J, \tag{13}$$

- Taking the covariance of college earnings with respect to a measurement equation

$$cov\left(lnY_{i,c,t}, M_{i,j} \mid X, X^{M}\right) = \alpha_{c,t}\alpha_{j}^{M}\sigma_{\theta}^{2}$$

- Risk aversion parameter identified using Euler equation arguments

$$ln\left(\frac{\widehat{\zeta}_{i,t+1}}{\widehat{\zeta}_{i,t}}\right) = \frac{1}{\psi}ln\frac{1+r}{1+\rho} - K_{i,t}\upsilon_{t} + K_{i,t+1}\upsilon_{t+1} + \pi_{i,t} \tag{14}$$

- Since the left hand side of

$$ln\widehat{\zeta}_{i,t} - K_{i,t}\upsilon_{t} = lng_{s,t}\left(\mathcal{I}_{i,t}\right) + \xi_{i,t}$$

is known and so is the distribution of $lng_{s,t}\left(\mathcal{I}_{i,t}\right)$ we can recover the distribution of $\xi_{i,t}$ by deconvolution.

- Estimated on a sample of white males that either graduated high school (and only high school) or are college graduates.
- Individuals from both the NLSY79 and PSID datasets pooled together. This requires integrating out missing information.
- Add test scores (M_j) as a function of θ_1:

$$M_{i,j} = X_i^M \beta_j^M + \theta_{i,1} \alpha_j^M + \varepsilon_{i,j}^M.$$

- Tests are assumed to be noisy proxies for **ability** which is given by θ_1.

- If unemployed: earnings are zero. If missing: imputed from the surrounding years.

- Simplify lifecycles to 5 periods: log of the present value of earnings for the period discounted at 3%.

- I find that $lnY_{i,s,t}$ being generated by a two factor model:

$$lnY_{i,s,t} = X_i\beta_{i,s,t} + \theta_{i,1}\alpha_{s,t,1} + \theta_{i,2}\alpha_{s,t,2} + \varepsilon_{i,s,t}.$$

  is enough to fit the data.

- Factors $\theta_{i,l}$ allowed to follow mixture of normals distribution. Uniquenesses $\varepsilon_{i,s,t}$ are assumed to be mixtures of normals, truncated between -4 and 4.

- Estimation of the model is done by maximum likelihood: combination of simulated annealing, the Nelder-Meade simplex method and the Broyden-Fletcher-Goldfarb-Shanno variable metric algorithm. The contribution of individual i who chooses schooling $S_i = s$

$$\int_{\Theta} \begin{bmatrix} \prod_{t=1}^{5} f_{\varepsilon_{i,s,t}} \left(lnY_{i,s,t} - \mu_{s,t}\left(X_{i,s,t}\right) - \theta_i \alpha_{s,t} | \theta_i, X_{i,s,t} \right) \\ f_{\varepsilon_{i,j}^M} \left(lnM_{i,j} - \mu_j^M \left(X_{i,j}^M \right) - \theta_i \alpha_j^M | \theta_i, X_{i,j}^M \right) \\ Pr\left(S_i = s | X_{i,s,t}, Z_i, \theta_i\right) \\ \prod_{t=1}^{4} f_{\xi_{i,t}} \left(ln\widehat{\zeta}_{i,t} - lng_{s,t}\left(\mathcal{I}_{i,t}\right) - K_{i,t} \upsilon_t | \theta_i, X_{i,s,t}, Z_i, K_{i,t} \right) \end{bmatrix} dF\left(\theta\right).$$

- Evaluation of the likelihood requires solution of dynamic program (for a given proposed $\mathcal{I}_{i,t}$)

- Computational issues associated with estimating a model like this one.
- Start with the solution of the dynamic program in equation (**??**) under different information sets.
- Individual who has already chosen his schooling level s. For simplicity

$$\mu_{s,t} = \mu_{s,t}\left(X_{s,t}\right).$$

- Let

$$R_{s,t}\left(\mu_{s,t}, \theta, \varepsilon_{s,t}, A_{s,t-1}\right) = Y_{s,t} + (1+r)\, A_{s,t-1}$$

denote the resources available to the agent at time t.

- Start with the last period T and assume that $\mathcal{I}_t$ contains all the elements of the L-dimensional vector θ.
- The value function at T is given by the utility the agent obtains from consuming his remaining resources

$$V^*_{s,T} = G_{s,T}\left(R_{s,T}\left(\mu_{s,T}, \theta, \varepsilon_{s,T}, A_{s,T-1}\right)\right) = u\left(R_{s,T}\right).$$

- At $T-1$ (still assuming the whole vector θ is in $\mathcal{I}_t$) the agent's optimization problem is

$$max_{A_{s,T-1}} u\left(R_{s,T-1}\left(\mu_{s,T-1},\theta,\varepsilon_{s,T-1},A_{s,T-2}\right)-A_{s,T-1}\right)$$
$$+\frac{1}{1+\rho}\int G_{s,T}\left(R_{s,T}\left(\mu_{s,T},\theta,\varepsilon_{s,T},A_{s,T-1}\right)\right)dF\left(\varepsilon_{s,T}\right),$$

so the agent expects the unknown $\varepsilon_{s,T}$ out. The solution to this problem is a function

$$A^*_{s,T-1}=\Gamma_{s,T-1}\left(R_{s,T-1}\left(\mu_{s,T-1},\theta,\varepsilon_{s,T-1},A_{s,T-2}\right),\mu_{s,T}+\theta\alpha_{s,T}\right)$$

that establishes, for a given amount of resources and a given mean log earnings at T, how much the agent will save for the next period via $A_{s,T-1}$.

- The associated value function is

$$V^*_{s,T-1}=G_{s,T-1}\left(R_{s,T-1}\left(\mu_{s,T-1},\theta,\varepsilon_{s,T-1},A_{s,T-2}\right),\mu_{s,T}+\theta\alpha_{s,T}\right)=$$

$$u\left(R_{s,T-1}-A^*_{s,T-1}\right)+\frac{1}{1+\rho}\int G_{s,T}\left(R_{s,T}\left(\mu_{s,T},\theta,\varepsilon_{s,T},A^*_{s,T-1}\right)\right)dF\left(\varepsilon_{s,T}\right)$$

- Proceeding sequentially we find that, if θ is known by the agent at $t+1$, the policy and value functions describing the solution at $t+1$ are

$$A^*_{s,t+1} = \Gamma_{s,t+1}\left(R_{s,t+1}, \left\{\mu_{s,\tau} + \theta\alpha_{s,\tau}\right\}_{\tau=t+2}^{T}\right)$$

$$V^*_{s,t+1} = G_{s,t+1}\left(R_{s,t+1}, \left\{\mu_{s,\tau} + \theta\alpha_{s,\tau}\right\}_{\tau=t+2}^{T}\right).$$

- Notice that the number of arguments in the function grows since we need to keep track of all future $\left\{\mu_{s,\tau} + \theta\alpha_{s,\tau}\right\}_{\tau=t+2}^{T}$ and not only next period's.

- Suppose that the information set proposed is such that at time t the agent does not know the last element of θ. In this case, the agent's consumption allocation problem is

$$max_{A_{s,t}} u\left(R_{s,t}\left(\mu_{s,t}, \{\theta_l\}_{l=1}^{L-1}, \varepsilon_{s,t}, A_{s,t-1}\right) - A_{s,t}\right)$$
$$+ \frac{1}{1+\rho} \int \int G_{s,t+1}\left(R_{s,t+1}\left(\mu_{s,t+1}, \theta, \varepsilon_{s,t+1}, A_{s,t}\right)\right) dF\left(\varepsilon_{s,t}\right) dF\left(\theta_L\right)$$

and the functions that describe the solution of the problem are

$$A^*_{s,t} = \Gamma_{s,t}\left(R_{s,t}\left(\mu_{s,t}, \{\theta_l\}_{l=1}^{L-1}, \varepsilon_{s,t}, A_{s,t-1}\right), \left\{\mu_{s,\tau} + \sum_{l=1}^{L-1} \theta_l \alpha_{s,\tau,l}\right\}_{\tau=t+1}^{T}\right)$$

$$V^*_{s,t} = G_{s,t}\left(R_{s,t}\left(\mu_{s,t}, \{\theta_l\}_{l=1}^{L-1}, \varepsilon_{s,t}, A_{s,t-1}\right), \left\{\mu_{s,\tau} + \sum_{l=1}^{L-1} \theta_l \alpha_{s,\tau,l}\right\}_{\tau=t+1}^{T}\right).$$

- Since the agent integrates θ_L out and assumptions (I-1) - (I-3) rule out the possibility of θ_L affecting earnings for $\tau \leq t$, neither the policy function nor the value function depend on θ_L.

- The solution of the problem from time period t on will be a function only of the elements of the model known to the agent so it will not include θ_L.

- Neither the value functions nor the policy functions have an algebraic solution.

- At any given period t, the dynamic program is solved for a grid of points on $R_{s,t}$ and $\left\{\mu_{s,\tau} + \sum_{l=1}^{L-1} \theta_l \alpha_{s,\tau,l}\right\}_{\tau=t+1}^{T}$.

- These solutions are then used to approximate the functions by regressing the solutions in the grid against polynomials on $R_{s,t}$ and $\left\{\mu_{s,\tau} + \sum_{l=1}^{L-1} \theta_l \alpha_{s,\tau,l}\right\}_{\tau=t+1}^{T}$.

- Once the functions that solve the allocation problem are found *for a given value of the parameters of the model*, the likelihood can be evaluated. The contribution to the likelihood of an individual who chooses, for example, $S = c$ is

$$\int_{\Theta} \begin{array}{c} \prod_{t=1}^{5} f_{\varepsilon_{c,t}} \left(lnY_{c,t} - \mu_{c,t}\left(X\right) - \theta\alpha_{c,t}|\theta, X_t\right) Pr\left(S = c|Z,\theta\right) \\ \prod_{t=1}^{4} f_{\xi_t} \left(ln\widehat{\zeta}_t - lng_{c,t}\left(\mathcal{I}_t\right) - K_t v_t|\theta, K\right) \end{array} dF\left(\theta\right).$$

- Evaluation of the likelihood requires the econometrician to solve the dynamic program in order to evaluate the schooling selection probability and the consumption policy function $g_{c,t}$.

- Notice that, since the econometrician never observes any element of θ, he has to integrate against its distribution when evaluating the likelihood.

Table 3.2

Tests for Misspecification: Model Selection Criteria

	θ_2 not known at periods 1 and 2	θ_2 not known at period 1	θ_1 and θ_2 not known at schooling decision date	θ_1 and θ_2 always known
Loglikelihood:	-18395.18	-18013.03	-16864.99	-16820.04
AIC:	18636.18	18254.03	17107.99	17060.04
BIC:	19388.18	19006.03	17866.22	17808.92

AIC (Akaike Information Criteria) = -Loglikelihood + Number of Parameters

BIC (Schwarz Criteria) = -Loglikelihood + 0.5*Number of Parameters*log(n)

In both cases a smaller number means we favor the selection of that model

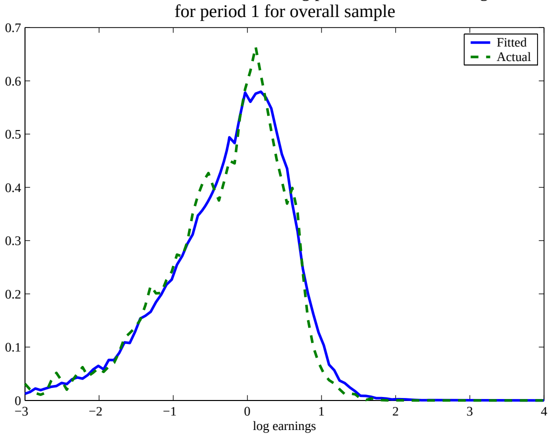
Table 3.1

Tests for Information Set Misspecification

Additional Parameters	θ_2 not known at periods 1 and 2: Schooling Choice	θ_2 not known at periods 1 and 2: Consumption Age 19 – 24	θ_2 not known at periods 1 and 2: Consumption Age 25 – 30	θ_2 not known at period 1: Schooling Choice	θ_2 not known at period 1: Consumption Age 19 – 24	θ_1 and θ_2 not known at schooling decision date: Schooling Choice
θ_1	-	-	-	-	-	0.33
Std. Error	-	-	-	-	-	0.00
$\theta_1\theta_1$	-	-	-	-	-	-0.20
Std. Error	-	-	-	-	-	0.00
θ_2	0.81	0.78	0.65	0.87	0.91	-1.10
Std. Error	0.02	0.03	0.03	0.02	0.03	0.00
$\theta_2\theta_2$	1.84	0.84	1.83	1.18	0.95	0.04
Std. Error	0.10	0.05	0.05	0.11	0.05	0.00
$\theta_1\theta_2$	-	-	-	-	-	-0.64
Std. Error	-	-	-	-	-	0.00

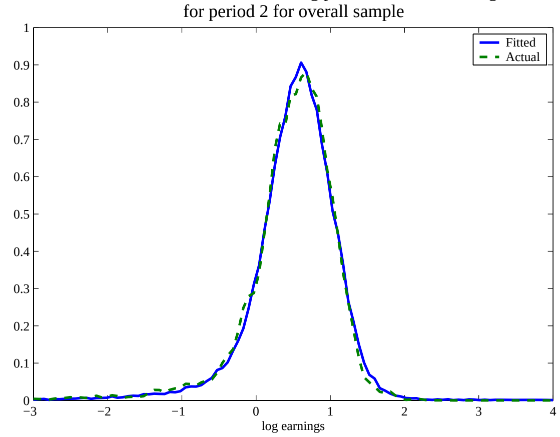
Let g(I) be the function describing the predicted choice in the model as a function of the assumed information set I. We then add the left out factors to the choice function (after the agent integrates them out) and test whether their associated parameters are different from zero.

Figure 2.1
Densities of fitted and actual log present value of earnings
for period 1 for overall sample



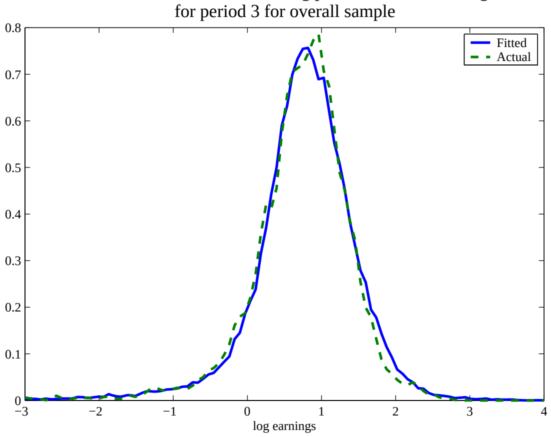
Log of present value of earnings from age 19 to 24 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(log(y)) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.2
Densities of fitted and actual log present value of earnings
for period 2 for overall sample



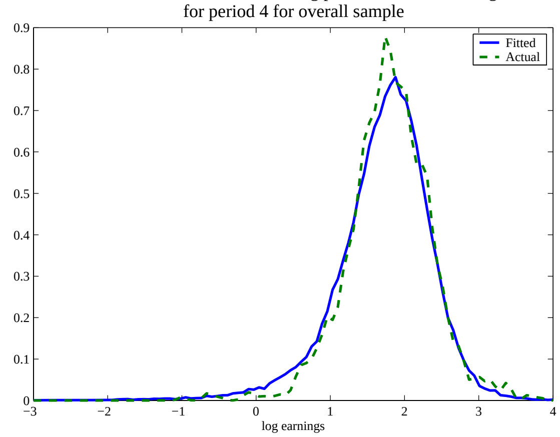
Log of present value of earnings from age 25 to 30 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(log(y)) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.3
Densities of fitted and actual log present value of earnings
for period 3 for overall sample



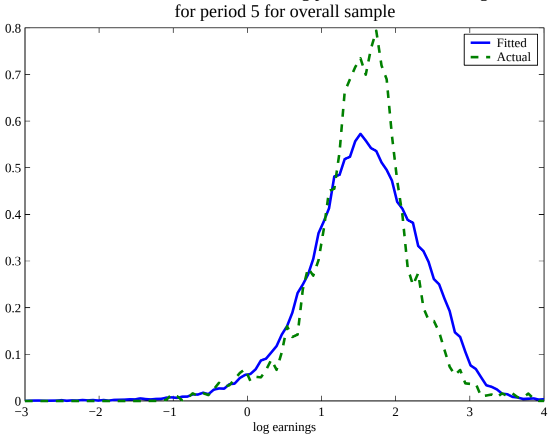
Log of present value of earnings from age 31 to 36 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(log(y)) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.4
Densities of fitted and actual log present value of earnings
for period 4 for overall sample



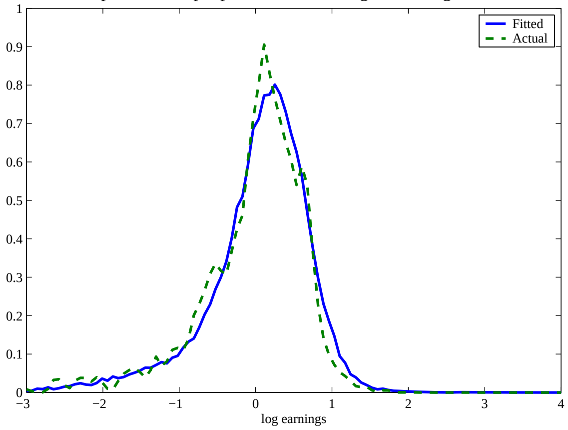
Log of present value of earnings from age 37 to 52 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(log(y)) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.5
Densities of fitted and actual log present value of earnings
for period 5 for overall sample



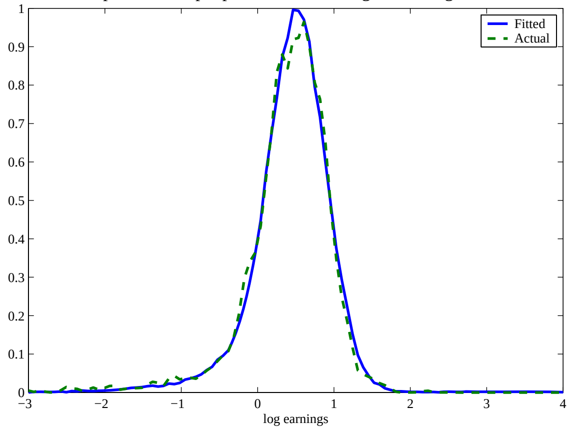
Log of present value of earnings from age 52 to 65 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(log(y)) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.1
Densities of fitted and actual log present value of earnings
for period 1 for people who choose to graduate high school



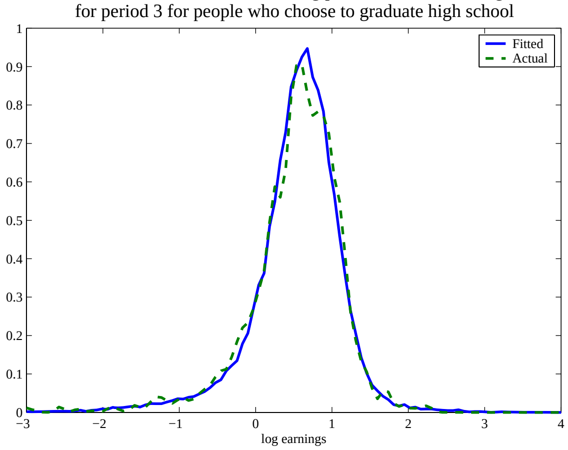
Log of present value of earnings from age 19 to 24 discounted using an interest rate of 3%. Let Y_0 denote. the present value of high school earnings for this period. Here we plot the density functions $f(\log(y_0)|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.2
Densities of fitted and actual log present value of earnings
for period 2 for people who choose to graduate high school



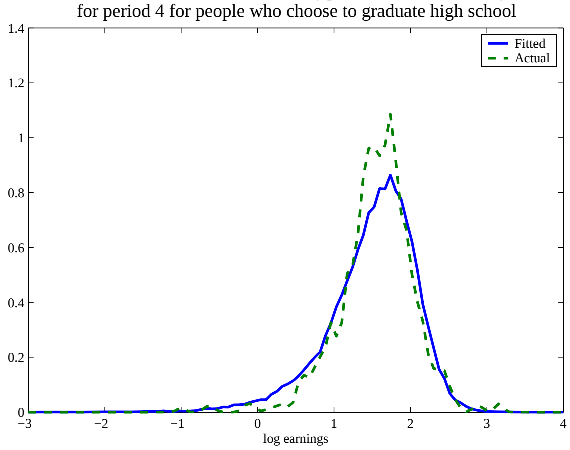
Log of present value of earnings from age 25 to 30 discounted using an interest rate of 3%. Let Y_0 denote. the present value of high school earnings for this period. Here we plot the density functions $f(\log(y_0)|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.3
Densities of fitted and actual log present value of earnings
for period 3 for people who choose to graduate high school



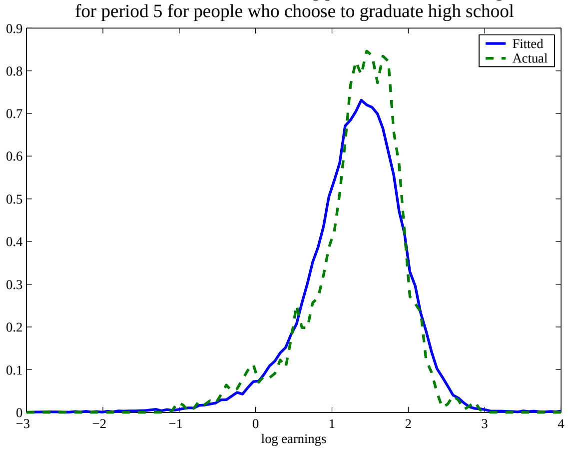
Log of present value of earnings from age 31 to 36 discounted using an interest rate of 3%. Let Y_0 denote. the present value of high school earnings for this period. Here we plot the density functions $f(\log(y_0)|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.4
Densities of fitted and actual log present value of earnings
for period 4 for people who choose to graduate high school



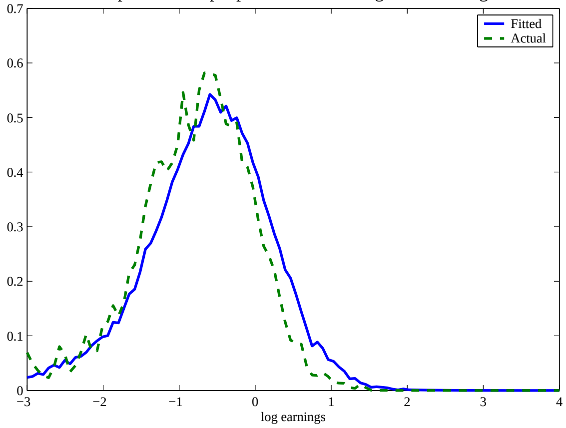
Log of present value of earnings from age 37 to 51 discounted using an interest rate of 3%. Let Y_0 denote. the present value of high school earnings for this period. Here we plot the density functions $f(\log(y_0)|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.5
Densities of fitted and actual log present value of earnings
for period 5 for people who choose to graduate high school



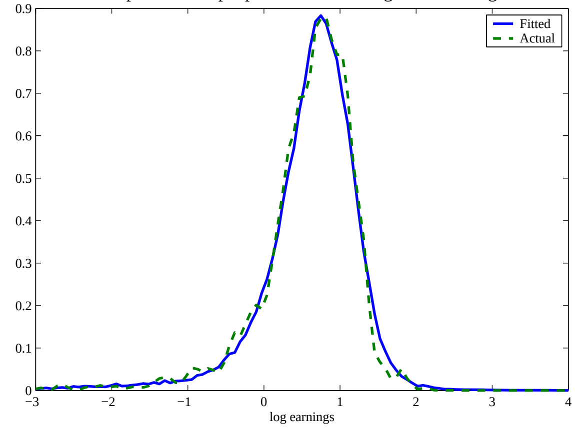
Log of present value of earnings from age 52 to 65 discounted using an interest rate of 3%. Let Y_0 denote. the present value of high school earnings for this period. Here we plot the density functions $f(\log(y_0)|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 4.1
Densities of fitted and actual log present value of earnings
for period 1 for people who choose to graduate college



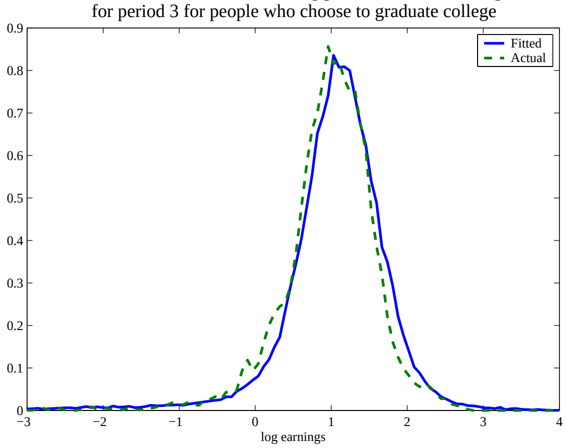
Log of present value of earnings from age 19 to 24 discounted using an interest rate of 3%. Let Y_1 denote. the present value of college earnings for this period. Here we plot the density functions $f(\log(y_1)|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 4.2
Densities of fitted and actual log present value of earnings
for period 2 for people who choose to graduate college



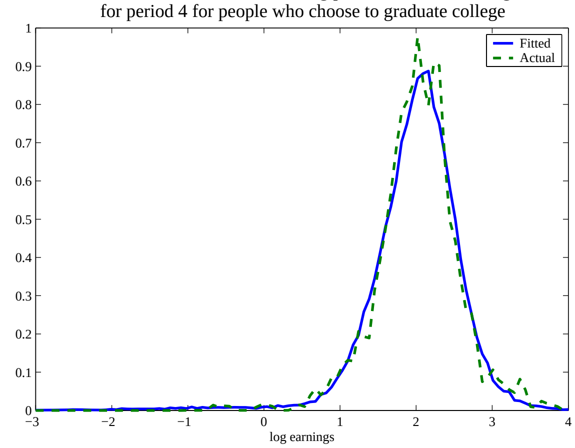
Log of present value of earnings from age 25 to 30 discounted using an interest rate of 3%. Let Y_1 denote. the present value of college earnings for this period. Here we plot the density functions $f(\log(y_1)|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 4.3
Densities of fitted and actual log present value of earnings
for period 3 for people who choose to graduate college



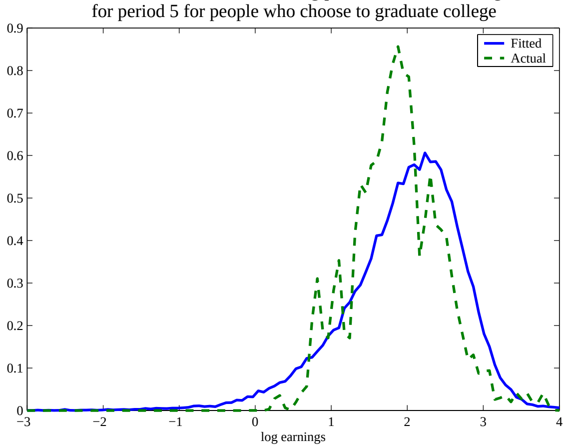
Log of present value of earnings from age 31 to 36 discounted using an interest rate of 3%. Let Y_1 denote. the present value of college earnings for this period. Here we plot the density functions $f(\log(y_1)|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 4.4
Densities of fitted and actual log present value of earnings
for period 4 for people who choose to graduate college



Log of present value of earnings from age 37 to 51 discounted using an interest rate of 3%. Let Y_1 denote. the present value of college earnings for this period. Here we plot the density functions $f(\log(y_1)|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 4.5
Densities of fitted and actual log present value of earnings
for period 5 for people who choose to graduate college



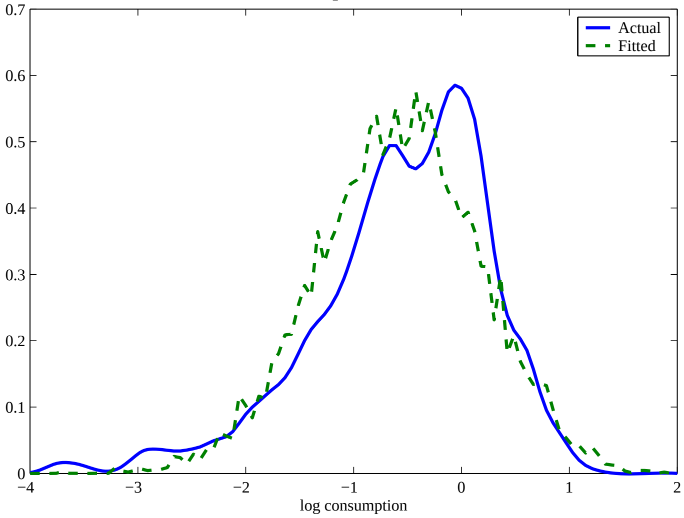
Log of present value of earnings from age 52 to 65 discounted using an interest rate of 3%. Let Y_1 denote. the present value of college earnings for this period. Here we plot the density functions $f(\log(y_1)|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Table 5
Goodness of Fit Tests of Equality Between Fitted and Actual Distributions of Log Present Value of Earnings

		High School	College	Overall
Age 19 – 24	χ^2 Statistic	134.3487	77.1154	181.9245
	Critical Value*	137.7015	85.9649	204.6902
Age 25 – 30	χ^2 Statistic	175.7894	133.9425	273.6058
	Critical Value*	186.1458	139.9208	308.2548
Age 31 – 36	χ^2 Statistic	151.8737	139.1731	288.8907
	Critical Value*	179.5806	157.6099	316.8185
Age 37 – 52	χ^2 Statistic	*92.2219*	55.8473	109.0211
	Critical Value*	67.5048	68.6693	110.8980
Age 53 – 65	χ^2 Statistic	25.4714	*40.2154*	*89.6596*
	Critical Value*	41.3371	27.5871	64.0011

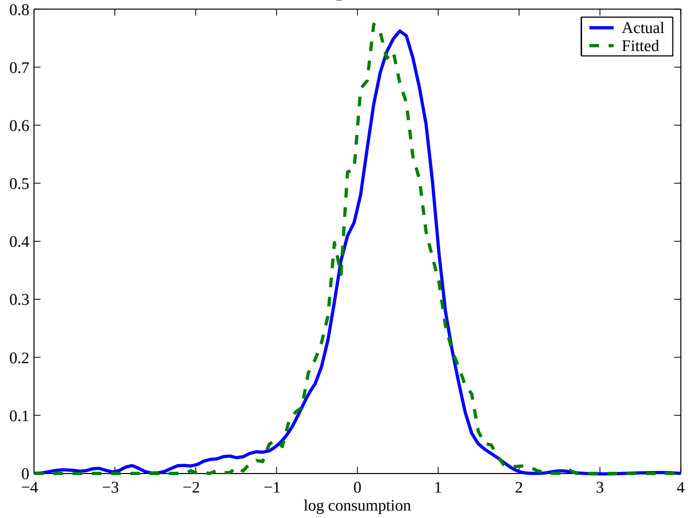
* 95% Confidence, equiprobable bins with aprox. 10 people per bin

Figure 5.1
Density of fitted and actual log consumption
for period 1



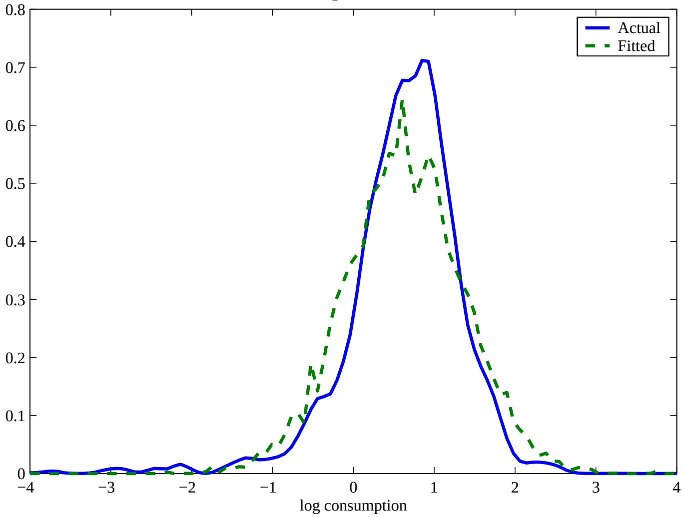
Log of the present value of consumption from age 19 to 24 discounted using an interest rate of 3%.
Density of fitted and actual log consumption
for period 1

Figure 5.2
Density of fitted and actual log consumption
for period 2



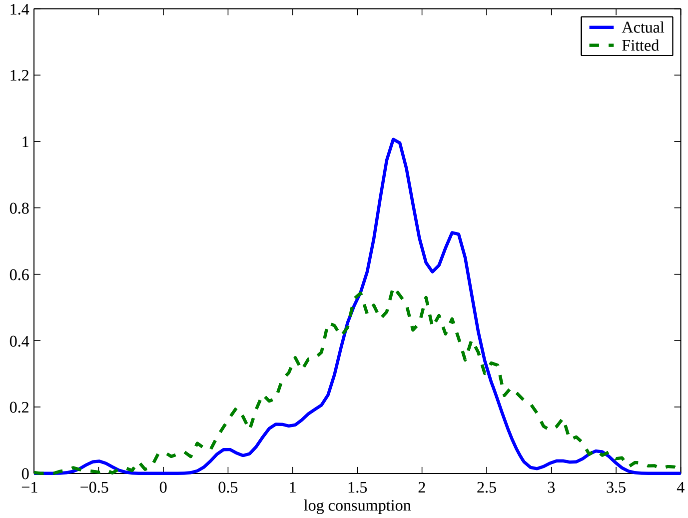
Log of the present value of consumption from age 25 to 30 discounted using an interest rate of 3%.
Density of fitted and actual log consumption
for period 2

Figure 5.3
Density of fitted and actual log consumption
for period 3



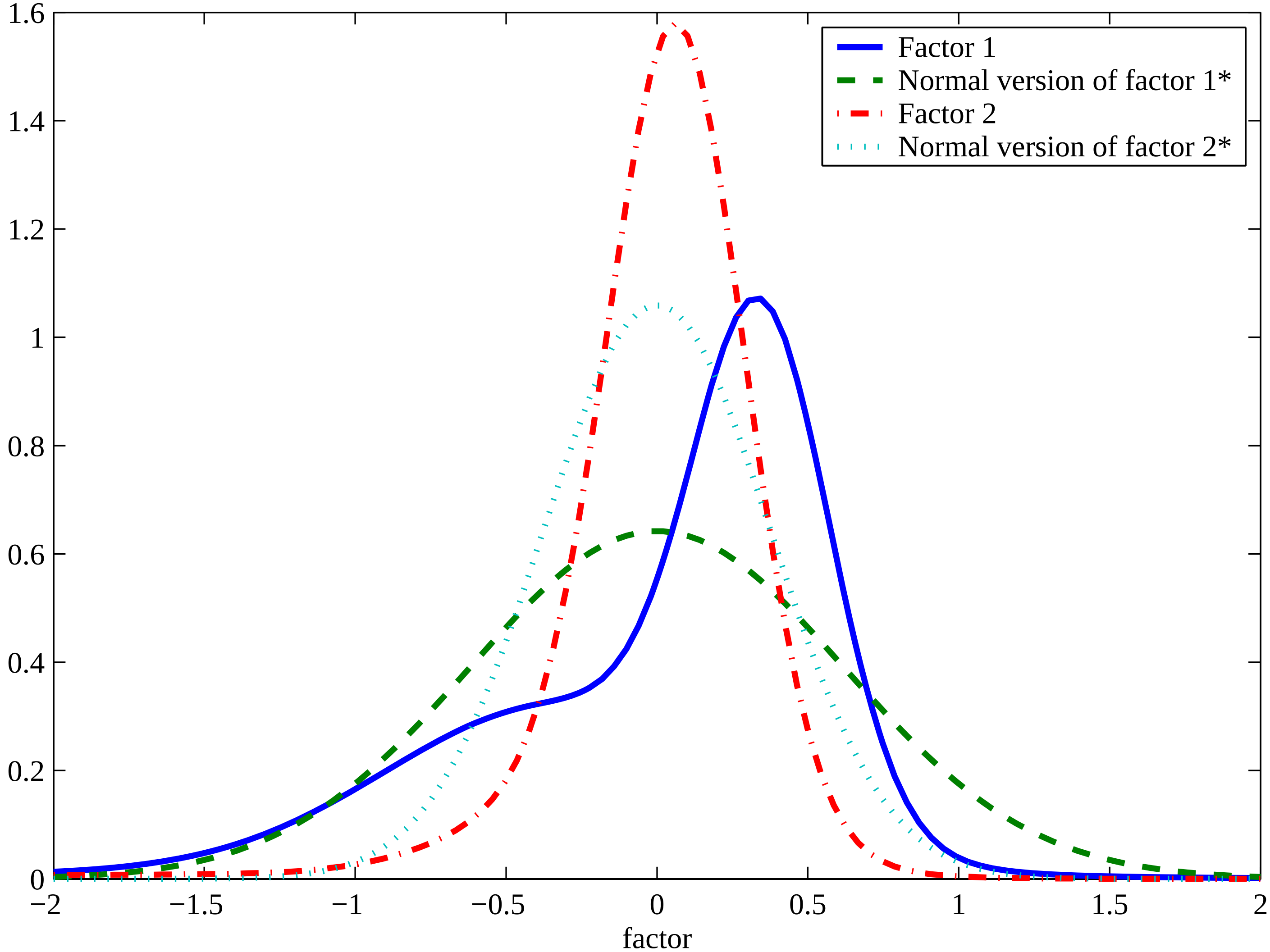
Log of the present value of consumption from age 31 to 36 discounted using an interest rate of 3%.
Density of fitted and actual log consumption
for period 3

Figure 5.4
Density of fitted and actual log consumption
for period 4



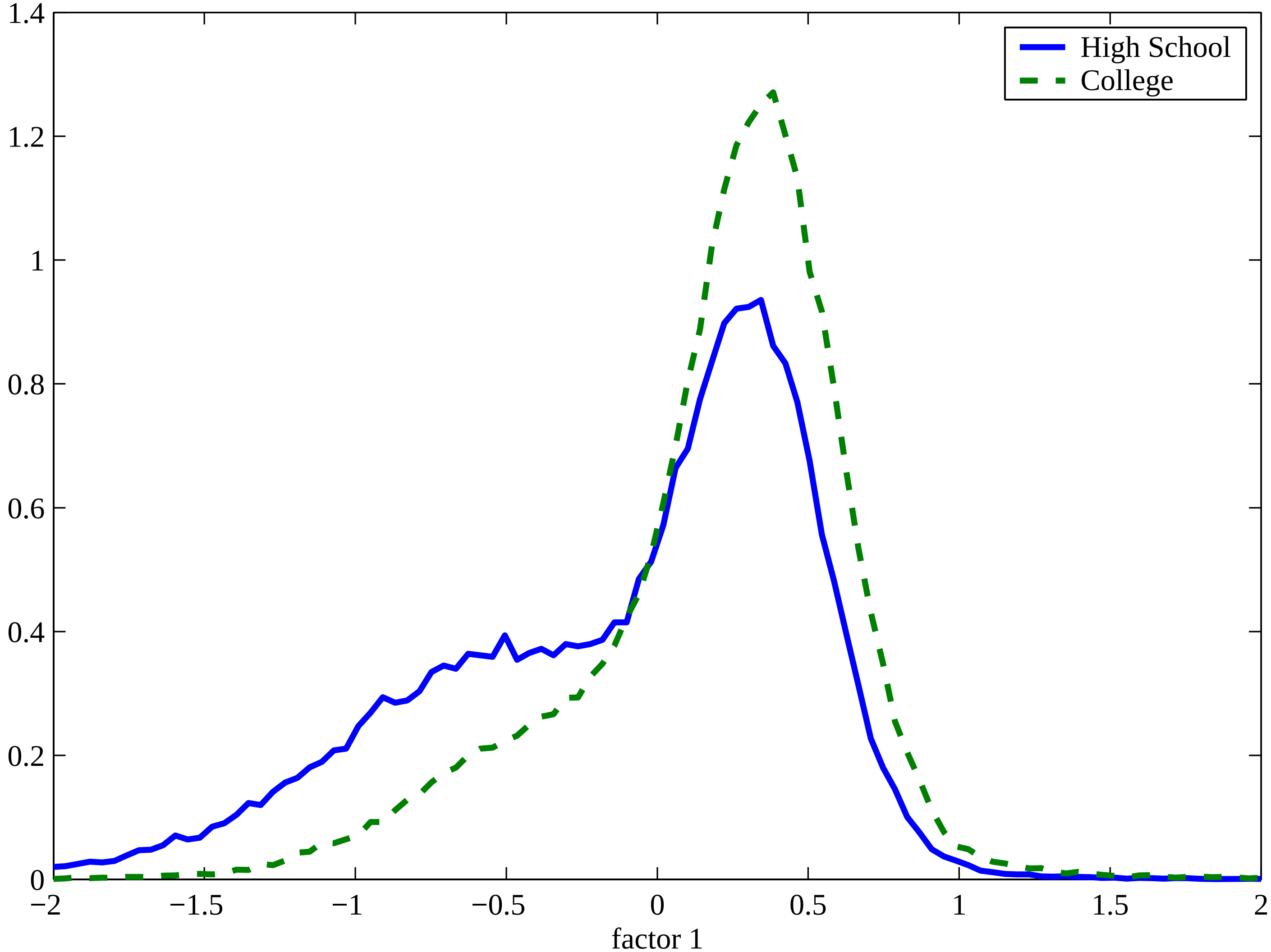
Log of the present value of consumption from age 37 to 51 discounted using an interest rate of 3%.
Density of fitted and actual log consumption
for period 4

Figure 6.1
Density of factors and their normal equivalents



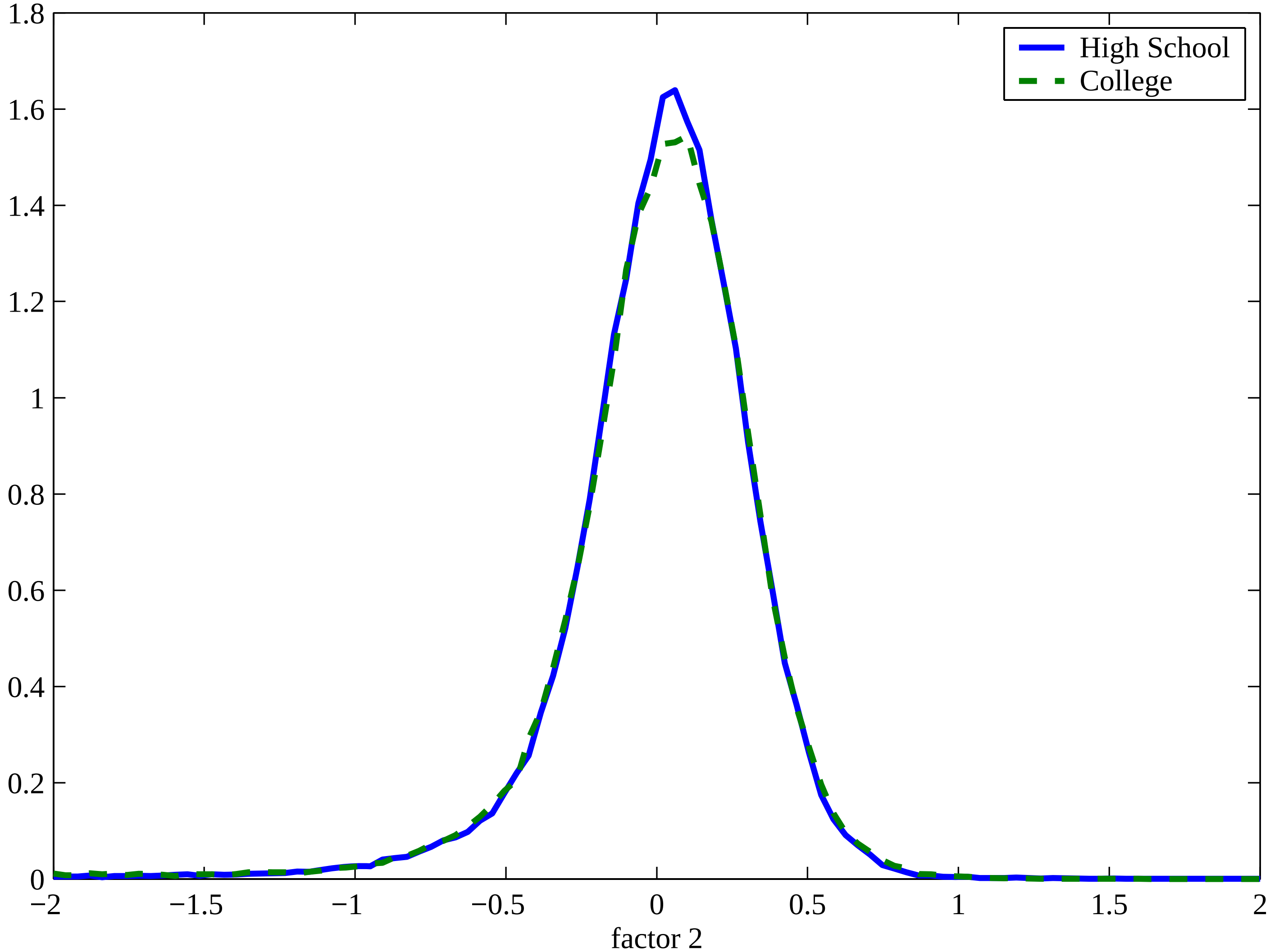
Let $f(\theta_1)$ denote the probability density function of factor θ_1. We allow $f(\theta_1)$ to be a mixture of normals. Assume $\mu_1=E(\theta_1)$, $\sigma_1=Var(\theta_1)$. Let $f(\mu_1,\sigma_1)$ denote the density of a normal random variable with mean μ_1 and variance σ_1. The solid curve is the 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 factor 2.

Figure 6.2
Density of Ability (Factor 1)
conditional on choice



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 6.3
Density of the Factor 2
conditional on choice



Let $f(\theta_2)$ denote the probability density function of factor θ_2. We allow that $f(\theta_2)$ 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, $f(\theta_2|\text{choice=college})$.

Table 6.1

Returns to college by decile of the ability distribution

Ability Decile	Choose high school		Choose college	
	Mean Return[1]	Proportion[2]	Mean Return[1]	Proportion[2]
1	-0.006	0.153	0.072	0.033
2	0.112	0.126	0.134	0.067
3	0.192	0.111	0.207	0.086
4	0.271	0.101	0.279	0.098
5	0.329	0.095	0.335	0.106
6	0.372	0.091	0.375	0.112
7	0.413	0.087	0.414	0.116
8	0.451	0.084	0.455	0.120
9	0.507	0.080	0.503	0.126
10	0.650	0.072	0.656	0.135

[1]Let Y_0 denote lifetime earnings in high school and Y_1 denote lifetime earnings in college. Then the return to college is $(Y_1-Y_0)/Y_0$.

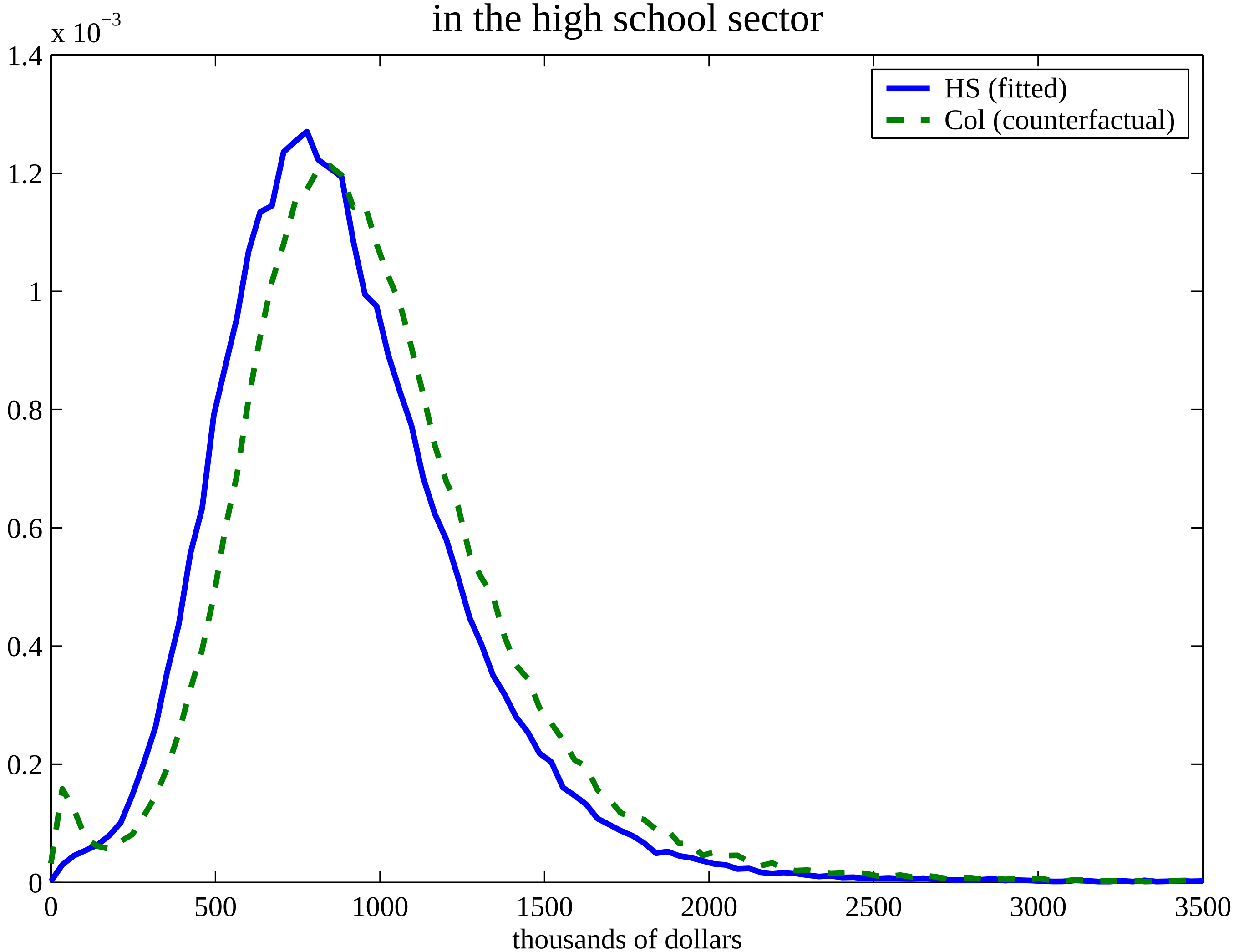
[2]Proportion of people who choose the schooling level indicated above and come from the ability decile to the left out of those who make the indicated choice. For example, 15.3% of those individuals who choose high school come from the first decile of the ability distribution.

Table 6.2
Distribution of college earnings conditional on high school earnings
$Pr(d_i<Y_c<d_{i+1} \mid d_j<Y_h<d_{j+1})$ where d_i is the ith decile of the college lifetime earnings distribution and d_j is the jth decile of the high school lifetime earnings distribution

Corrrelation(Y_c,Y_h) = 0.659171

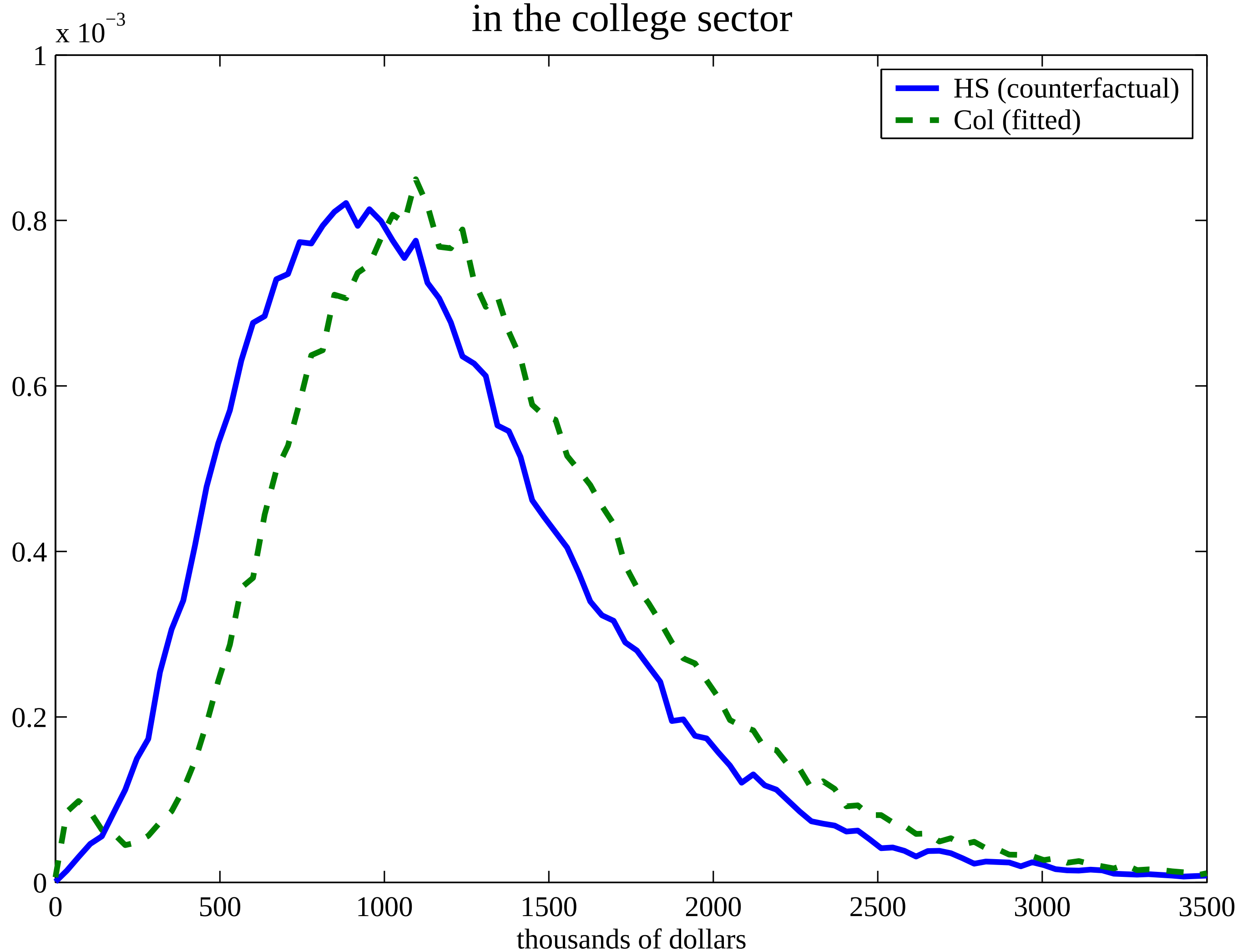
	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.68	0.20	0.07	0.03	0.01	0.01	0.00	0.00	0.00	0.00
2	0.22	0.33	0.20	0.12	0.06	0.04	0.02	0.01	0.00	0.00
3	0.06	0.23	0.24	0.19	0.13	0.07	0.04	0.02	0.01	0.00
4	0.02	0.13	0.20	0.21	0.16	0.12	0.08	0.04	0.02	0.01
5	0.01	0.06	0.14	0.18	0.19	0.17	0.12	0.07	0.04	0.02
6	0.00	0.03	0.09	0.14	0.18	0.19	0.16	0.13	0.07	0.03
7	0.00	0.01	0.04	0.08	0.14	0.18	0.19	0.17	0.13	0.06
8	0.00	0.00	0.02	0.04	0.07	0.13	0.19	0.22	0.20	0.12
9	0.00	0.00	0.00	0.02	0.04	0.08	0.14	0.21	0.29	0.22
10	0.00	0.01	0.01	0.01	0.01	0.03	0.05	0.11	0.24	0.54

Figure 7.1
Density of present value of earnings from age 19 to 65 in the high school sector



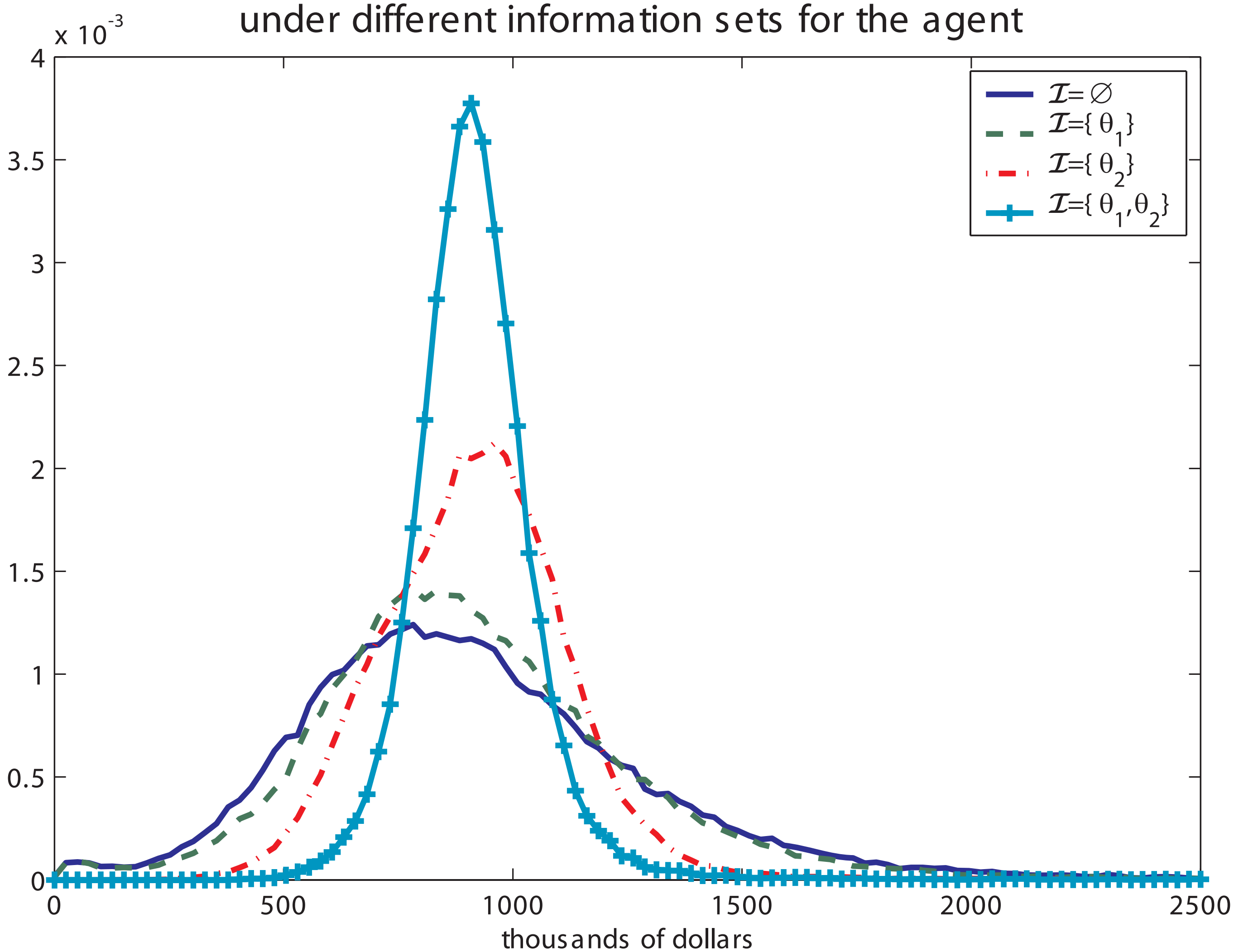
Let Y_0 denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(y_0)$ denote its density function. The solid curve plots the fitted Y_0 density for those agents who choose high school, that is, $f(y_0|\text{choice=high school})$, while the dashed line shows the counterfactual density function of Y_0 for those agents who are actually college graduates, that is, $f(y_0|\text{choice=college})$.

Figure 7.2
Density of present value of earnings from age 19 to 65
in the college sector



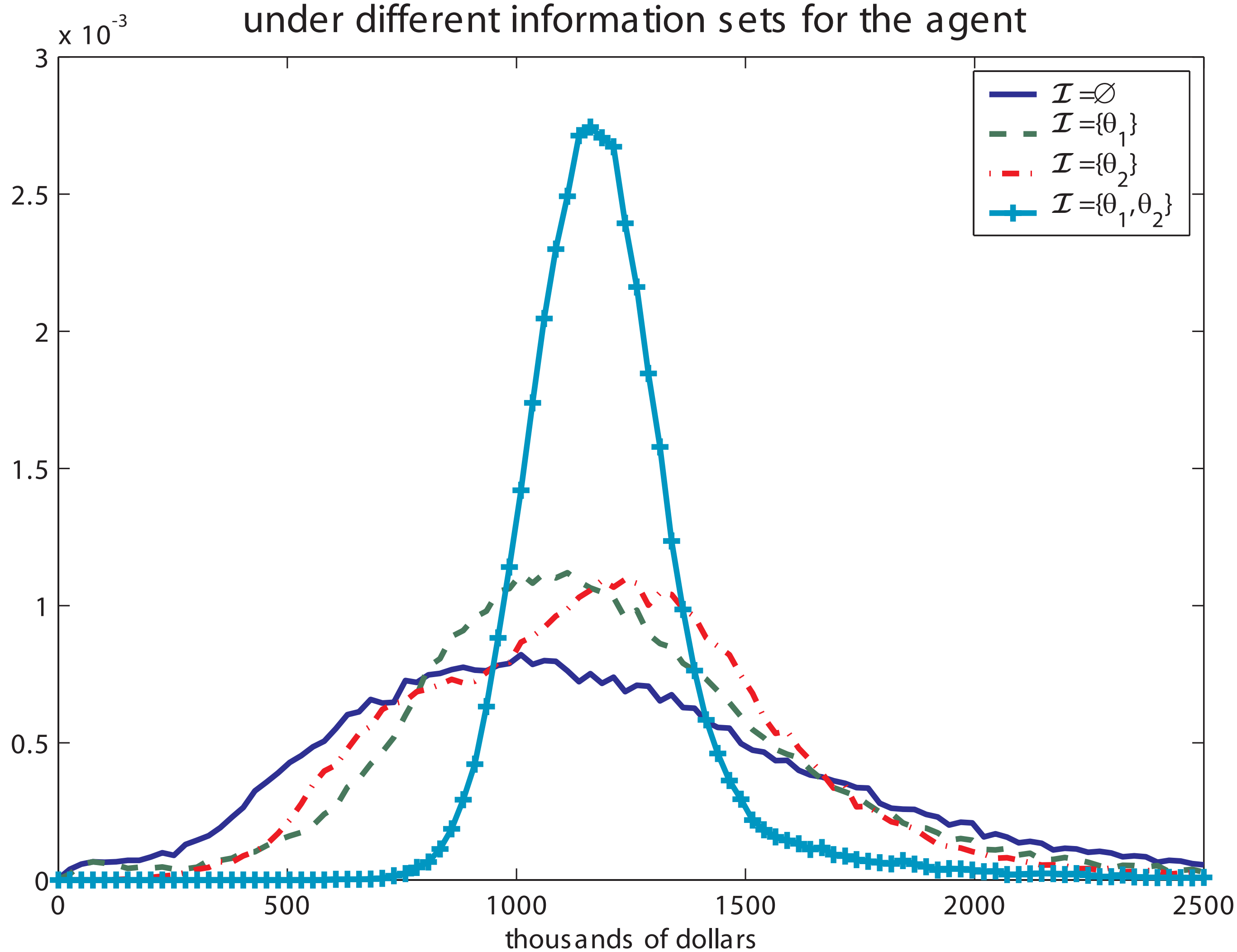
Let Y_1 denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(y_1)$ denote its density function. The solid curve plots the counterfactual Y_1 density for those agents who choose high school, that is, $f(y_1|\text{choice=high school})$, while the dashed line shows the fitted density function of Y_1 for those agents who are actually college graduates, that is, $f(y_1|\text{choice=college})$.

Figure 8.1
Density of present value of high school earnings
under different information sets for the agent



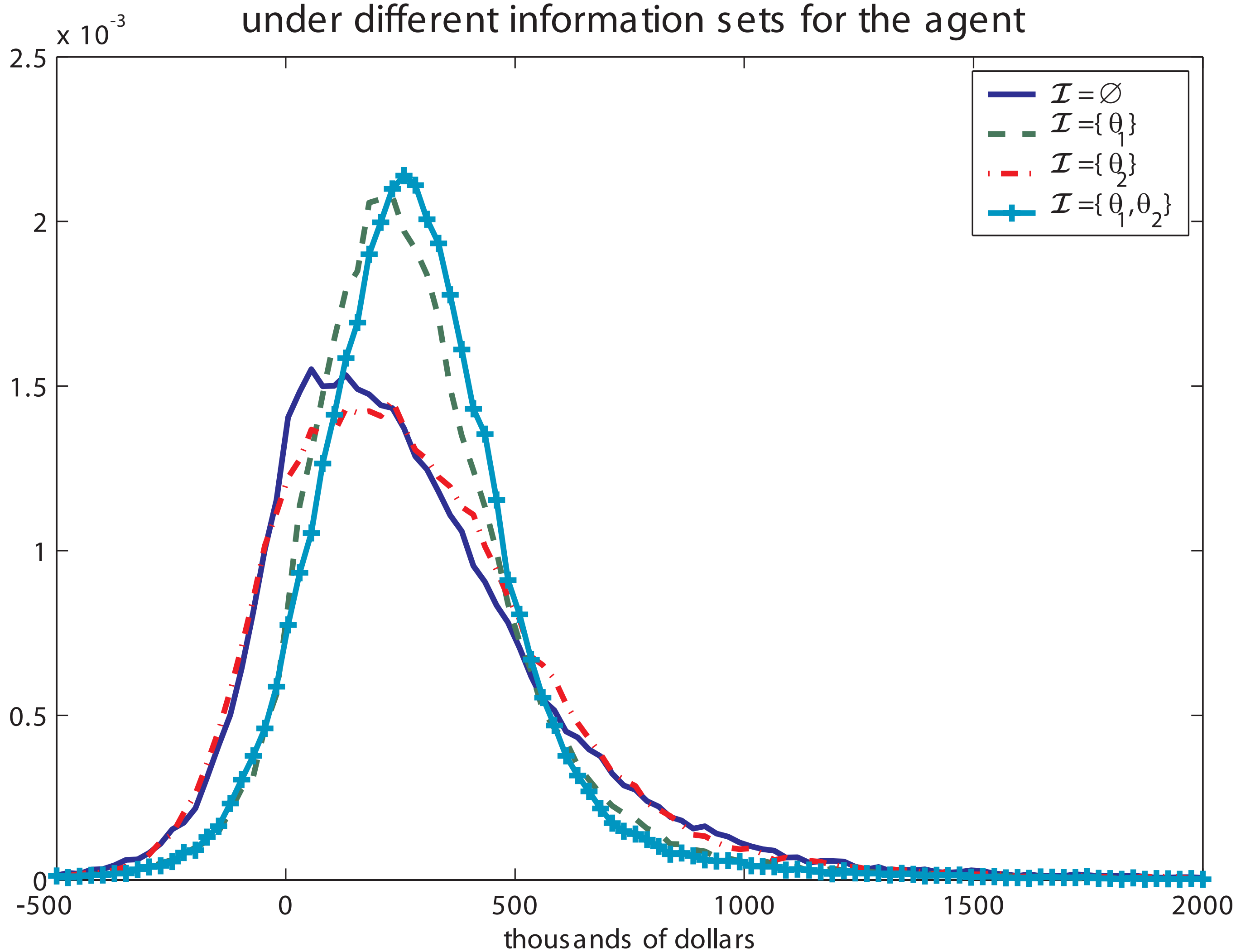
Let Y_0 denote the present value of earnings in the high school sector discounted at a 3% interest rate
Let I denote the agent's information set and $f(Y_0|\mathcal{I})$ denote the density of the present value of earnings in high school conditioned on the information set I. We plot $f(Y_0|\mathcal{I})$ under no information, with each factor in the information set, and with both factors in the information set. We use kernel density estimation to smooth these functions.

Figure 8.2
Density of present value of college earnings
under different information sets for the agent



Let Y_1 denote the present value of earnings in the college sector discounted at a 3% interest rate
Let $\mathcal{I}$ denote the agent's information set and $f(Y_1|\mathcal{I})$ denote the density of the present value of earnings in college conditioned on the information set I. We plot $f(Y_1|\mathcal{I})$ under no information, with each factor in the information set, and with both factors in the information set. We use kernel density estimation to smooth these functions.

Figure 8.3
Density of the difference between the present value of college and high school earnings under different information sets for the agent



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 $D=Y_1-Y_0$. Let $\mathcal{I}$ denote the agent's information set and $f(d|\mathcal{I})$ denote the density of the difference in present value of earnings in the college and high school sectors conditioned on the information set $\mathcal{I}$. We plot $f(Y_0|\mathcal{I})$ under no information, with each factor in the information set, and with both factors in the information set. We use kernel density estimation to smooth these functions.

Table 7.2

Agent's Forecast[1] of the Variance of the Present Value of Earnings[2]

Under Different Information Sets at Schooling Choice Date

		$Var(Y_h)$	$Var(Y_c)$	$Var(Y_c - Y_h)$
	Variance with $\mathcal{I} = \emptyset$	213657.03	360423.15	186705.84
$\mathcal{I}_1 = \theta_1$	Variance	186568.88	215538.98	131538.31
	Fraction of the variance[3] with $\mathcal{I} = \emptyset$ explained by $\mathcal{I}_1$	12.68%	40.20%	29.55%
$\mathcal{I}_2 = \theta_2$	Variance	90262.93	185409.75	165322.26
	Fraction of the variance with $\mathcal{I} = \emptyset$ explained by $\mathcal{I}_2$	57.75%	48.56%	11.45%
$\mathcal{I}_3 = \{\theta_1, \theta_2\}$	Variance	64535.63	49739.06	114352.54
	Fraction of the variance with $\mathcal{I} = \emptyset$ explained by $\mathcal{I}_3$	69.79%	86.20%	38.75%

[1]Variance of the unpredictable component of earnings between age 19 and 65 as predicted at age 19

[2]We use an interest rate of 3% to calculate the present value of earnings.

[3]So the variance of the unpredictable component of high school earnings with $\mathcal{I}_1 = \theta_1$ is $(1 - 0.1268) * 213657.03 = 186568.88$

Table 7.1

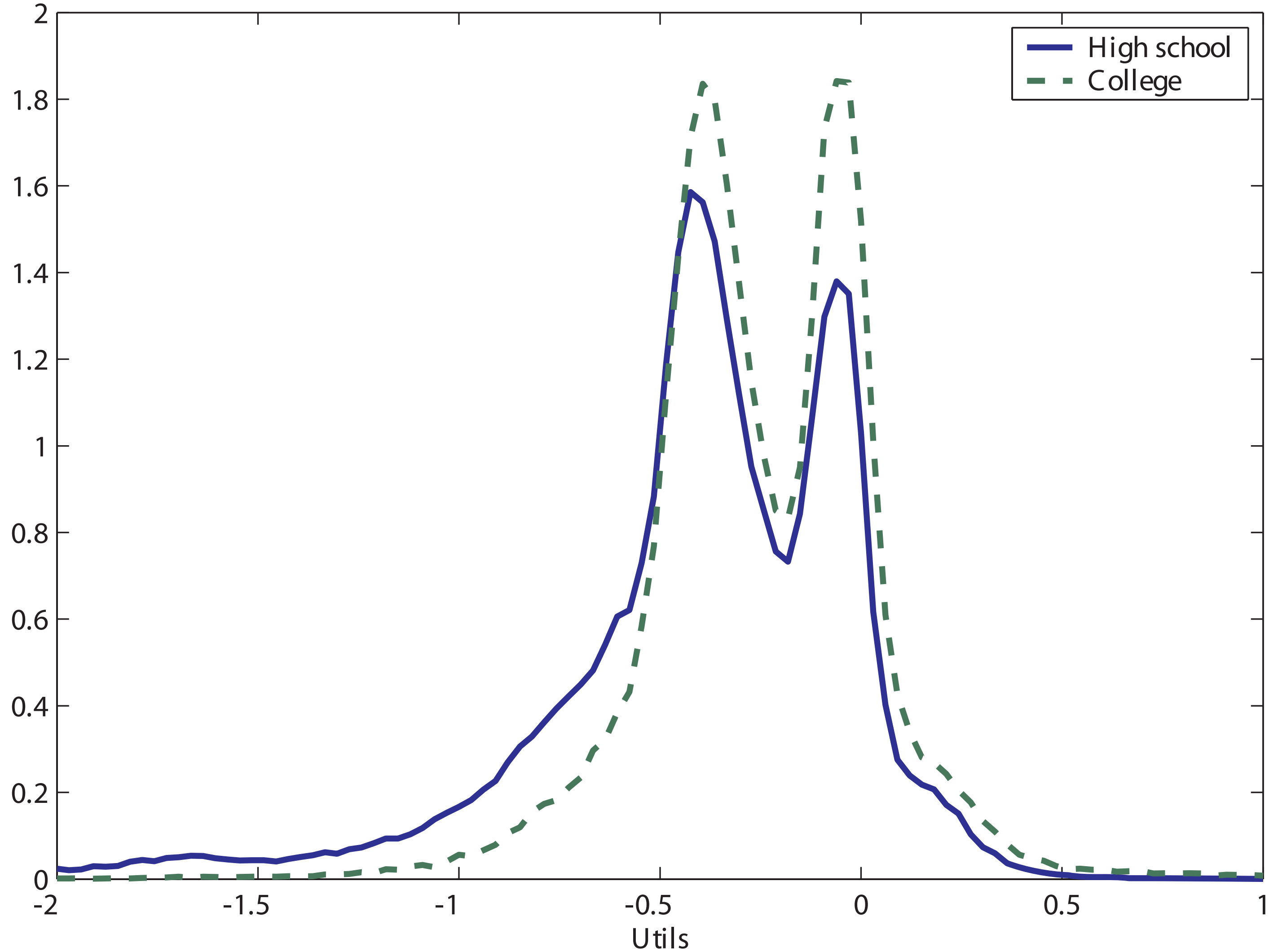
Proportion of people who, after observing their realized outcomes, regret their choice

	Choice under certainty	
Choice under uncertainty	**High School:** Average Annual Return: 5.65%	**College** Average Annual Return: 11.46%
High School: Average Annual Return 7.22%	81.02%	18.98%
College: Average Annual Return 9.61%	22.84%	77.16%

* For example, out of those that select high school under uncertainty 81% would still choose high school if they were to choose based on their realized earnings. The average return for people who choose high school under uncertainty is 7.22%. It would be 5.65% if people were to choose after their outcomes are realized.

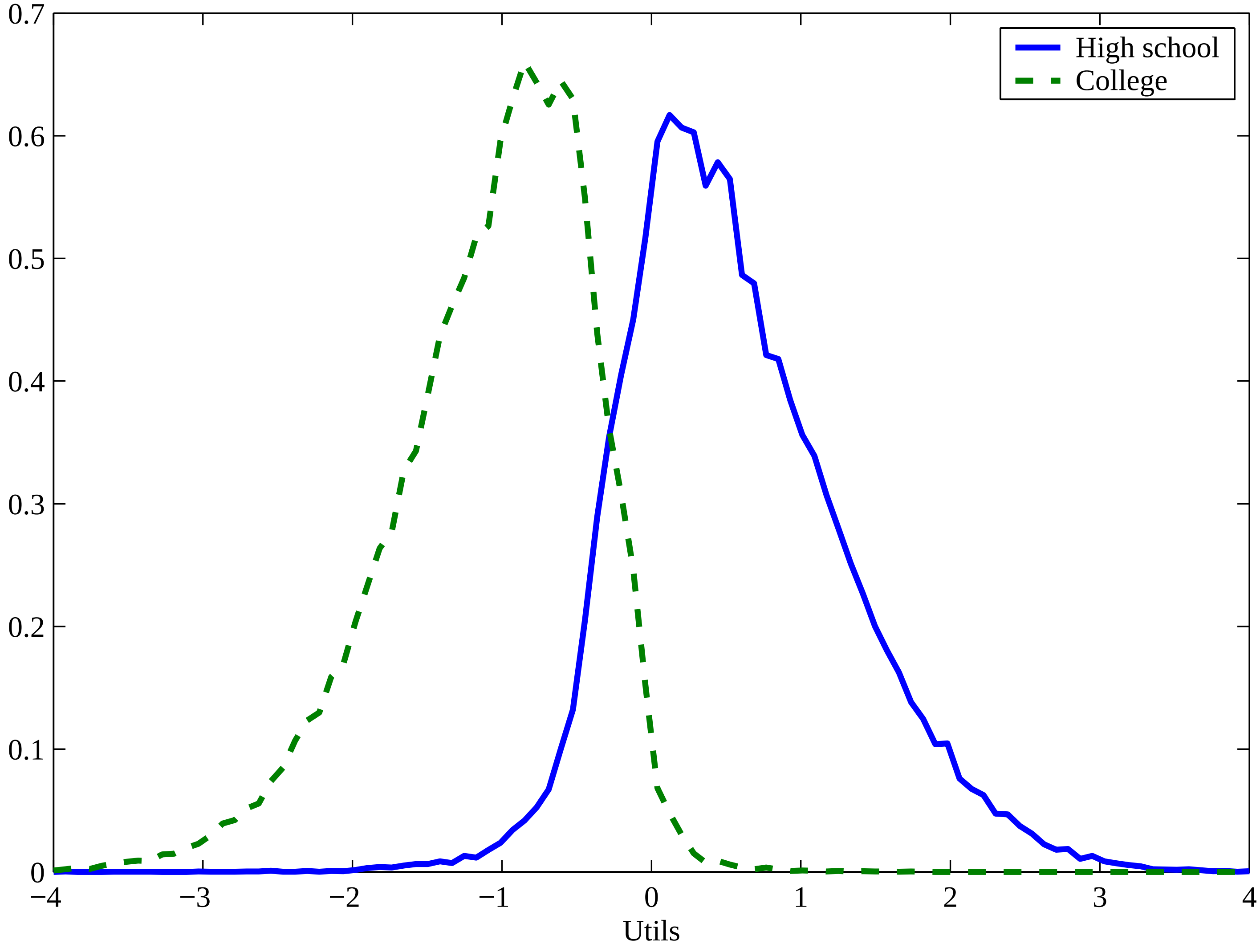
Figure 9.1

Density of expected gross utility differences conditional on choice



Let $(V_{h,1}, V_{c,1})$ denote the value functions for the high school and college at period 1.
Define the ex-ante gross utility difference, $D=E(V_{c,1}-V_{h,1}|\mathcal{I}_0)$ where the expectation
is taken with respect to the information available at period 0. The solid line shows the density of D for agents who choose high school (i.e., f(d|choice=high school)), and the dashed line shows the density of D for agents who choose college (i.e., f(d|choice=college))

Figure 9.2
Density of psychic costs
conditional on choice



Let C denote psychic costs. Let f(c) denote its density function. The solid line shows the density of psychic costs for high school graduates, that is f(c|choice=high school). The dashed line shows the density of psychic costs for college graduates, that is, f(c|choice=college).

Table 8
Percentage of people who choose college when tuition is set to zero

Subsidy	*Overall*
None	44.15%
Zero Tuition Economy	49.50%

Table A-1
Number of observations[1] by variable
White Males from NLSY79

Variable	Number of observations
ASVAB test scores	2293
Enrolled in school at test date	2438
Age at test date	2439
Highest grade completed at test date	2386
Mother's highest grade completed[2]	2439
Father's highest grade completed[2]	2439
Broken home dummy	2434
Number of siblings	2439
Lived in an urban environment at age 14	2433
In state tuition for 4 year college at age 17	2438
Education	2397
Year of birth	2439

[1] We begin with a sample of 2439 white males. We drop any individuals who are missing any of the variables listed above. This results in a sample with 2196 people. We also drop anyone for whom we cannot impute earnings when they are missing from the survey. This produces a sample of 2167 individuals.

[2] Includes imputed values.

Table A-2

Evolution of attrition over time

White males from PSID

Age	19	25	31	37	52
19	842	722	527	202	0
25	0	2254	1721	839	41
31	0	0	2267	1143	173
37	0	0	0	1661	538
52	0	0	0	0	877

*For example, out of 842 individuals whose earnings are observed at 19, 722 are also observed at age 25, 527 at age 31, and so on.

Table A-3

Evolution of attrition over time

White male high school and college graduates from PSID

Age	19	25	31	37	52
19	483	420	325	122	0
25	0	1500	1185	570	24
31	0	0	1542	761	96
37	0	0	0	1076	309
52	0	0	0	0	508

*For example, out of 483 individuals whose earnings are observed at 19, 420 are also observed at age 25, 325 at age 31, and so on.

Table A-4

Earnings with and without imputed values[1]
For NLSY from age 19 to 43

		Observations	*Mean*	*StdDev*
High School and College	Original	20431	32061.32	29858.5
	Imputed	26646	30927.7	30536.12
High School	Original	11103	28298.47	21258.04
	Imputed	14601	26911.42	22043.71
College	Original	9328	36540.19	37121.4
	Imputed	12045	35796.25	37822.81

For PSID from age 19 to 65

		Observations	*Mean*	*StdDev*
High School and College	Original	40488	48024.54	36116.65
	Imputed	50929	47041.02	40626.4
High School	Original	21434	37621.36	21932.5
	Imputed	27307	36640.38	26673.98
College	Original	19054	59727.15	44407.22
	Imputed	23622	59064.15	49663.37

[1]To impute earnings, we take a weighted mean of the observed earnings for each individual with weights given by the Epanechnikov kernel.

Table A-5

Earnings and consumption for each period[1]

		Overall			High School			College		
Period[2]	*Variable*[3]	*Obs*	*Mean*	*StdDev*	*Obs*	*Mean*	*StdDev*	*Obs*	*Mean*	*StdDev*
1	Earnings	1848	0.93	0.64	1153	1.15	0.64	695	0.58	0.45
	Consumption	1139	0.75	0.51	659	0.92	0.54	480	0.51	0.36
2	Earnings	2865	1.84	0.92	1598	1.67	0.78	1267	2.06	1.02
	Consumption	898	1.63	0.96	482	1.50	0.80	416	1.78	1.10
3	Earnings	2908	2.50	1.57	1551	1.96	1.07	1357	3.11	1.81
	Consumption	931	2.27	1.49	507	1.88	1.21	424	2.73	1.65
4	Earnings	1076	7.04	4.47	523	5.27	2.60	553	8.71	5.17
	Consumption	107	7.23	4.40	28	4.66	1.92	79	8.14	4.67
5	Earnings	509	5.58	4.25	308	4.29	2.28	201	7.56	5.60
	Consumption	123	7.22	4.37	62	5.65	3.28	61	8.82	4.77

[1] Hundreds of thousands of dollars.

[2] Periods are defined as ages: 19-24, 25-30, 31-36, 37-51 and 52-65.

[3] Earnings are defined as the present value of earnings for the ages included in the period discounted using an interest rate of 3%. Consumption is defined as the difference between available resources at the beginning of the period (earnings plus assets) and discounted assets the next period.

Table A-6

Comparison of variables from PSID and NLSY

People born between 1957 and 1964

Variable	PSID Mean	PSID Std. Dev.	NLSY Mean	NLSY Std. Dev.
Tuition for 4 year college	2.00	0.53	2.16	0.82
Number of Siblings	3.11	1.98	2.82	1.82
PV Earnings from 19 to 24	1.40	0.70	0.77	0.55
PV Earnings from 25 to 30	1.89	0.83	1.70	0.97
PV Earnings from 31 to 36	2.63	1.65	2.37	1.67
Consumption from 19 to 24	1.23	0.54	0.69	0.49
Consumption from 25 to 30	1.80	0.83	1.57	0.91
Consumption from 31 to 36	2.51	1.46	2.15	1.47
Assets at age 25	0.15	0.54	0.13	0.43
Assets at age 31	0.43	0.92	0.53	1.08
Assets at age 37	0.71	0.95	0.98	1.52
Categorical Variable	*Proportion*		*Proportion*	
Mother Dropped out of High School	0.175		0.195	
Mother High School Graduate	0.538		0.525	
Mother Attended Some College	0.139		0.144	
Mother College Graduate	0.149		0.136	
Father Dropped out of High School	0.256		0.251	
Father High School Graduate	0.371		0.353	
Father Attended Some College	0.124		0.147	
Father College Graduate	0.249		0.249	
College Graduate	0.401		0.453	
Urban at age 14	0.873		0.742	
Divorced Parents	0.150		0.155	