Schooling and Returns to Schooling
Part 4A. Decomposing Trends in Inequality in Earnings into Forecastable and Uncertain Components

From Cunha & Heckman (2015), Econ 350, Winter 2015

James J. Heckman
University of Chicago

AEA Continuing Education Program
ASSA Course: Microeconomics of Life Course Inequality
San Francisco, CA, January 5-7, 2016

Introduction

- Variability in wages across people over time is not the same as uncertainty in wages.

Approach is based on the following simple idea:

- A decision variable C_1, say consumption in the first period, may depend on outcomes $Y_1, \ldots, Y_T$ over horizon T of another variable (e.g., income) that are realized after the consumption choice is taken.
- Abstracting from measurement errors, under the permanent income hypothesis the correlation between C_1 and future Y_t is a measure of how much of future Y_t is known *and acted on* when agents make their decisions (Flavin, 1981).
- When making their consumption decisions, agents only imperfectly predict their future earnings using information $\mathcal{I}_1$.

- Suppose that
  - C_1 depends on future Y_t only through expected present value $E\left(PV|\mathcal{I}_1\right)$
  - $PV = \sum_{t=1}^{T} \frac{Y_t}{(1+\rho)^{t-1}}$
  - ρ is the discount rate.
- If, after the choice of C_1 is made, we observe $Y_1, \ldots, Y_T$, we can construct PV *ex-post*.
- If information set properly specified, the residual corresponding to the component of PV that is not forecastable in the first period, $V = PV - E\left(PV|\mathcal{I}_1\right)$, should not predict C_1.
- $E\left(PV|\mathcal{I}_1\right)$ is predictable.
- V arises from uncertainty.
- The variance in PV that is unpredictable using $\mathcal{I}_1$ is a measure of uncertainty as of the first period.

- Use college attendance choices as its decision variable to estimate uncertainty.
- Not essential — Can use consumption, labor supply, educational choices or all together.

The Model

Earnings Equations

- Roy model (1951).
- Agents possess two *ex-post* earnings streams, $(Y_{0,t}, Y_{1,t})$, $t = 1, \ldots, T$, for schooling levels "0" and "1" respectively.
- They are assumed to have finite means.
- For conditioning variables X, we write:

$$Y_{0,t} = X\beta_{0,t} + U_{0,t} \tag{1}$$

$$Y_{1,t} = X\beta_{1,t} + U_{1,t}, \qquad t = 1, \ldots, T. \tag{2}$$

- $U_{s,t}$ are defined to satisfy $E\left(U_{s,t} \mid X\right) = 0$, $s = 0, 1, t = 1, \cdots, T$.
- For any individual, only observe one of the two possible earnings streams.

Choice Equations

- **?**

$$I = E\left[\sum_{t=1}^{T}\left(\frac{1}{1+\rho}\right)^{t-1}(Y_{1,t} - Y_{0,t}) - C \middle| \mathcal{I}_1\right] \quad (3)$$

- C is the cost of attending college
- Costs include both pecuniary and psychic costs
- Let Z and U_C denote, respectively, measured and unmeasured determinants of costs respectively

$$C = Z\gamma + U_C \quad (4)$$

$$\mu_I(X,Z) = \sum_{t=1}^{T} \left(\frac{1}{1+\rho}\right)^{t-1} X\left(\beta_{1,t} - \beta_{0,t}\right) - Z\gamma,$$

$$U_I = \sum_{t=1}^{T} \left(\frac{1}{1+\rho}\right)^{t-1} \left(U_{1,t} - U_{0,t}\right) - U_C \ ,$$

and substituting in (1), (2), and (4) into decision rule (3) we obtain

$$I = E\left[\mu_I(X,Z) + U_I \middle| \mathcal{I}_1\right]. \tag{5}$$

- $E(U_I|\mathcal{I}_1)$ is the error term in the choice equation and it may or may not include $U_{1,t}$, $U_{0,t}$, or U_C, depending on what is in the agent's information set.
- Similarly, $\mu_I(X, Z)$ may only be based on expectations of future X and Z at the time schooling decisions are made.
- People go to college if the expected present value of earnings is positive:

$$S = \mathbf{1}[I \geq 0]. \quad (6)$$

Cognitive Ability

- Usually captured by a test score.
- M_k denotes an agent's score on the k^{th} test.

$$M_k = X^M \beta_k^M + U_k^M \text{ and } E\left(U_k^M \mid X^M\right) = 0, k = 1, 2, \ldots, K. \quad (7)$$

Heterogeneity and Uncertainty

$$Y_{s,t} = E\left(Y_{s,t} \mid \mathcal{I}_1\right) + V_{s,t}, \qquad s = 0, 1, \; t = 1, \dots, T.$$

- The component $E\left(Y_{s,t} \mid \mathcal{I}_1\right)$ is available to the agent to help predict schooling choices.
- Varies across people because of information, including constraints, ability, etc.

Factor Models

- Develop our factor structure, starting with the earnings and choice equations.
- Let factors and factor loadings be $\theta = (\theta_1, \ldots, \theta_K)$ and $\alpha_{s,t} = (\alpha_{1,s,t}, \ldots, \alpha_{K,s,t})$.
- The idiosyncratic error terms, $\varepsilon_{s,t}, s \in \{0,1\}, t \in \{1, \cdots, T\}$, affect only the period-t, schooling-s earnings equation.

$$U_{s,t} = \theta\alpha_{s,t} + \varepsilon_{s,t} \quad s = 0, 1, \quad t = 1, \ldots, T. \tag{8}$$

- $\varepsilon_{s,t}$, $s = 0, 1$ and $t = 1, \ldots, T$, are mutually independent.

- Psychic and pecuniary cost is decomposed in a fashion similar to the earnings equations,

$$C = Z\gamma + \theta\alpha_C + \varepsilon_C. \tag{9}$$

$$I = E\left[\ \sum_{t=1}^{T}\left(\frac{1}{1+\rho}\right)^{t-1} X\left(\beta_{1,t} - \beta_{0,t}\right) - Z\gamma + \theta\alpha_I + \sum_{t=1}^{T}\left(\frac{1}{1+\rho}\right)^{t-1}\left(\varepsilon_{1,t} - \varepsilon_{0,t}\right) - \varepsilon_C\ \middle|\ \mathcal{I}_1\right] \tag{10}$$

where we define

$$\alpha_I = \sum_{t=1}^{T} \frac{1}{(1+\rho)^{t-1}}\left(\alpha_{1,t} - \alpha_{0,t}\right) - \alpha_C.$$

Test Score Equations

$$M_k = X^M \beta_k^M + \theta_1 \alpha_k^M + \varepsilon_k^M, k = 1, \ldots, K \quad (11)$$

The Estimation of Predictable Components of Future Earnings

- For expositional simplicity, *in this section alone* we assume that X, Z, $\beta_{s,t}$ ($s = 0, 1,\ t = 1, \ldots, T$) and ε_C are in the information set $\mathcal{I}_1$.
- Suppose that there are two factors, θ_1 (ability) and θ_2.
- Suppose that it is claimed that both θ_1 and θ_2 are known by the agent when schooling choices are made but the $\varepsilon_{s,t}$ are not, *i.e.* $\{\theta_1, \theta_2\} \subset \mathcal{I}_1$, but $\varepsilon_{s,t} \notin \mathcal{I}_1$ for all s and t.
- If this is true,

$$I = \mu_I(X, Z) + \alpha_{1,I}\theta_1 + \alpha_{2,I}\theta_2 + \varepsilon_C. \tag{12}$$

- Standard results in the theory of discrete choice (see Matzkin, 1992 or Heckman and Vytlacil, 2007 for precise conditions), can proceed as if we observe I in equations (6) and (12) up to an unknown positive scale.
- From the correlation between S and realized incomes, we can form (up to scale) the covariance between I and $Y_{s,t}$, $t = 1, \ldots, T$ for $s = 0$ or 1.
- Conditional on X, Z this covariance is

$$\mathrm{Cov}\,(I, Y_{s,t}|X, Z) = \alpha_{1,I}\alpha_{1,s,t}\sigma^2_{\theta_1} + \alpha_{2,I}\alpha_{2,s,t}\sigma^2_{\theta_2}, \qquad s = 0, 1, t = 1, \cdots, T. \quad (13)$$

- Suppose next that θ_2 is not known, or is known and not acted on by the agent when schooling choices are made.
- In this case, $\alpha_{2,I} = 0$.
- If neither θ_2 nor θ_1 is known, $\alpha_{1,I} = \alpha_{2,I} = 0$.

Empirical Results

- White males born between 1957 and 1964, National Longitudinal Survey of Youth (NLSY/1979).
- White males born between 1941 and 1952, National Longitudinal Survey (NLS/1966).
- Refer to the samples as NLSY/1979 and NLS/1966, respectively.

- Figures 1 and 2 display, respectively, the mean earnings by age of high school and college graduates for NLSY/1979 and NLS/1966.

Figure 1: Mean Earnings Profile NLSY/66, Comparison Across Schooling Within Cohorts

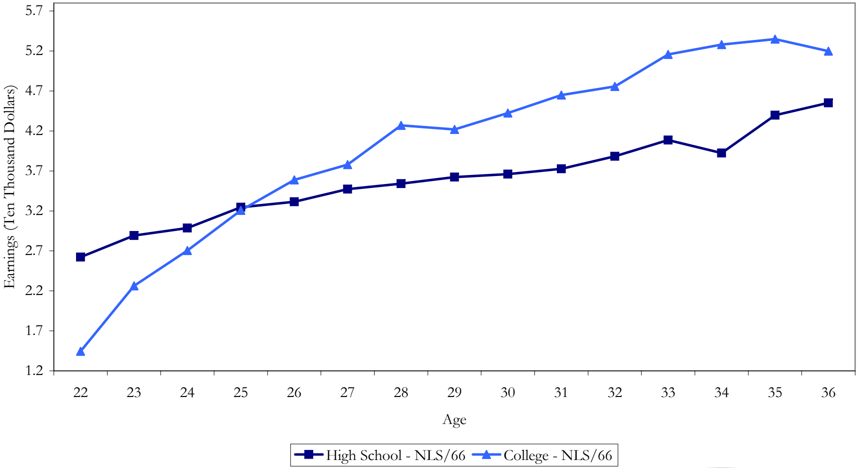
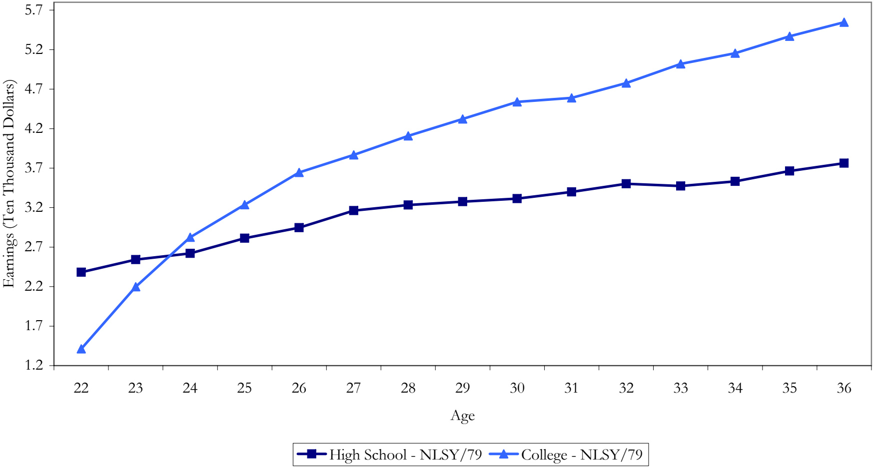


Figure 2: Mean Earnings Profile NLSY/79, Comparison Across Schooling Within Cohorts



- Plot the standard deviation of earnings by age for high school graduates (Figure 3) and college graduates (Figure 4) for both cohorts.
- Both data sets.

Figure 3: Standard Deviation of Earnings, High School Sample, Comparison Within Schooling Groups Across Cohorts

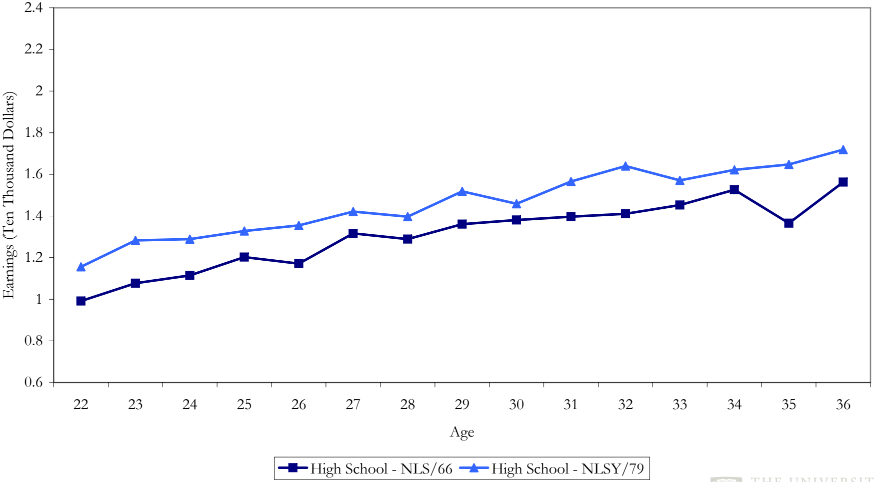


Figure 4: Standard Deviation of Earnings, College Sample, Comparison Within Schooling Groups Across Cohorts

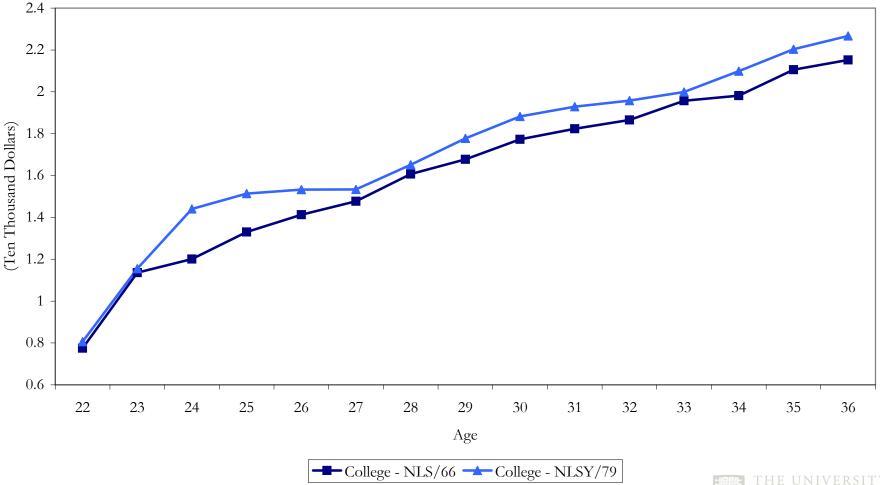
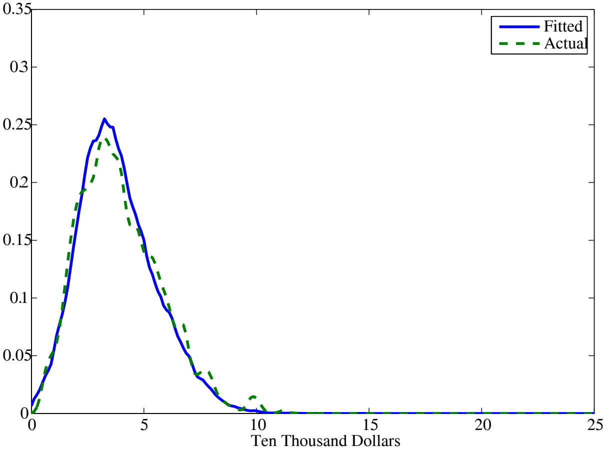


Table 1: Schooling Choice and Rates of Return per Year of College: Comparison Across Cohorts

	NLS/66	**NLSY/79**
High School Graduates	58.17%	64.19%
College Graduates	41.83%	35.81%
Mincer Returns to College[1]	9.01%	11.96%
Mincer Returns to College[2]	10.17%	12.41%
Mincer Returns to College[3]	8.17%	11.00%

[1]Pooled OLS Regression, controlling only for Mincer Experience and Mincer Experience Squared
[2]Pooled OLS Regression, controlling for Mincer Experience, Mincer Experience Squared, and Year Dummies
[3]Pooled OLS Regression, controlling for Mincer Experience, Mincer Experience Squared, Cognitive Skills, Urban and South Residence at Age 14, and Year Dummies (Dependent Variable: Log Earnings).

Figure 5: Densities of Earnings at Age 33, Overall Sample NLSY/79

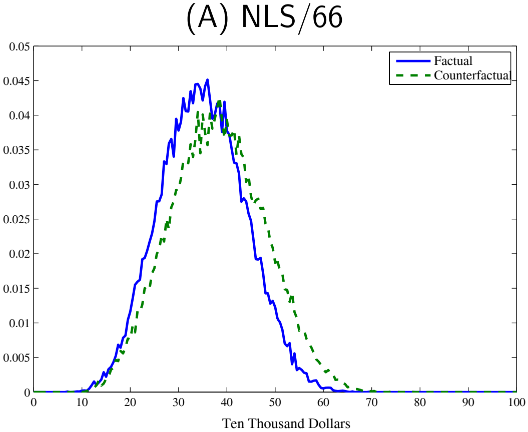


Let Y denote earnings at age 33 in the overall sample. Here we plot the density functions $f(y)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

The Evolution of Joint Distributions of Earnings and the Returns to College

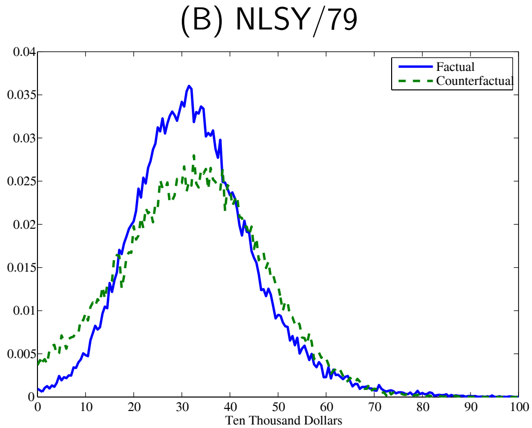
- Density of the present value of truncated *ex post* earnings

Figure 6: Densities of Present Value Earnings, High School Sample



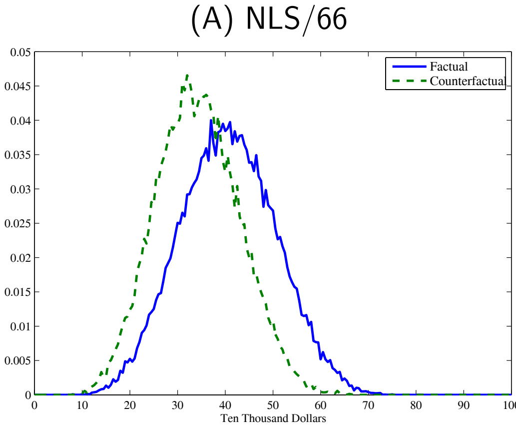
Present value of earnings from age 22 to 36 for High School Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_0|S=0)$ (the solid curve), against the counterfactual density function $f(y_1|S=0)$ (the dashed line). We use kernel density estimation to smooth these functions.

Figure 6: Densities of Present Value Earnings, High School Sample



Present value of earnings from age 22 to 36 for High School Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_0|S=0)$ (the solid curve), against the counterfactual density function $f(y_1|S=0)$ (the dashed line). We use kernel density estimation to smooth these functions.

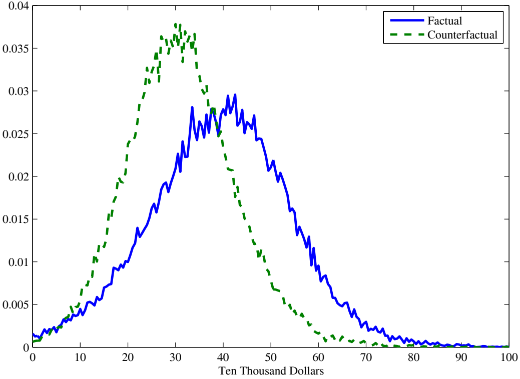
Figure 7: Densities of Present Value of Earnings, College Sample



Present value of earnings from age 22 to 36 for College Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_1|S=1)$ (the solid curve), against the counterfactual density function $f(y_0|S=1)$ (the dashed line). We use kernel density estimation to smooth these functions.

Figure 7: Densities of Present Value of Earnings, College Sample

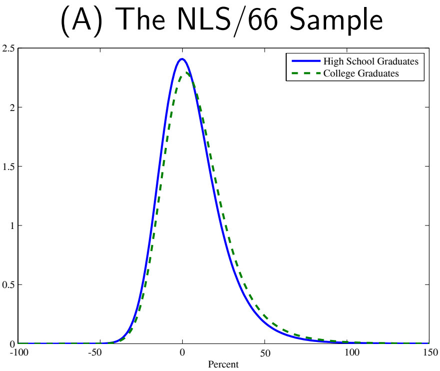
(B) NLSY/79



Present value of earnings from age 22 to 36 for College Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_1|S=1)$ (the solid curve), against the counterfactual density function $f(y_0|S=1)$ (the dashed line). We use kernel density estimation to smooth these functions.

- From such distributions, we can generate the distribution of *ex post* gross rate of return R to college (excluding costs) as $R = \frac{Y_1 - Y_0}{Y_0}$.

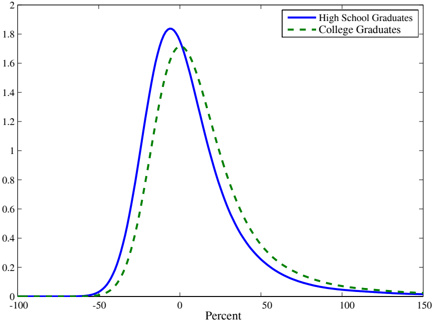
Figure 8: Densities of Returns to College



Let Y_0, Y_1 denote the present value of earnings from age 22 to age 36 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 the innovations $\varepsilon_{s,t}$ for $s =$ high school, college and $t = 22, \cdots, 36$.

Figure 8: Densities of Returns to College

(B) The NLSY/79 Sample



Let Y_0, Y_1 denote the present value of earnings from age 22 to age 36 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 the innovations $\varepsilon_{s,t}$ for $s =$ high school, college and $t = 22, \cdots, 36$.

Table 2: Mean Rates of Return per Year of College by Schooling Group

		NLS/66		NLSY/79
Schooling Group	Mean Returns	Standard Error	Mean Returns	Standard Error
High School Graduates	0.0592	0.0046	0.0955	0.0063
College Graduates	0.0877	0.0070	0.1355	0.0080
Individuals at the Margin	0.0750	0.0178	0.1184	0.0216

The Evolution of Uncertainty and Heterogeneity

- Under risk neutrality, the valuation or net utility function for schooling is

$$I = E\left(\sum_{t=1}^{T^*} \frac{Y_{1,t} - Y_{0,t}}{(1+\rho)^{t-1}} \middle| \mathcal{I}_1\right) - E\left(C_{T^*} \middle| \mathcal{I}_1\right),$$

where

$$C_{T^*} = -\sum_{t=T^*+1}^{\bar{T}} \frac{1}{(1+\rho)^{t-1}} \left(Y_{1,t} - Y_{0,t}\right) + C.$$

- T^* is last age of observation of earnings.

Total Residual Variance and Variance of Unforecastable Components

$$P_s = \sum_{t=1}^{T^*} \frac{\theta_3 \alpha_{3,s,t} + \mathcal{T}_t \phi + \varepsilon_{s,t}}{(1+\rho)^{t-1}} \tag{14}$$

- $\mathcal{T}_t$: year dummies on future values of the macro economy
- θ_1 and θ_2: estimated to be known to the agent

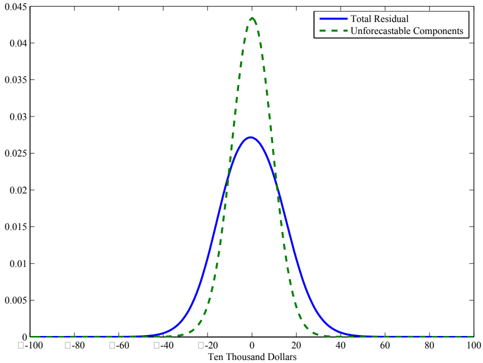
THE UNIVERSITY OF CHICAGO

Table 3: Evolution of Uncertainty

	College	High School	Returns
NLS/1966			
	College	High School	Returns
Total Variance	195.882	136.965	611.245
Variance of Unforecastable Components	76.332	31.615	167.187
Variance of Forecastable Components	119.550	105.350	444.058
NLS/1979			
	College	High School	Returns
Total Variance	292.368	165.350	823.200
Variance of Unforecastable Components	84.464	48.137	221.976
Variance of Forecastable Components	207.904	117.214	601.223
Evolution			
Percentage Increase in Total Variance	49.26%	20.72%	34.68%
Percentage Increase in Variance of Unforecastable Components	10.65%	52.26%	32.77%
Percentage Increase in Variance of Forecastable Components	73.90%	11.26%	35.39%
Percentage Increase in Total Variance by Source			
	College	High School	Returns
Percentage Increase in Total Variance due to Unforecastable Components	8.43%	58.20%	25.85%
Percentage Increase in Total Variance due to Forecastable Components	91.57%	41.80%	74.15%

Figure 9: The Densities of Total Residual vs. Unforecastable Components in Present Value of High School Earnings

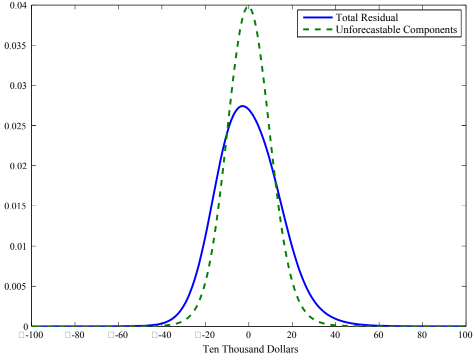
(A) The NLS/66 Sample



In this figure we plot the density of total residual (the solid curve) against the density of the unforecastable components (the dashed curve) for the present value of high school earnings from ages 22 to 36 for the NLS/66 and NLSY/79 samples of white males. The present value of earnings is calculated using a 5% interest rate.

Figure 9: The Densities of Total Residual vs. Unforecastable Components in Present Value of High School Earnings

(B) The NLSY/79 sample

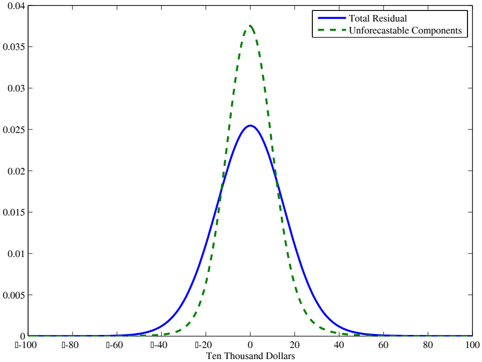


In this figure we plot the density of total residual (the solid curve) against the density of the unforecastable components (the dashed curve) for the present value of high school earnings from ages 22 to 36 for the NLS/66 and NLSY/79 samples of white males. The present value of earnings is calculated using a 5% interest rate.

- Figures 10A and 10B make the analogous comparison for present values of college earnings for the 1979 and 1966 samples.

Figure 10: The Densities of Total Residual vs. Unforecastable Components in Present Value of College Earnings

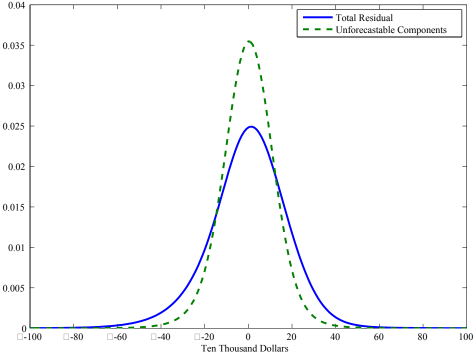
(A) The NLS/66 sample



In this figure we plot the density of total residual (the solid curve) against the density of the unforecastable components (the dashed curve) for the present value of college earnings from ages 22 to 36 for the NLS/66 and NLSY/79 samples of white males. The present value of earnings is calculated using a 5% interest rate.

Figure 10: The Densities of Total Residual vs. Unforecastable Components in Present Value of College Earnings

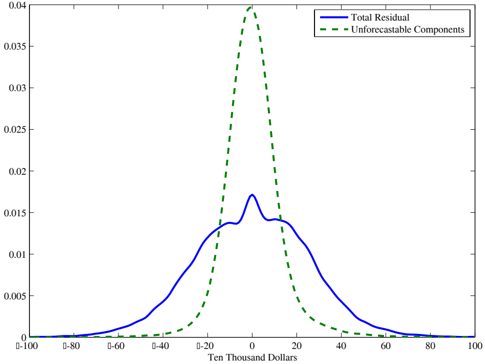
(B) The NLSY/79 sample



In this figure we plot the density of total residual (the solid curve) against the density of the unforecastable components (the dashed curve) for the present value of college earnings from ages 22 to 36 for the NLS/66 and NLSY/79 samples of white males. The present value of earnings is calculated using a 5% interest rate.

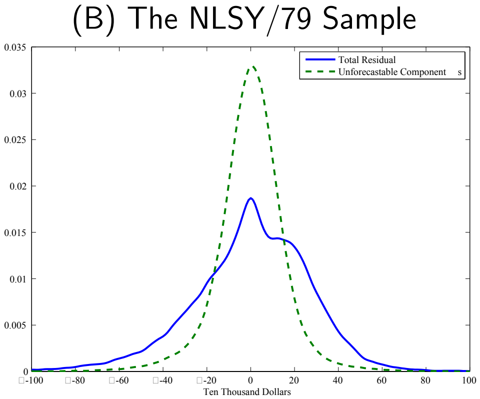
Figure 11: The Densities of Total Residual vs. Forecastable Components Returns College vs. High School

(A) The NLS/66 Sample



In this figure we plot the density of total residual (the solid curve) against the density of the unforecastable components (the dashed curve) for the present value of earnings differences (or returns to college) from ages 22 to 36 for the NLS/66 and NLSY/79 samples of white males. The present value of earnings is calculated using a 5% interest rate.

Figure 11: The Densities of Total Residual vs. Forecastable Components Returns College vs. High School



In this figure we plot the density of total residual (the solid curve) against the density of the unforecastable components (the dashed curve) for the present value of earnings differences (or returns to college) from ages 22 to 36 for the NLS/66 and NLSY/79 samples of white males. The present value of earnings is calculated using a 5% interest rate.

The Variance of the Unforecastable and Forecastable Components by Age

- Figure 12A plots the variances of unforecastable components by age in high school earnings in NLS/1966, and NLSY/1979.

Figure 12: Profile of Variance of Uncertainty

(A) High School Sample, NLS/66 vs NLSY/79

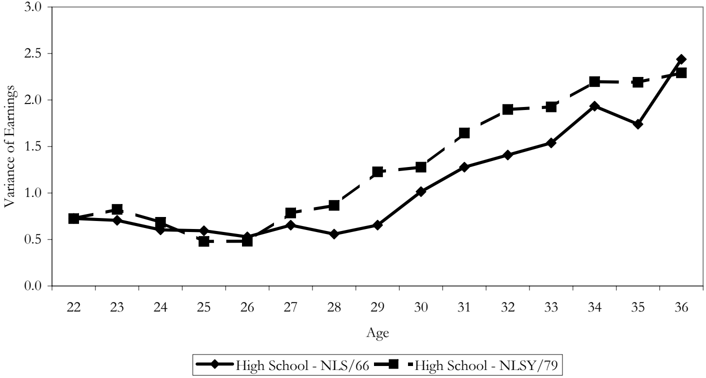


Figure 12: Profile of Variance of Uncertainty

(B) College Sample, NLS/66 vs NLSY/79

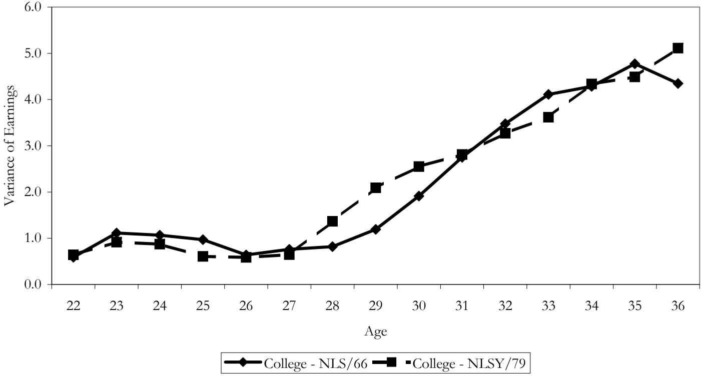


Figure 13: Profile of Variance of Heterogeneity

(A) High School Sample, NLS/66 vs NLSY/79

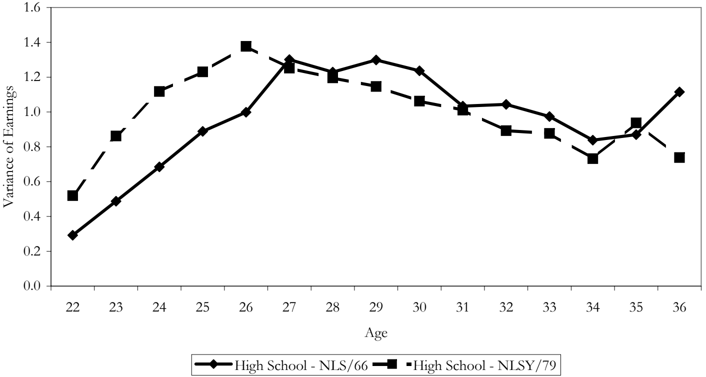
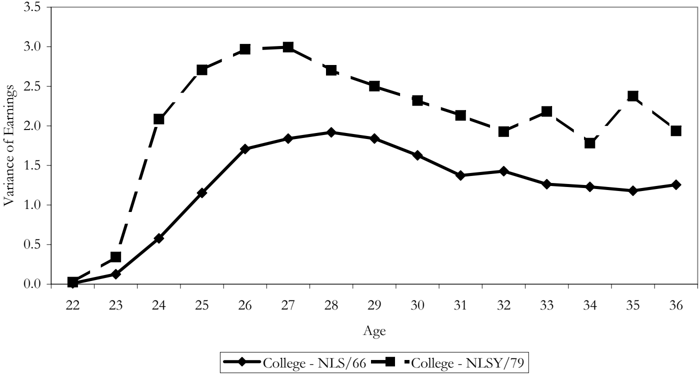


Figure 13: Profile of Variance of Heterogeneity

(B) College Sample, NLS/66 vs NLSY/79



Accounting for Macro Uncertainty

Table 4: Share of Variance of Business Cycle in Total Variance of Unforecastable Components

	NLS/1966		NLSY/1979	
	Point Estimate	Standard Error	Point Estimate	Standard Error
High School	0.1111	0.0147	0.0156	0.0020
College	0.0452	0.0077	0.0392	0.0052
Overall	0.0679	0.0107	0.0328	0.0042

Risk Aversion and More General Market Structures

- A basic question is *What can be identified in more general environments?*
- Basic identification problem: how to separate
  (a) Uncertainty and UN
  (b) Credit constraints
  (c) Preferences

Accounting for Inequality

We show that the distribution of $Y_{k,i}$ for each cohort, displayed in the first row of

- Table 5A (for the Gini index)
- Table 5B (for the Theil index)
- Table 6 (for the Atkinson index)

Table 5: Predictable Heterogeneity

A. Gini Decomposition			
	NLS/66	NLSY/79	% Growth
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.1803	0.2088	15.85%
Counterfactual: Predictable Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.1591	0.1825	14.73%

B. The Theil Entropy Index T (Overall)			
	NLS/66	NLSY/79	% Growth
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.0502	0.0693	37.98%
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.0390	0.0522	33.76%
Within Schooling Groups			
	NLS/66	NLSY/79	% Change
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.0491	0.0631	28.53%
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.0378	0.0465	22.85%
Between Schooling Groups			
	NLS/66	NLSY/79	% Change
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.0011	0.0062	447.37%
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.0011	0.0057	394.22%

[1] Let $Y_{k,s,t,i}$ denote the earnings of an agent i, $i = 1, ..., n_k$, at age t, $t = 22, ..., 36$, in schooling level s, $s =$ high school, college, and cohort k, $k = NLS/1966, NLSY/1979$. We model earnings $Y_{k,s,t,i}$ as:

$$Y_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i} + \theta_{3,k,i}\alpha_{3,k,s,t,i} + \varepsilon_{k,s,t,i}. \qquad (1)$$

The present value of earnings at schooling level s, $Y_{k,s,i}$, is $Y_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y_{k,s,t,i}}{(1+\rho)^{t-1}}$. The observed present value of earnings satisfies $Y_{k,i} = S_{k,i}Y_{k,1,i} + \left(1 - S_{k,i}\right) Y_{k,0,i}$ where $S_{k,i} = 1$ if agent i in cohort k graduates college, and $S_{k,i} = 0$ if the person graduates high school. Let $C_{k,i}$ denote the direct costs for individual i in cohort k.

The schooling choice is:

$$S_{k,i} = 1 \Leftrightarrow E\left(Y_{k,1,i} - Y_{k,0,i} - C_{k,i} \middle| \mathcal{I}_k \right) \geq 0. \tag{2}$$

This is the factual economy. In this row, we show the inequality measure in the subtitle.
[2]We simulate the economy by replacing (1) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}\left(X_k\right) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are as determined in (2). In this row, we show the inequality measure for the concept given in the subtitle for the observed truncated present value of earnings $Y^h_{k,s,i}$ when we constrain schooling choices to be the same as in the economy that generates the first row.

THE UNIVERSITY OF CHICAGO

Table 6: Atkinson Index

		$\varepsilon = 0.5$	
	NLS/66	NLSY/79	%Change
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.0276	0.0389	0.4111
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.0213	0.0286	0.3437
		$\varepsilon = 1.5$	
	NLS/66	NLSY/79	%Change
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.0968	0.1467	0.5147
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.0716	0.0980	0.3687

$$\mu_k = \frac{1}{n_k}\sum_{i=1}^{n_k} Y_{k,i}$$

- Atkinson index (1970)
- $\bar{Y}_k$

$$\frac{\left(\bar{Y}_k\right)^{1-\epsilon}-1}{1-\epsilon} = \frac{1}{n_k}\sum_{i=1}^{n_k}\frac{Y_{k,i}^{1-\epsilon}-1}{1-\epsilon}$$

- The Atkinson index A is defined as:

$$A = 1-\left(\frac{\bar{Y}_k}{\mu_k}\right)$$

Table 6: Atkinson Index

		$\varepsilon = 1.0$	
	NLS/66	NLSY/79	%Change
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.0586	0.0847	0.4446
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.0447	0.0604	0.3503
		$\varepsilon = 2.0$	
	NLS/66	NLSY/79	%Change
Factual Economy: Predictable Heterogeneity and Uncertainty[1]	0.1627	0.2627	0.6149
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Predictable Heterogeneity Only[2]	0.1060	0.1506	0.4205

[1] Let $Y_{k,s,t,i}$ denote the earnings of an agent i, $i = 1, ..., n_k$, at age t, $t = 1, ..., T$, in schooling level s, $s =$ high school, college, and cohort k, $k = NLS/1966, NLSY/1979$. We model earnings $Y_{k,s,t,i}$ as:

$$Y_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i} + \theta_{3,k,i}\alpha_{3,k,s,t,i} + \varepsilon_{k,s,t,i}. \quad (1)$$

The present value of earnings in schooling level s, $Y_{k,s,i}$, is $Y_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y_{k,s,t,i}}{(1+\rho)^{t-1}}$. The observed truncated present value of earnings is $Y_{k,i} = S_{k,i}Y_{k,1,i} + \left(1 - S_{k,i}\right) Y_{k,0,i}$. Let $C_{k,i}$ denote the direct costs for individual i in cohort k. The schooling choice is:

$$S_{k,i} = 1 \Leftrightarrow E\left(Y_{k,1,i} - Y_{k,0,i} - C_{k,i} \middle| \mathcal{I}_k \right) \geq 0. \quad (2)$$

THE UNIVERSITY OF CHICAGO

This is the factual economy. We then compute the average present value of earnings across all individuals in cohort k, $\mu_k = \frac{1}{n}\sum_{i=1}^{n_k} Y_{k,i}$. For a given inequality aversion parameter ϵ, we compute the level of permanent income $\bar{Y}_k(\epsilon)$ that generates the same welfare as the social welfare of the actual distribution in cohort k:

$$\frac{\left[\bar{Y}_k(\epsilon)\right]^{1-\epsilon} - 1}{1-\epsilon} = \frac{1}{n_k}\sum_{i=1}^{n_k} \frac{\left(Y_{k,i}\right)^{1-\epsilon} - 1}{1-\epsilon}$$

For each value of ϵ, the Atkinson Index is $A(\epsilon) = 1 - \frac{\bar{Y}_k(\epsilon)}{\mu_k}$. In this row, we show the Atkinson Index for the observed present value of earnings $Y_{k,i}$ for different values of ϵ.
[2]We simulate the economy by replacing (1) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only predictable heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are as determined in (2). In this row, we show the Atkinson Index for the observed present value of earnings $Y^h_{k,i}$ for different values of ϵ when we constrain schooling choices, $S_{k,i}$, to be observed in the factual economy.

THE UNIVERSITY OF CHICAGO

- Sharp differences in the contribution of rising uncertainty to inequality for different schooling groups.
- High school graduates' earnings variability is due to a substantial rise in inequality due to uncertainty.
- Uncertainty in college graduate earnings has not increased substantially, although predictable components have become more variable.

Table 7: Ex-Ante Conditional Distributions for the NLSY/79 (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 Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution Individuals fix unknown θ at their means, so $\theta_3 = 0$ Correlation $(Y_C, Y_H) = -0.8730$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0002	0.0000	0.0014	0.0053	0.0175	0.0323	0.0645	0.0987	0.1825	0.5976
2	0.0000	0.0020	0.0106	0.0248	0.0524	0.0808	0.0968	0.1598	0.2978	0.2750
3	0.0008	0.0072	0.0276	0.0488	0.0782	0.0976	0.1368	0.2304	0.2774	0.0952
4	0.0006	0.0234	0.0608	0.0814	0.0932	0.1350	0.1932	0.2230	0.1678	0.0216
5	0.0052	0.0430	0.0872	0.0900	0.1362	0.1880	0.2146	0.1744	0.0574	0.0040
6	0.0144	0.0662	0.1038	0.1400	0.1798	0.2166	0.1778	0.0856	0.0158	0.0000
7	0.0318	0.1108	0.1443	0.1895	0.2273	0.1693	0.0960	0.0276	0.0034	0.0000
8	0.0792	0.1609	0.2219	0.2711	0.1757	0.0698	0.0188	0.0020	0.0006	0.0000
9	0.1765	0.3325	0.2931	0.1437	0.0400	0.0112	0.0026	0.0002	0.0002	0.0000
10	0.6064	0.3216	0.0637	0.0078	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000

Table 8: Ex-Ante Conditional Distributions for the NLSY/66 (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 Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution Individuals fix unknown θ at their means, so $\theta_3 = 0$ Correlation $(Y_C, Y_H) = -0.9332$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0012	0.0020	0.0004	0.0004	0.0010	0.0010	0.0056	0.0396	0.2065	0.7422
2	0.0014	0.0000	0.0000	0.0010	0.0034	0.0210	0.0966	0.2664	0.4118	0.1984
3	0.0012	0.0002	0.0004	0.0048	0.0342	0.1148	0.2464	0.3142	0.2322	0.0516
4	0.0004	0.0004	0.0048	0.0326	0.1302	0.2442	0.2726	0.1942	0.1146	0.0060
5	0.0002	0.0014	0.0318	0.1218	0.2354	0.2658	0.1876	0.1256	0.0302	0.0002
6	0.0002	0.0130	0.1034	0.2494	0.2618	0.1862	0.1288	0.0514	0.0058	0.0000
7	0.0020	0.0774	0.2590	0.2864	0.1898	0.1216	0.0550	0.0088	0.0000	0.0000
8	0.0236	0.2616	0.3410	0.2042	0.1186	0.0436	0.0072	0.0000	0.0002	0.0000
9	0.1992	0.4510	0.2260	0.0966	0.0252	0.0018	0.0002	0.0000	0.0000	0.0000
10	0.7669	0.1961	0.0337	0.0028	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000

Table 9: Ex-Post Conditional Distributions for the NLSY/79 (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 Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution
Full Information Set Correlation$(Y_C, Y_H) = -0.8752$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0000	0.0000	0.0004	0.0029	0.0123	0.0263	0.0466	0.0936	0.2046	0.6132
2	0.0000	0.0010	0.0072	0.0220	0.0470	0.0762	0.1110	0.1818	0.2968	0.2570
3	0.0010	0.0074	0.0270	0.0454	0.0788	0.1172	0.1722	0.2250	0.2464	0.0796
4	0.0012	0.0160	0.0440	0.0784	0.1084	0.1528	0.2076	0.2184	0.1476	0.0256
5	0.0032	0.0368	0.0754	0.1108	0.1536	0.1816	0.1950	0.1596	0.0766	0.0074
6	0.0108	0.0668	0.1108	0.1644	0.1892	0.1950	0.1506	0.0850	0.0264	0.0010
7	0.0264	0.1129	0.1649	0.2084	0.2028	0.1547	0.0895	0.0326	0.0078	0.0000
8	0.0630	0.1994	0.0287	0.2195	0.1559	0.0794	0.0255	0.0068	0.0018	0.0000
9	0.1780	0.3341	0.2716	0.1409	0.0517	0.0177	0.0043	0.0012	0.0004	0.0000
10	0.5622	0.3368	0.0806	0.0157	0.0035	0.0012	0.0000	0.0000	0.0000	0.0000

Table 10: Ex-Post Conditional Distributions for the NLS/66 (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 Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution
Full Information Set Correlation$(Y_C, Y_H) = -0.8869$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0002	0.0016	0.0012	0.0006	0.0016	0.0067	0.0224	0.0668	0.2331	0.6657
2	0.0006	0.0002	0.0016	0.0064	0.0216	0.0526	0.1280	0.2388	0.3354	0.2148
3	0.0002	0.0012	0.0060	0.0250	0.0692	0.1360	0.2112	0.2526	0.2218	0.0768
4	0.0004	0.0046	0.0262	0.0694	0.1302	0.1938	0.2234	0.2024	0.1240	0.0256
5	0.0016	0.0156	0.0636	0.1326	0.1906	0.2166	0.1864	0.1262	0.0580	0.0086
6	0.0032	0.0452	0.1294	0.2028	0.2090	0.1832	0.1270	0.0734	0.0242	0.0026
7	0.0188	0.1112	0.2180	0.2306	0.1894	0.1260	0.0684	0.0310	0.0060	0.0006
8	0.0620	0.2348	0.2666	0.1966	0.1368	0.0650	0.0284	0.0082	0.0014	0.0002
9	0.2220	0.3639	0.2204	0.1190	0.0488	0.0192	0.0048	0.0016	0.0002	0.0000
10	0.6762	0.2320	0.0699	0.0178	0.0027	0.0010	0.0004	0.0000	0.0000	0.0000

Table 11: Percentage that Regret Schooling Choices

Schooling Group	NLS/1966	NLSY/1979
Percentage of High School Graduates who Regret Not Graduating from College	0.1522	0.1659
Percentage of College Graduates who Regret Graduating from College	0.1402	0.1495