Counterfactual Analysis of Inequality and Social Mobility

Flavio Cunha, James Heckman and Salvador Navarro
(2005)

Econ 374, Spring 2012
This draft, May 28, 2012

- Move beyond aggregate summary measures of policy outcomes to gauge the effects of a policy on subgroups defined by *unobserved* potential outcomes within the overall population distribution.
- Move beyond traditional inequality and social mobility analysis to consider how a policy shifts persons from a position in one potential outcome distribution to another even though joint potential outcome distributions cannot be directly measured, but must be derived from marginal outcome distributions for program participants and nonparticipants.

Welfare Economics and Social Choice

- The standard model of welfare economics postulates a social welfare function V defined over the utilities u_i under policy j of the N members of society,

$$V_j = V\left[u_1^j, u_2^j, \ldots, u_N^j\right],$$

  where V_j is the society's welfare under policy j and u_i^j is the utility of person i under policy j.
- One common example is the Benthamite social welfare function $V_j = \sum_{i=1}^N u_i^j$. Another possibility is the Rawlsian social welfare function $V_j = \min\left\{u_1^j, u_2^j, \ldots, u_N^j\right\}$.

- Standard criteria used to evaluate policies and compare income distributions including those based on the preceding welfare criteria, as well as conventional cost benefit analysis, invoke:
- **Anonymity Axiom** (Cowell, 2000).
- Let $\left(Y_1^A, Y_2^A, \ldots, Y_N^A\right)$ and $\left(Y_1^B, Y_2^B, \ldots, Y_N^B\right)$
- The subscripts denote individuals.
- Assume that under policy B everything else is the same as under policy A, except that the outcomes for agents 1 and 2 under policy B are exchanged:

$$\left(Y_1^B, Y_2^B, \ldots, Y_N^B\right) = \left(Y_2^A, Y_1^A, \ldots, Y_N^A\right).$$

- According to the anonymity axiom, any social welfare ordering over these two policies or states of affairs should be indifferent between policies A and B, since overall inequality is the same.

- Adopting the anonymity axiom is empirically convenient because its implementation only requires information on the marginal distributions of outcomes under different policies, and not the joint distributions of outcomes across policy states.
- The anonymity axiom makes strong assumptions.
- The main problem is that individual outcomes under alternative policies are either assumed to be independent or any such dependence of outcomes across policy states is assumed to be irrelevant in assessing the merits of alternative policies.
- The initial position of persons is assumed not to affect judgments about final outcomes of a policy.

- However, if the joint distributions of policy outcomes can be recovered, we can assess how the median voter would evaluate a proposed reform, both *ex post* and *ex ante*, and see what percentage of a population would favor the reform given their initial position—the desiderata of modern positive political economy (See Persson and Tabellini, 2000).
- If only the two marginal distributions (pre- and post-policy) are available, we cannot assess how the median voter, who is interested in how a policy affects his movements from the baseline to the final states, would evaluate that policy unless one assumes something about the dependence of outcomes for persons across policies.

- In any actual policy setting, it is likely that persons, or groups of persons, have at least partial knowledge about how they will fare under different policy regimes.
- Thus, even if outcomes in alternative policy regimes are not completely known, outcomes under the policy in place are known.
- The outcomes in different regimes are likely to be dependent so that persons who benefit under one policy are also likely to benefit under another.
- However, due to uncertainty, these outcomes are unlikely to be perfectly dependent.

- Consequently, for a variety of actual social choice mechanisms, both the initial and final positions of each agent are relevant for evaluation of social policy, but the exact dependence is unknown to the analyst.
- Below we show how the methodology presented here can be applied to identify people who gain or lose from each policy at various deciles of initial or final distributions, relaxing the anonymity axiom.
- We can do such analyses for factual or counterfactual distributions.
- We also allow for uncertainty in the evaluation of outcome states not yet experienced. Thus we can distinguish between *ex ante* and *ex post* evaluations of a reform.

The Evaluation of Social Programs: Choices Within Policy States and Comparisons Across Policy States

- "*ex ante*" and "*ex post*"

- There are two possible outcomes within each policy regime.
- Let $S = 0$, and $S = 1$ denote nonreceipt and receipt of education, respectively, within a policy regime.
- In our empirical analysis, $S = 0$ denotes a worker who is a high school graduate, and $S = 1$ a worker who is a college graduate.
- $S = 0$
- $S = 1$

- Associated with each level of education is a potential outcome.
- Let (Y_0, Y_1) denote potential outcomes in state $S = 0$ and $S = 1$, respectively within a given policy regime.
- Each person has a (Y_0, Y_1) pair.
- We assume that (Y_0, Y_1) have finite means and can be expressed in terms of conditioning variables X in the following manner:

$$Y_0 = \mu_0(X) + U_0 \tag{1a}$$
$$Y_1 = \mu_1(X) + U_1, \tag{1b}$$

  where $E(Y_0 \mid X) = \mu_0(X)$, $E(Y_1 \mid X) = \mu_1(X)$ and $E(U_0 \mid X) = E(U_1 \mid X) = 0$.
- The gain for an individual who moves from the $S = 0$ to $S = 1$ within a policy regime is Δ, where $\Delta \equiv Y_1 - Y_0$.

- An evaluation problem within a policy regime arises because we do not observe the pair (Y_0, Y_1) for anybody.
- The econometric approach features the use of choice data in constructing counterfactuals.

- For simplicity, and in accordance with a well established tradition in econometrics, we write index I as a net utility

$$I = Y_1 - Y_0 - C \tag{2}$$

where C is the cost of participation in sector 1.

- We write $C = \mu_C(Z) + U_C$ where the Z are observed (by the analyst) determinants of cost and U_C denotes unobserved determinants of C from the point of view of the analyst.
- In reduced form (substituting out for Y_1, Y_0 and C), we may write

$$I = \mu_I(X, Z) + U_I$$

where

$$\mu_I(X, Z) = \mu_1(X) - \mu_0(X) - \mu_C(Z)$$

and

$$U_I = U_1 - U_0 - U_C.$$

- We write
$$S = 1 \text{ if } I \geq 0; \; S = 0 \text{ otherwise.} \tag{3}$$
- Thus if the net utility of state 1 is positive, $S = 1$ is chosen.
- Other decision rules may be used, but the model of (Y_1, Y_0, S) is sufficiently rich to serve our purposes.

- Overall income within the policy regime is $Y = SY_1 + (1-S)Y_0$.
- Traditional analyses of inequality compare the distribution of Y across policy regimes that are observed.
- We consider the consequences of choices ($S = 0$ or $S = 1$) within policy regimes and how alternative policies cause people to change their S decisions and relocate into different portions of the overall distribution.
- We can do a parallel analysis for those who switch from $S = 1$ to $S = 0$, reversing the roles of Y_0 and Y_1.
- We can do this for counterfactuals as well as for factuals.
- We can also do counterfactual social mobility analysis.

- Traditionally, the literature on program evaluation has focused on estimating mean impacts of S and not distributions.
- The most commonly studied parameter in the literature is the average treatment effect:

$$ATE = E\left(\Delta \mid X\right) = E\left(Y_1 - Y_0 \mid X\right).$$

- Another popular parameter is the effect of treatment on the treated,

$$TT = E\left(\Delta \mid X, S = 1\right) = E\left(Y_1 - Y_0 \mid X, S = 1\right).$$

- The modern literature allows for the possibility that the gains to switching from $S = 0$ to $S = 1$, $Y_1 - Y_0$, are heterogenous across agents even conditioning on X.
- Further, the agents act on this difference when choosing S.
- In the analysis of this model, two problems emerge.

- The proportion of people taking schooling that benefit from it in terms of gross returns $\Delta \left(= Y_1 - Y_0\right)$ is $\Pr\left(\Delta > 0 \mid S = 1\right)$.
- This parameter is one way to measure how widely program gains are distributed among participants.
- The proportion of the total population benefiting from participating in schooling is $\Pr\left(\Delta > 0 \mid S = 1\right) \cdot \Pr\left(S = 1\right)$.
- It is of interest to determine how many people in society at large benefit (in the sense of $Y_1 - Y_0$ gains) from participating in schooling.

- The distribution of gains from schooling for agents who are at selected base state values is $\Pr(\Delta \leq a \mid S = 1, Y_0 = y_0)$.
- This measure interests Rawlsian evaluators who seek to determine the impact of schooling on recipients in the lower tail of the base state distribution.

- The increase in the level of outcomes above a certain threshold, say the poverty line $\bar{y}$, due to schooling is $\Pr(Y_1 > \bar{y} \mid S = 1) - \Pr(Y_0 > \bar{y} \mid S = 1)$.
- This is a parameter that describes how the distribution of the outcomes for the participants compares to the distribution of the outcomes for the same agents if they had not participated in schooling.

- We can also form measures for people affected by a specific policy.
- Let A and B denote two policy states, say a high tuition and a low tuition policy, respectively.
- The proportion of people who benefit from a policy that induces them into schooling (*e.g.*, a reduction in tuition) is $\Pr(\Delta > 0 \mid S_A = 0, S_B = 1)$, where the measure of benefit is a gross gain measure and S^A and S^B are choice indicators under policy A and B, respectively.
- We can also measure the proportion of the total population that benefits from the policy:
  $\Pr(\Delta > 0 \mid S_A = 0, S_B = 1) \cdot \Pr\left(S_A = 0, S_B = 1\right).$
- Our empirical analysis reports these and other measures of impact that we can define and estimate both within a policy and across policy regimes.

- It is fruitful to distinguish between two kinds of policies: (a) those that affect potential outcomes $\left(Y_0^A, Y_1^A\right)$ for outcomes and costs $\left(C^A\right)$ under policy regime A through price and quality effects and (b) those that affect sectorial choices (through C^A), but do not affect potential outcomes.
- Tuition and educational access policies that do not produce general equilibrium effects fall into the second category of policy.
- It is the second kind of policy that receives the most attention in empirical work on the economics of education, either when estimating gains to schooling under a policy regime $\left(Y_1^A - Y_0^A\right)$ (see *e.g.*, Card, 1999) or evaluating schooling policies (*e.g.*, Kane, 1994).

- Consider two general policy environments denoted A and B.
- These policies might affect the costs of schooling including access to it.
- In the general case, we could have $\left(Y_0^A, Y_1^A, C^A\right)$ and $\left(Y_0^B, Y_1^B, C^B\right)$ for each person.
- There might be general equilibrium policies or policies that operate in the presence of social interactions that affect both costs and outcomes.

- A special case of this policy produces two social states for outcomes that we wish to compare.
- However, in this special case, interventions have no effect on potential outcomes and can be described as producing two choice sets $\left(Y_0, Y_1, C^A\right)$ and $\left(Y_0, Y_1, C^B\right)$ for each person.
- They affect costs and the choice of outcomes, but not the potential outcomes as a full-fledged general equilibrium or social interaction analysis would do.
- We focus most of our attention on policies that keep potential schooling outcomes unchanged but that vary C in selecting who takes schooling.

- Two sets of counterfactuals: (a) (Y_0^A, Y_1^A) within policy regime A and (Y_0^B, Y_1^B) within policy regime B, and (b) aggregate income across policy regimes $\left(Y^A, Y^B\right)$ where $Y^A = Y_1^A S^A + Y_0^A(1 - S^A)$ is the observed income under regime A and $Y^B = Y_1^B S^B + Y_0^B(1 - S^B)$ is the income under regime B, where $S^A = 1$ if a person chose $S = 1$ under regime A and S^B is defined in an analogous fashion.

- We can construct counterfactual distributions of $\left(Y_1^A, Y_0^A\right)$ and $\left(Y_1^B, Y_0^B\right)$ within each policy regime and can also construct comparisons across policy states based on $Y_1^A S^A + Y_0^A(1 - S^A)$ and $Y_1^B S^B + Y_0^B(1 - S^B)$.

Identifying Counterfactual Distributions Using Factor Models

- As shown by Heckman (1990) and Heckman and Smith (1998), under the assumptions that (Z, X) are statistically independent from (U_0, U_1, U_I), $\mu_I(X, Z)$ is a nontrivial function of Z given X, and full support on $\mu_0(X), \mu_1(X)$ and $\mu_I(X, Z)$, and an assumption that the elements of the pairs $(\mu_0(X), \mu_I(X, Z))$ and $(\mu_1(X), \mu_I(X, Z))$ can be varied independently of each other, then one can identify the joint distributions of $\left(U_0, \frac{U_I}{\sigma_I}\right)$and $\left(U_1, \frac{(U_I)}{\sigma_I}\right)$ and also $\mu_0(X)$, $\mu_1(X)$, and $\frac{\mu_I(X,Z)}{\sigma_I}$.
- Thus, one can identify the joint distributions of (Y_0, I^*) and (Y_1, I^*) given X and Z where $I^* = I/\sigma_I$.
- One cannot recover the conditional (on X, Z) joint distribution of (Y_0, Y_1) or (Y_0, Y_1, I^*) without further assumptions.

- We provide an intuitive motivation for why $F(Y_0, I^*)$ and $F(Y_1, I^*)$ are identified, drawing on standard results in the semiparametric discrete choice literature.
- The thrust of this literature is that under the stated conditions, we can identify the distribution of I up to a factor of proportionality, σ_I.
- We can also identify $F(Y_0, I | I < 0, X, Z) = F(Y_0 | D = 0, X, Z) \Pr(D = 0 | X, Z)$ and $F(Y_1, I | I \geq 0, X, Z) = \Pr(Y_1 | D = 1, X, Z) \Pr(D = 1 | X, Z)$.
- By varying X, Z we can trace out the full distributions of $F(Y_0, I)$ and $F(Y_1, I)$ respectively.
- Once we estimate the distributions, we perform conventional factor analysis on (Y_0, I^*) and (Y_1, I^*) because, effectively, we observe these two distributions.

Factor Models

- (U_0, U_1, U_{I^*}) is generated by a scalar factor.

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

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

Recovering the Factor Loadings

The Case When There is Information Only on Y_0 for $I < 0$ and Y_1 for $I > 0$ but the Decision Rule is (2)–(3)

- From these distributions one can identify the left hand sides of the following two equations:

$$
\begin{aligned}
Cov\left(U_0, U_{I^*}\right) &= \alpha_0 \alpha_{I^*} \sigma_\theta^2 \\
Cov\left(U_1, U_{I^*}\right) &= \alpha_1 \alpha_{I^*} \sigma_\theta^2 .
\end{aligned}
$$

- As previously noted, the scale of the unobserved I is normalized, a standard condition for discrete choice models.
- A second normalization that we need to impose is that $\sigma_\theta^2 = 1$.
- This is required since the factor is not observed and we must set its scale.
- That is, since $\alpha\theta = k\alpha\frac{\theta}{k}$ for any constant k, we need to set the scale by, say, normalizing the variance of θ.
- We could alternatively normalize some α_0 or α_1 to one.
- Finally, if we set $\alpha_{I^*} = 1$ (something we can relax, as noted below and in the next section), then we identify α_1 and α_0 from the known covariances above.

- Since
$$Cov\left(U_1, U_0\right) = \alpha_1 \alpha_0 \sigma_\theta^2$$
we can identify the covariance between Y_1 and Y_0 even though we do not observe both Y_0 and Y_1 for anyone.
- We then use the variances $Var\left(U_1\right), Var\left(U_0\right)$ and the normalization $Var\left(U_{I^*}\right) = 1$ to recover the variances of the uniquenesses $\sigma_{\varepsilon_0}^2, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_{I^*}}^2$.

- The fact that we needed to normalize both $\sigma_\theta^2 = 1$ and $\alpha_{I^*} = 1$ is a consequence of our assumption that we have only one observation for Y_1 or Y_0 for each person.
- If we have access to more observations (say from panel data) or to more equations that depend on the factor (as in the next section), we can relax the normalizations, say $\sigma_\theta^2 = 1$, since then we could form for a panel of length T, the left hand sides of the following equations:

$$\begin{aligned} \frac{Cov\left(Y_{1,t'}, I^*\right)}{Cov\left(Y_{1,t'}, Y_{1,t}\right)} &= \alpha_{1,t}, \; t = 1, ..., T \\ \frac{Cov\left(Y_{0,t'}, I^*\right)}{Cov\left(Y_{0,t'}, Y_{0,t}\right)} &= \alpha_{0,t}, \; t = 1, ..., T \end{aligned}$$

  and recover σ_θ^2 from, say, $Cov\left(Y_{1,t}, I^*\right) = \alpha_{1,t}\sigma_\theta^2$, given the normalization $\sigma_{I^*}^2 = 1$.
- The variances of the uniquenesses follow as before.

- The crucial idea motivating this identification strategy is that even though we never observe (Y_0, Y_1) as a pair, both Y_0 and Y_1 are linked to S through the choice equation.
- From information on choice S we can recover I^* from a standard identification argument in econometrics.
- Thus, we essentially observe (Y_0, I^*) and (Y_1, I^*).
- The common low dimensional dependence of Y_0 and Y_1 on I^* secures identification of the joint distribution of Y_0, Y_1, I^*.
- This plays the role of I^* and in certain respects identification with a measurement is more transparent and more traditional.

Adding a Measurement Equation

- Measured ability M is

$$M = \mu_M(X) + U_M.$$

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

- We assume $\alpha_M \neq 0$.

$$\begin{aligned} Cov\left(U_M, U_0\right) &= \alpha_M \alpha_0 \sigma_\theta^2 \\ Cov\left(U_M, U_1\right) &= \alpha_M \alpha_1 \sigma_\theta^2 \\ Cov\left(U_M, U_{I^*}\right) &= \alpha_M \alpha_{I^*} \sigma_\theta^2. \end{aligned}$$

$$Y_0 - \mu_0\left(X\right) = U_0,$$

$$Y_1 - \mu_1\left(X\right) = U_1$$

- If we impose the normalization $\alpha_M = 1$.

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

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

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

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

- We can use the identified variances $Var\left(U_0\right)$, $Var\left(U_1\right)$, $Var\left(U_{I^*}\right) = 1$, and $Var\left(U_M\right)$ to recover the variance of the uniquenesses $\sigma_{\varepsilon_0}^2, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_{I^*}}^2$, and $\sigma_{\varepsilon_M}^2$.

- Identifying the distributions of the unobservables.

Recovering the Distributions Nonparametrically

Theorem

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

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

with θ, v_1, v_2 mutually statistically independent, $E(\theta) < \infty$, $E(v_1) = E(v_2) = 0$, that the conditions for Fubini's theorem are satisfied for each random variable, and the random variables possess nonvanishing characteristic functions, then the densities $f_\Theta(\theta), f_{V_1}(v_1)$, and $f_{V_2}(v_2)$ are identified.

Proof.

See Kotlarski (1967). $\square$

- This Theorem improved in Cunha, Heckman, and Schennach (2010)

- $\alpha_M = 1$
- The system is

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

- This system can be rewritten as

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

- We can identify the densities of $\theta, \frac{\varepsilon_{I*}}{\alpha_{I*}}, \frac{\varepsilon_0}{\alpha_0}, \frac{\varepsilon_1}{\alpha_1}, \varepsilon_M$.
- Since we know α_{I^*}, α_0 and α_1 we can identify the densities of $\theta, \varepsilon_{I^*}, \varepsilon_0, \varepsilon_1, \varepsilon_M$
- Thus, we can identify the distributions of all of the error terms.
- Finally, to recover the joint distribution of (Y_1, Y_0) given X, denoted $F(Y_1, Y_0 \mid X)$, note that

$$F(Y_1, Y_0 \mid X) = \int F(Y_1, Y_0 \mid \theta, X)\, dF(\theta),$$

  where $F(\theta)$ is the distribution of θ.
- From Kotlarski's theorem, $F(\theta)$ is known.
- Because of the factor structure, Y_1, Y_0 and S are independent once we condition on θ.

- So
$$F(Y_1, Y_0 \mid \theta, X) = F(Y_1 \mid \theta, X)\, F(Y_0 \mid \theta, X).$$
- But $F(Y_1 \mid \theta, X)$ and $F(Y_0 \mid \theta, X)$ are identified once we condition on the factors since
$$\begin{aligned} F(Y_1 \mid \theta, X, S = 1) &= F(Y_1 \mid \theta, X) \\ F(Y_0 \mid \theta, X, S = 0) &= F(Y_0 \mid \theta, X). \end{aligned}$$

- Our method generalizes matching by allowing the variables that would produce the conditional independence assumed in matching to be unobserved by the analyst.

Distinguishing between Heterogeneity and Uncertainty

- $$Y_{1,i} = \sum_{t=0}^{T} \frac{y_{1,i,t}}{(1+r)^t},$$
  $$Y_{0,i} = \sum_{t=0}^{T} \frac{y_{0,i,t}}{(1+r)^t},$$
  $$S_i = \begin{cases} 1, & \text{if } E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_i\right) \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

  $$\begin{aligned} y_{0,i,t} &= X_{i,t}\beta_{0,t} + v_{0,i,t} \\ y_{1,i,t} &= X_{i,t}\beta_{1,t} + v_{1,i,t} \\ C_i &= Z_i\gamma + v_{i,C}. \end{aligned}$$

- Suppose there exists a vector of factors $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)$

$$
\begin{aligned}
v_{0,i,t} &= \vec{\theta}_i \alpha_{0,t} + \varepsilon_{0,i,t} \\
v_{1,i,t} &= \vec{\theta}_i \alpha_{1,t} + \varepsilon_{1,i,t},
\end{aligned}
$$

$$C_i = Z_i \gamma + \vec{\theta}_i \alpha_C + \varepsilon_{i,C}.$$

$$I_i = E\left(\sum_{t=0}^{T} \frac{\left(X_{i,t}\beta_{1,t} + \vec{\theta}_i\alpha_{1,t} + \varepsilon_{1,i,t}\right) - \left(X_{i,t}\beta_{0,t} + \vec{\theta}_i\alpha_{0,t} + \varepsilon_{0,i,t}\right)}{(1+r)^t} - \left(Z_i\gamma + \vec{\theta}_i\alpha_C + \varepsilon_{i,C}\right) \;\middle|\; \mathcal{I}_i\right).$$

$S_i = 1$ if $I_i \geq 0$; $S_i = 0$ otherwise.

- If there is an element of the vector $\vec{\theta}_i$, say $\theta_{i,2}$ (factor 2), that has nonzero loadings on future earnings, say at age 40, in either counterfactual state, $\alpha_{2,s,40} \neq 0$, for $s = 0$ or 1 and factor $\theta_{i,2}$ is a determinant of schooling choices, then one can say that at the time of the schooling choice, the agent knew the unobservable captured by the factor 2 that affects future earnings.
- If $\theta_{i,2}$ does not enter the choice equation but explains future earnings, then $\theta_{i,2}$ is uncertain (not predictable) at the age the decisions are made.
- By assumption $\varepsilon_{i,C}$ is predictable but the future $\varepsilon_{1,i,t}$ and $\varepsilon_{0,i,t}$ are not predictable.

- The idea of our test is thus very simple: the components of future earnings that are forecastable are captured by the factors that are known by the agents when they make their educational choices.
- The predictable factors are estimated with a nonzero loading in the choice equation.
- The uncertainty in the decision regarding college is captured by the factors that the agent does not act on when making the decision of whether to attend college or not.
- In this case, the loadings (coefficients on these factors) in the choice equation would be zero.
- Carneiro, Hansen and Heckman (2003) provide exact conditions for identifying the factor loadings.
- Cunha, Heckman, and Navarro (2005) develop this analysis further.

Empirical Results

- $Y_{s,t}$ is generated by a two factor model,

$$Y_{s,t} = X\beta_{s,t} + \theta_1\alpha_{s,t,1} + \theta_2\alpha_{s,t,2} + \varepsilon_{s,t}. \tag{4}$$

Table 1
List of Covariates

Variable Name	Cost Function (Z)	Test System (X_{T})	PV Earnings* (X)
South at age 14	Yes	Yes	Yes
Urban at age 14	Yes	Yes	Yes
Parents Divorced	Yes	Yes	No
Number of Siblings	Yes	Yes	No
Mother's education	Yes	Yes	No
Father's Education	Yes	Yes	No
Family Income age 17	Yes	Yes	No
Dummy 1957	Yes	No	Yes
Dummy 1958	Yes	No	Yes
Dummy 1959	Yes	No	Yes
Dummy 1960	Yes	No	Yes
Dummy 1961	Yes	No	Yes
Dummy 1962	Yes	No	Yes
Dummy 1963	Yes	No	Yes
Dummy 1964	No	No	No
Age in 1980	No	Yes	No
Grade Completed 1980	No	Yes	No
Enrolled in 1980	No	Yes	No
Distance to College	Yes	No	No
Tuition at age 17	Yes	No	No

*Present Value of Earnings in thousands of dollars.

- We normalize $\alpha_{h,1,2} = 1$.

- Formally, let T_j denote the test score j :

$$T_j = X_T \omega_j + \theta_1 \alpha_{test_j,1} + \varepsilon_{test_j}. \tag{5}$$

- The cost function C is given by:

$$C = Z\gamma + \theta_1 \alpha_{C,1} + \theta_2 \alpha_{C,2} + \varepsilon_C \tag{6}$$

- $$V = E\left(Y_{c,1} + \frac{Y_{c,2}}{1+r} - Y_{h,1} - \frac{Y_{h,2}}{1+r} \mid X, \bar{\theta}\right) - E\left(C \mid Z, X, \bar{\theta}, \varepsilon_C\right) \tag{7}$$

- We assume that each factor k, $k \in \{1, 2\}$ is generated by a mixture of J_k normal distributions:

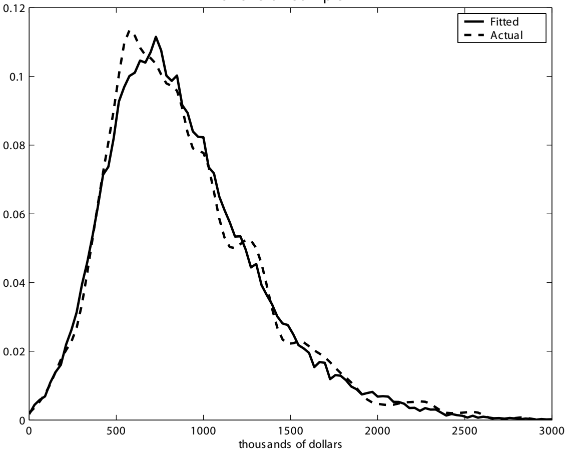
$$\theta_k \backsim \sum_{j=1}^{J_k} p_{k,j} \phi\left(f_k \mid \mu_{k,j}, \tau_{k,j}\right)$$

where $\phi\left(\eta \mid \mu_j, \tau_j\right)$ is a normal density for η with mean μ_j and variance τ_j.

Results

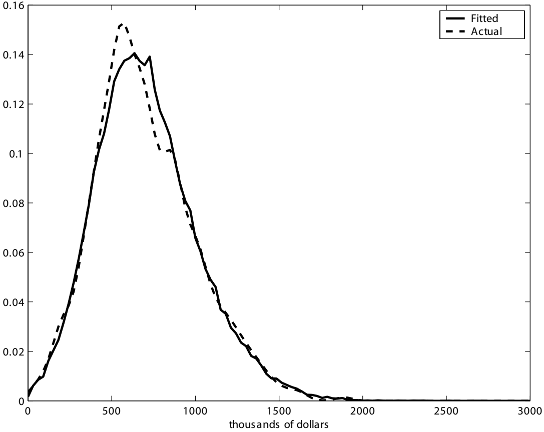
How the Model Fits the Data

Figure 1
Densities of fitted and actual present values of earnings
for overall sample



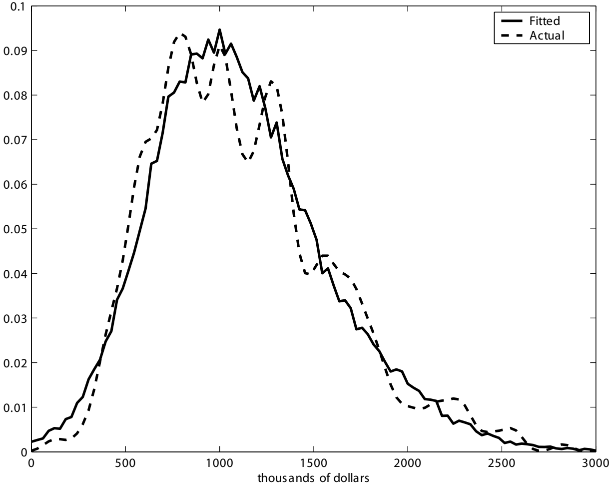
Present value of earnings from age 17 to 65 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let $S = 0$ denote high school sector, and $S = 1$ denote college sector. Define observed earnings as $Y = SY_1 + (1-S)Y_0$. Let $f(y)$ denote the density function of observed earnings. Here we plot the density functions f generated from the data (the dashed line), against that fitted by the model (the solid curve). We use kernel density estimation to produce these functions.

Figure 2
Densities of fitted and actual present values of earnings for people who choose to graduate high school



Present value of earnings from age 17 to 65 discounted using an interest rate of 3%. Let (Y_0) denote the potential outcome in the high school sector. Let S = 0 denote choice of the high school sector. Let f(y | S=0) denote the density function of observed earnings conditioned on agents that are high school graduates. Here we plot the density functions f(y | S=0) generated from the data (the dashed line), against that fitted by the model (the solid curve). We use kernel density estimation to produce these functions.

Figure 3
Densities of fitted and actual present values of earnings
for people who choose to graduate college



Present value of earnings from age 17 to 65 discounted using an interest rate of 3%. Let (Y_1) denote the potential outcome in the college sector. Let S=1 denote choice of the college sector. Let f(y | S=1) denote the density function of observed earnings conditioned on agents that are college graduates. Here we plot the density functions f(y | S=1) generated from the data (the dashed line), against that fitted by the model (the solid curve). We use kernel density estimation to produce these functions.

Table 2

Goodness of Fit Test for Lifetime Earnings

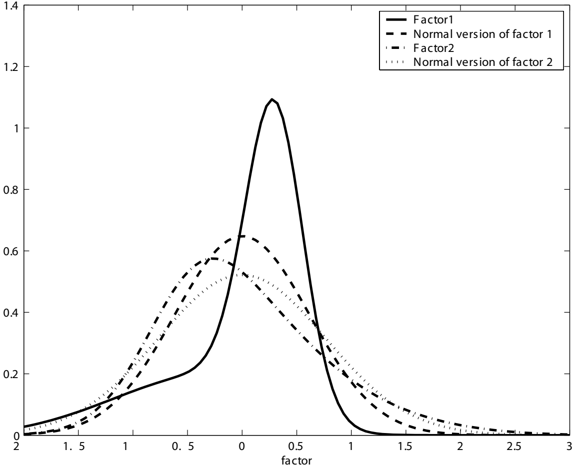
	χ^2 Statistic	Critical Value*
Overall	48.9251	53.1419
High School	25.4820	26.0566
College	32.2506	33.2562

*95% confidence, equiprobable bins with approximately 23 people per bin

The Factors: Non-normality and Evidence on Selection

Figure 4

Densities of factors and their normal equivalents



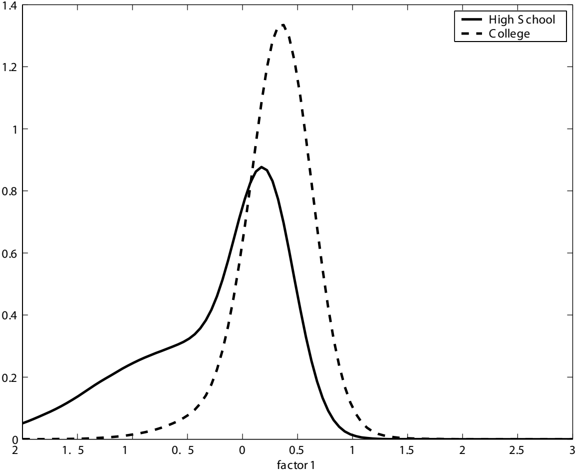
Let $f(\theta_1)$ denote the density function of factor θ_1.

We assume that $f(\theta_1)$ is a mixture of normals. Assume $\mu_1 = E(\theta_1)$, $\sigma_1 = Var(\theta_1)$.

Let $\phi(\mu_1, \sigma_1)$ denote the density of a normal random variable with mean μ_1 and variance σ_1. The solid curve is the estimated density of factor θ_1, $f(\theta_1)$, while the dashed curve is the density of a normal random variable with mean and variance of factor θ_1, $\phi(\mu_1, \sigma_1)$. We proceed similarly for factor 2, where the fitted density is plotted with dots and dashes and the normal version is plotted with dots.

- Figure 5 plots the density of factor 1 conditional on educational choices.

Figure 5
Densities of "ability" (factor 1) by schooling level



Let $f(\theta_1)$ denote the density function of factor θ_1.
We assume that $f(\theta_1)$ is a mixture of normals.
The solid line is the estimated density of factor 1 conditional on choosing the high school sector, that is, $f(\theta_1 \mid \text{Choice} = \text{High School})$. The dashed line plots the density of factor 1 conditional on choosing the college sector, that is, $f(\theta_1 \mid \text{Choice} = \text{College})$.

Estimating Joint Distributions of Counterfactuals: Returns, Costs and Ability as Determinants of Schooling

- Table 3 presents the conditional distribution of *ex post* potential college earnings given *ex post* potential high school earnings decile by decile.

Table 3

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

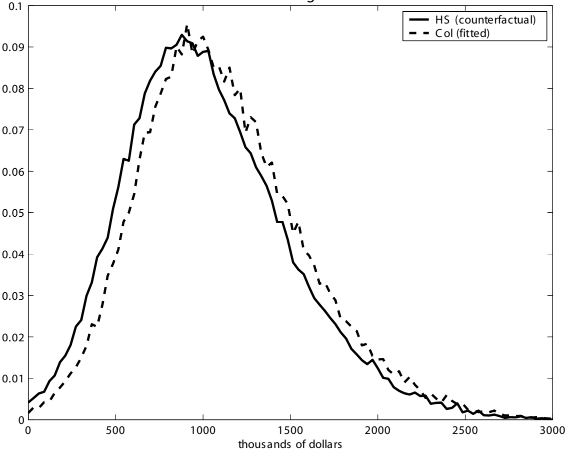
$\Pr(d_i < Y_c < d_i + 1 \mid d_j < Y_h < d_j + 1)$*

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

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

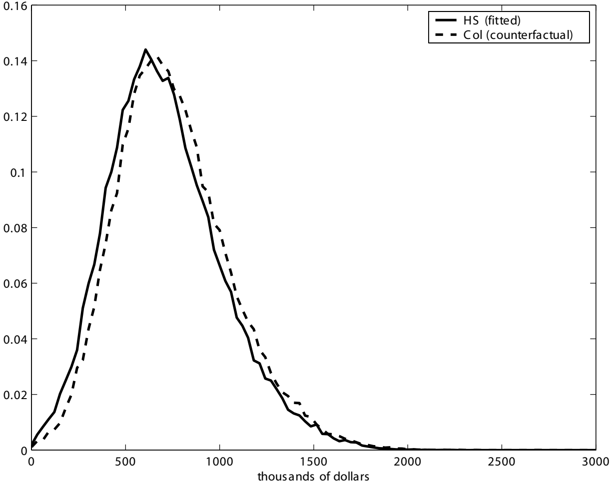
- Figures 6 and 7 present the marginal densities of predicted and counterfactual earnings for college (figure 6) and high school (figure 7).

Figure 6
Densities of present value of earnings
in the college sector



Let Y_1 denote present value of earnings (discounted at a 3% interest rate) in the college sector. Let $f(y_1)$ denote its density function. The dashed line plots the fitted Y_1 density conditioned on choosing college, that is, $f(y_1 \mid S = 1)$, while the solid line shows the estimated counterfactual density function of Y_1 for those agents who are actually high school graduates, that is, $f(y_1 \mid S = 0)$.

Figure 7
Densities of present value of earnings
in the high school sector



Let Y_0 denote present value of earnings (discounted at a 3% interest rate) in the high school sector. Let $f(y_0)$ denote its density function. The solid curve plots the fitted Y_0 density conditioned on choosing high school, that is, $f(y_0 \mid S = 0)$, while the dashed line shows the counterfactual density function of Y_0 for those agents who are actually college graduates, that is, $f(y_0 \mid S = 1)$.

- Tables 4 and 5 provide further evidence against the hypothesis of perfect dependence across counterfactual distributions.

Table 4
Average present value of earnings[1] for high school graduates
Fitted and Counterfactual
White males from NLSY79

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

Average returns[3] for high school graduates

	High School vs Some College
High School	0.4600
Std. Err.	0.1401

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

[2]It defines the result of taking a person at random from the population regardless of his schooling choice.

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

Table 5
Average present value of earnings[1] for college graduates
Fitted and Counterfactual
White males from NLSY79

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

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

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

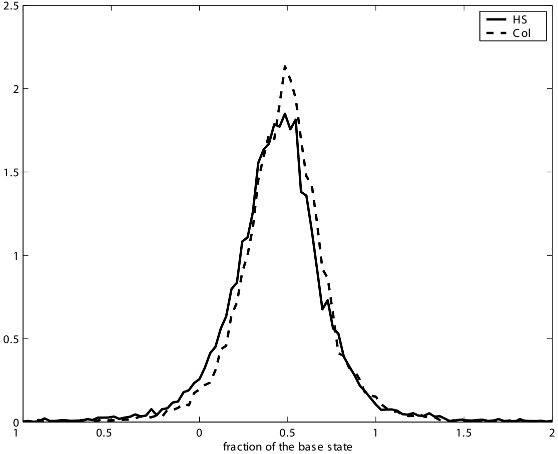
[2]It defines the result of taking a person at random from the population regardless of his schooling choice.

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

- Figure 8 plots the density of returns to education for agents who are high school graduates (the solid curve), and the density of returns to education for agents who are college graduates (the dashed curve).

Figure 8

Densities of ex post returns to college by schooling level chosen



Let Y_0, Y_1 denote the present value of earnings in high school and college sectors, respectively. Define ex post returns to college as the ratio $R = (Y_1 - Y_0)/Y_0$. 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, $f(r \mid S=0)$. The dashed line is the density of ex post returns to college for college graduates, that is, $f(r \mid S=1)$.

- Table 6 reveals that the average individual who is just indifferent between a college education and a high school diploma earns \$743.40 thousand dollars as a high school graduate or \$1,089.97 thousand dollars as a college graduate.

Table 6

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

Conditional on education level

White males from NLSY79

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

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

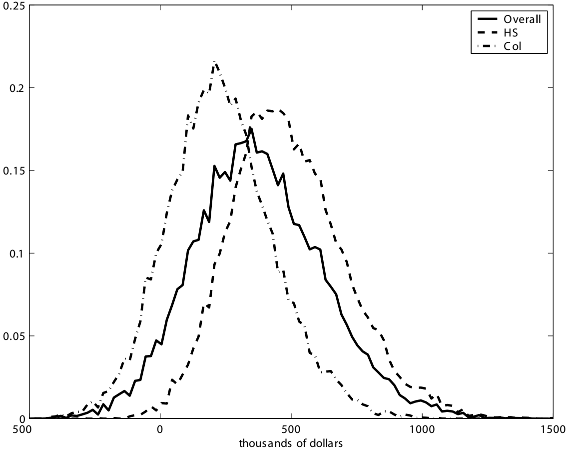
	High School vs Some College
Average	0.4800
Std. Err.	0.0853

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

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

- Figure 9 shows the estimated density of the monetary value of this cost both overall and by schooling level.
- Table 7 explores this point in more detail by presenting the mean total cost of attending college (first rows) and the mean cost that is due to ability (*i.e.,* factor 1), given in the second rows.

Figure 9
Density of monetary value of psychic cost
both overall and by schooling level



In this figure we plot the monetary value of psychic costs. Let C denote the monetary value of psychic costs.

The monetary value of psychic costs is given by:

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

The contribution of ability to the costs of attending college, in monetary value is $\theta_1 \alpha_{C1}$.

Table 7

Mean monetary value of total cost of attending college

High School	College	Overall
488.24	232.56	375.27

Mean monetary value of cost of attending college due to ability

High School	College	Overall
40.97	-51.27	0.0

Let C denote the monetary value of psychic costs. Then C is given by:

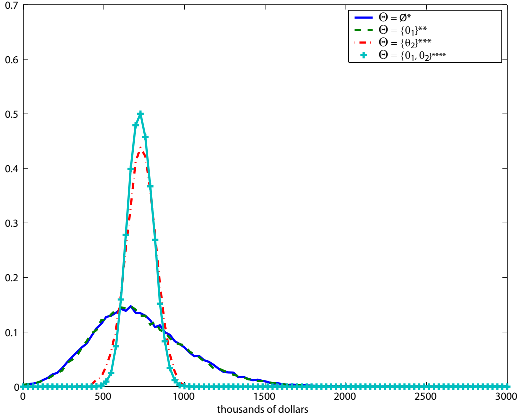
$$C = Z\gamma + \theta_1\alpha_{C1} + \theta_2\alpha_{C2} + \varepsilon_C$$

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

Mobility and Heterogeneity versus Uncertainty

- In figures 10 through 12, we separate the effect of heterogeneity from uncertainty in earnings.
- The information set of the agent is $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$ where Θ contains some or all of the factors.
- Focusing on figure 10 we start by assuming that the agents do not know their factors; consequently, $\Theta = \varnothing$.

Figure 10
Densities of agent's forecast of the present value of high school earnings under different information sets: $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$



Let Y_0 denote the agent's forecast of present value of earnings in the high school sector . These are formed over the whole population, not just the subpopulation who go to high school. We assume that agents know all coefficients. Let $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$ denote the agents information set. Let $f(y_0|\mathcal{I})$ denote the density of the agent's forecast of present value of earnings in high school conditioned on the information set $\mathcal{I}$. Then:

*Plot of $f(y_0|\mathcal{I})$ under no element of θ in the information set, i.e., $\Theta = \varnothing$.

** Plot of $f(y_0|\mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.

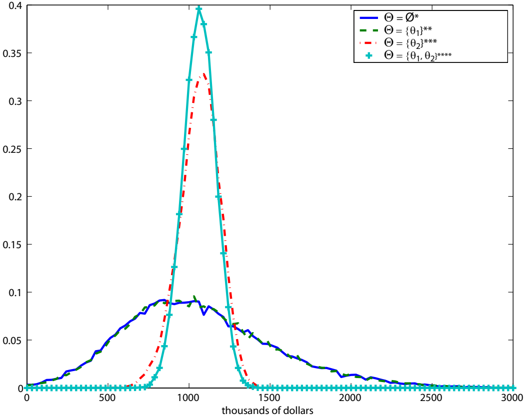
*** Plot of $f(y_0|\mathcal{I})$ when only factor 2 is in the information set, i.e., $\Theta = \{\theta_2\}$.

**** Plot of $f(y_0|\mathcal{I})$ when both factors are in the information set, i.e., $\Theta = \{\theta_1, \theta_2\}$.

- Figure 11 reveals much the same story about college earnings.

Figure 11

Densities of agent's forecast of the present value of college earnings under different information sets: $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$



Let Y_1 denote the agent's forecast of present value of earnings in the college sector . These are formed over the whole population, not just the subpopulation who go to college. We assume that agents know all coefficients. Let $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$ denote the agents information set. Let $f(y_1 | \mathcal{I})$ denote the density of the agent's forecast of present value of earnings in college conditioned on the information set $\mathcal{I}$. Then:

*Plot of $f(y_1 | \mathcal{I})$ under no element of θ in the information set, i.e., $\Theta = \emptyset$.

** Plot of $f(y_1 | \mathcal{I})$ when only factor 1 is in the information set, i.e.,$\Theta = \{\theta_1\}$.

*** Plot of $f(y_1 | \mathcal{I})$ when only factor 2 is in the information set, i.e.,$\Theta = \{\theta_2\}$.

**** Plot of $f(y_1 | \mathcal{I})$ when both factors are in the information set, i.e.,$\Theta = \{\theta_1, \theta_2\}$.

Table 8
Agent's Forecast Variance of Present Value of Earnings
Under Different Information Sets: $I = \{ X, Z, X_T, \varepsilon_C, \Theta \}$
(as a fraction of the variance when no information is available)

	$Var(Y_c)$	$Var(Y_h)$	$Var(Y_c - Y_h)$	$Cov(Y_c , Y_h)$
For time period 1:[+]				
Variance when $\Theta = \emptyset$	7167.20	5090.46	3073.94	4591.86
Percentage of variance remaining after controlling for the indicated factor:				
$\Theta = \{\theta_1\}$	97.50%	98.34%	99.43%	97.33%
$\Theta = \{\theta_2\}$	18.50%	32.83%	89.52%	2.67%
$\Theta = \{\theta_1, \theta_2\}$	16.01%	31.17%	88.94%	0.00%
For time period 2:[++]				
Variance when $\Theta = \emptyset$	49690.64	167786.87	41137.80	88169.85
Percentage of variance remaining after controlling for the indicated factor:				
$\Theta = \{\theta_1\}$	97.18%	97.54%	98.25%	97.28%
$\Theta = \{\theta_2\}$	7.39%	4.73%	16.55%	2.72%
$\Theta = \{\theta_1, \theta_2\}$	4.57%	2.27%	14.80%	0.00%
For lifetime:[+++]				
Variance when $\Theta = \emptyset$	56857.84	172877.33	44211.74	92761.72
Percentage of variance remaining after controlling for the indicated factor:				
$\Theta = \{\theta_1\}$	97.22%	97.57%	98.33%	97.28%
$\Theta = \{\theta_2\}$	8.79%	5.56%	21.62%	2.72%
$\Theta = \{\theta_1, \theta_2\}$	6.01%	3.13%	19.95%	0.00%

We use an interest rate of 3% to calculate the present value of earnings. In all cases, the information set of the agent is $I = \{ X, Z, X_T, \varepsilon_C, \Theta \}$ and we change the contents of Θ.

[+]Variance of the unpredictable component of earnings between age 17 and 28 as predicted at age 17.

[++]Variance of the unpredictable component of earnings between age 29 and 65 as predicted at age 17.

[+++]Variance of the unpredictable component of earnings between age 17 and 65 as predicted at age 17.

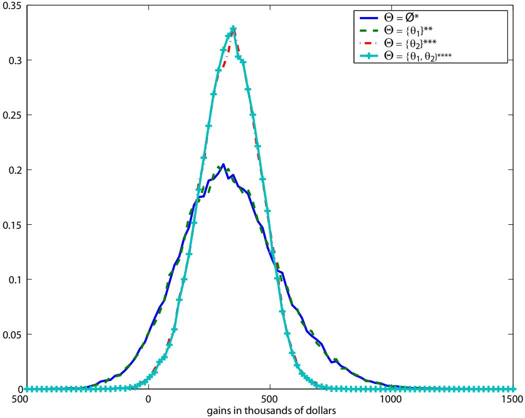
So we would say that the variance of the unpredictable component of period 1 college earnings when using factor 1 in the prediction is 97.5% of the variance when no information is available (*i.e.*, 0.975*7167.2).

- Figure 12 presents an exercise for returns to college ($Y_1 - Y_0$) similar to that presented in figures 10 and 11 regarding information sets available to the agent.

Figure 12

Densities of agent's forecast gains in present value of earnings

(Y_1 -Y_0) under different information sets: $\mathcal{I} = \{ X, Z, X_T, \varepsilon_C, \Theta \}$



Let Y_0, Y_1 denote the agent's forecast of present value of earnings in the high school and college sectors, respectively. We define the difference in present value of earnings as $\Delta = Y_1 - Y_0$. We assume that agents know all coefficients. Let $\mathcal{I} = \{X, Z, X_T, \varepsilon_C, \Theta\}$, $f(\Delta \mid \mathcal{I})$ denote the agents information set and the density of the agent's forecast of gains in present value of earnings in choosing college conditioned on the information set $\mathcal{I}$, respectively. These are defined over the entire population, then:

*Plot of $f(\Delta \mid \mathcal{I})$ under no element of θ in the information set, i.e., $\Theta = \emptyset$.

** Plot of $f(\Delta \mid \mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.

*** Plot of $f(\Delta \mid \mathcal{I})$ when only factor 2 is in the information set, i.e., $\Theta = \{\theta_2\}$.

**** Plot of $f(\Delta \mid \mathcal{I})$ when both factors are in the information set, i.e., $\Theta = \{\theta_1, \theta_2\}$.

Ex Ante versus Ex Post

- Once the distinction between heterogeneity and uncertainty is made, we can talk about the distinction between *ex ante* and *ex post* decision making.

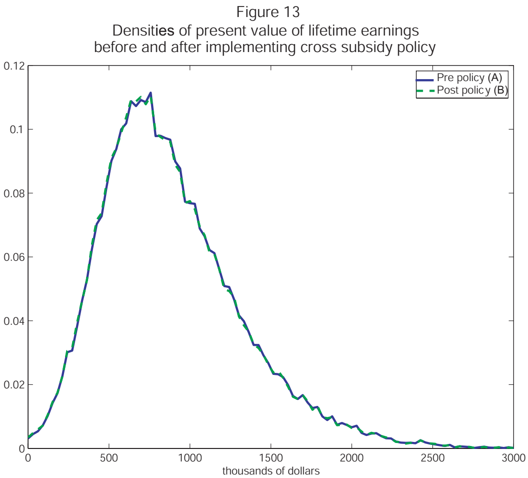
$$
\begin{aligned}
V &= Y_{c,1} + \frac{Y_{c,2}}{1+r} - Y_{h,1} - \frac{Y_{h,2}}{1+r} - C > 0 \\
S &= 1 \text{ if } V > 0;\ S = 0 \text{ otherwise,}
\end{aligned}
$$

- In our empirical model, if individuals could pick their schooling level using their *ex post* information (*i.e.*, after learning their luck components in earnings) 13.81% of high school graduates would rather be college graduates and 17.15% of college graduates would have stopped their schooling at the high school level.

Analyzing a Cohort Specific Cross-Subsidized Tuition Policy: Constructing Joint Distributions of Counterfactuals Across Policy Regimes

- Total tuition raised covers the cost K of educating each student.
- Thus if there are N_P poor students and N_R rich students, total costs are $(N_P + N_R)\, K$. In the proposed policy, the poor pay nothing.
- So each rich person is charged a tuition $T = (K)\left(1 + \frac{N_P}{N_R}\right)$.
- To determine T, notice that $N_P = N_P(T)$; $N_R = N_R(T)$.
- We iterate to find the unique self financing T.
- Notice that $N_P\,(T)$, the number of poor people who attend college when tuition is zero, is the same for all values of T $(N_P\,(T) = N_P\,(0)$ for all $T)$. N_R is sensitive to the tuition level charged.

- Figure 13 shows that the marginal distributions of income in both the prepolicy state and the postpolicy state are essentially identical.
- Under anonymity we would judge these two situations as equally good using Lorenz measures or second order stochastic dominance.
- We move beyond anonymity and analyze the effect that the policy has on what Fields (2003) calls "positional" mobility.



Let Y^A, Y^B denote the observed present value of earnings pre and post policy, respectively. Define $f(y^A)$, $g(y^B)$ as the marginal densities of present value of earnings pre and post policy. In this figure we plot $f(y^A)$, $g(y^B)$.

- Panel 1 of table 9 presents this analysis by describing how the 9.2% of the people who are affected by the policy move between deciles of the distribution of income.
- Moving beyond the anonymity postulate (which instructs us to examine only marginal distributions), we learn much more about the effects of the policy on different groups.

Table 9

Mobility of People Affected by Cross Subsidizing Tuition

Overall: Fraction of Total Population who Switch Schooling Levels: 0.0932											
		Probability of Moving to a Different Decile of the Lifetime Earnings Distribution									
Fraction by Decile of Origin	Deciles of Origin	1	2	3	4	5	6	7	8	9	10
0.0728	1	0.5565	0.2011	0.1220	0.0634	0.0283	0.0074	0.0012	0.0000	0.0000	0.0000
0.0867	2	0.2079	0.1712	0.1715	0.1690	0.1585	0.0870	0.0322	0.0025	0.0002	0.0000
0.0955	3	0.1148	0.1489	0.0935	0.1137	0.1573	0.1888	0.1387	0.0409	0.0034	0.0000
0.0998	4	0.0619	0.1557	0.0910	0.0534	0.0764	0.1615	0.2084	0.1557	0.0360	0.0000
0.1032	5	0.0296	0.1495	0.1387	0.0630	0.0304	0.0571	0.1411	0.2456	0.1396	0.0055
0.1050	6	0.0066	0.0959	0.1726	0.1471	0.0520	0.0142	0.0415	0.1671	0.2605	0.0425
0.1084	7	0.0006	0.0336	0.1411	0.1956	0.1269	0.0420	0.0082	0.0348	0.2346	0.1827
0.1089	8	0.0000	0.0046	0.0519	0.1765	0.2211	0.1495	0.0388	0.0034	0.0513	0.3029
0.1101	9	0.0000	0.0000	0.0055	0.0421	0.1570	0.2733	0.2302	0.0447	0.0014	0.2459
0.1069	10	0.0000	0.0000	0.0000	0.0002	0.0041	0.0517	0.2082	0.3242	0.2490	0.1626

Table 9

Mobility of People Affected by Cross Subsidizing Tuition

High school: Fraction of Total Population who Switch from High School to College due to the policy: 0.0450											
0.1012	1	0.3954	0.2557	0.1775	0.0936	0.0417	0.0110	0.0018	0.0000	0.0000	0.0000
0.1279	2	0.0382	0.1220	0.2176	0.2325	0.2200	0.1210	0.0448	0.0035	0.0003	0.0000
0.1369	3	0.0023	0.0188	0.0692	0.1536	0.2244	0.2701	0.1984	0.0584	0.0049	0.0000
0.1367	4	0.0000	0.0016	0.0088	0.0368	0.1116	0.2417	0.3123	0.2332	0.0540	0.0000
0.1285	5	0.0000	0.0000	0.0007	0.0052	0.0277	0.0903	0.2324	0.4047	0.2300	0.0090
0.1122	6	0.0000	0.0000	0.0000	0.0004	0.0024	0.0151	0.0792	0.3209	0.5004	0.0816
0.1017	7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0101	0.0761	0.5133	0.3997
0.0797	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0067	0.1440	0.8493
0.0557	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0032	0.9968
0.0173	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 9

Mobility of People Affected by Cross Subsidizing Tuition

College: Fraction of Total Population who Switch from College to High School due to the policy: 0.0473											
0.0459	1	0.8941	0.0866	0.0055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0475	2	0.6423	0.2972	0.0534	0.0062	0.0009	0.0000	0.0000	0.0000	0.0000	0.0000
0.0560	3	0.3763	0.4510	0.1501	0.0211	0.0015	0.0000	0.0000	0.0000	0.0000	0.0000
0.0647	4	0.1860	0.4648	0.2559	0.0868	0.0059	0.0007	0.0000	0.0000	0.0000	0.0000
0.0791	5	0.0753	0.3801	0.3518	0.1522	0.0347	0.0059	0.0000	0.0000	0.0000	0.0000
0.0982	6	0.0138	0.2001	0.3602	0.3064	0.1059	0.0133	0.0004	0.0000	0.0000	0.0000
0.1148	7	0.0011	0.0618	0.2598	0.3603	0.2337	0.0766	0.0066	0.0000	0.0000	0.0000
0.1366	8	0.0000	0.0071	0.0807	0.2744	0.3436	0.2323	0.0603	0.0015	0.0000	0.0000
0.1618	9	0.0000	0.0000	0.0073	0.0559	0.2084	0.3628	0.3056	0.0593	0.0008	0.0000
0.1920	10	0.0000	0.0000	0.0000	0.0002	0.0044	0.0561	0.2260	0.3519	0.2702	0.0911

Note: Cross subsidy consists in making tuition zero for people with family income below average and making the budget balance by raising tuition for college students with family income above the average. For example, we read from the first panel row 1, column 1 that 7.28% of the people who switch schooling levels come from the lowest decile. Out of those, 55% are still in the first decile after the policy while 2.83% jump to the fifth decile. Panel 2 has the same interpretation but it only looks at people who switch from high school to college while panel 3 looks at individuals who switch from college to high school.

- Table 10 and panels 2 and 3 of table 9 reveal that not only 9.2% of the population is affected by the policy but that actually about half of them moved from high school into college (4.5% of the population) and half moved from college into high school (4.7% percent of the population).

Table 10
Mobility of people affected by cross subsidizing tuition
Fraction of the total population who switch schooling levels: 0.0932

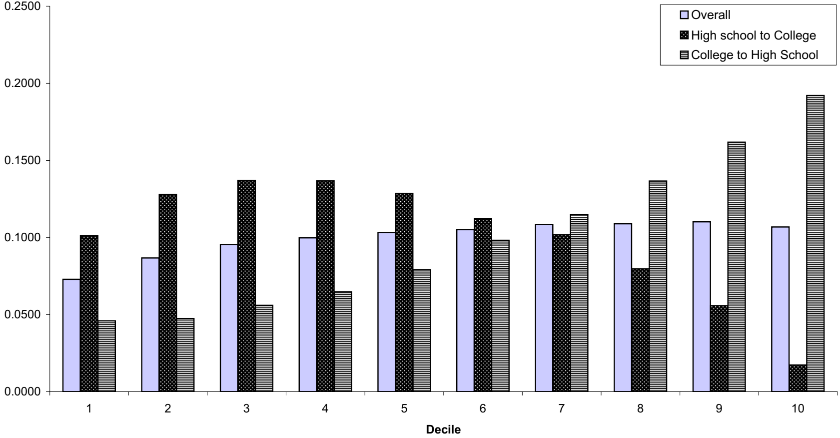
Pre-policy Choice:	Fraction of High School Graduates:	
	Do not switch	Become College graduates
High School	0.9197	0.0803
	Fraction of College Graduates:	
	Do not switch	Become High School graduates
College	0.8923	0.1077

Note: Cross subsidy consists in making tuition zero for people with family income below average and making the budget balance by raising tuition for college students with family income above the average.

- This translates into saying that, of those affected by the policy, 92% of the high school graduates stay in high school in the post-policy regime while only 89% of college graduates stay put.
- Thus the policy is slightly biased against college attendance.
- We can form the joint distributions of lifetime earnings by initial schooling level.

- Figure 14 summarizes some of the evidence presented in table 10.
- The figure 14 the panels 2 and 3 of table 9 show that the policy affects very few high school graduates at the top end of the income distribution (only 1.7% of those affected come from the 10^{th} percentile) and a lot of college graduates in the same situation (19% of college graduates affected come from the top decile).
- Table 11 shows where in the prepolicy distribution of high school earnings persons induced to go to college come from and where in the postpolicy distribution of college earnings they go to.

Figure 14
Fraction of people who switch schooling levels when tuition is cross subsidized by decile of origin from the lifetime earnings distribution



*Cross subsidy consists in making tuition zero for people with family income below average and making the budget balance by raising tuition for college students with family income above the average.

Table 11

Mobility of People Affected by Cross Subsidizing Tuition Across Counterfactual Distributions

Highschool: Fraction of Total Population who from High School to College after the policy: 0.0450

Fraction by Decile of Origin in the Prepolicy High School Distribution	Deciles of Origin	Probability of Moving to a Different Decile of the Post Policy College Lifetime Earnings Distribution									
		1	2	3	4	5	6	7	8	9	10
0.0667	1	0.8266	0.1227	0.0140	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0811	2	0.4044	0.4110	0.1490	0.0296	0.0055	0.0000	0.0000	0.0000	0.0000	0.0000
0.0908	3	0.1488	0.3544	0.3059	0.1419	0.0445	0.0039	0.0005	0.0000	0.0000	0.0000
0.0998	4	0.0401	0.2343	0.3096	0.2490	0.1234	0.0379	0.0053	0.0004	0.0000	0.0000
0.1047	5	0.0089	0.0713	0.2081	0.3053	0.2348	0.1282	0.0365	0.0068	0.0000	0.0000
0.1058	6	0.0004	0.0202	0.0950	0.2155	0.2761	0.2416	0.1273	0.0239	0.0000	0.0000
0.1062	7	0.0000	0.0033	0.0243	0.0896	0.1888	0.3026	0.2662	0.1155	0.0096	0.0000
0.1116	8	0.0000	0.0004	0.0016	0.0159	0.0630	0.1690	0.3220	0.3228	0.1024	0.0028
0.1138	9	0.0000	0.0000	0.0000	0.0016	0.0043	0.0293	0.1227	0.3271	0.4568	0.0582
0.1173	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0027	0.0333	0.2626	0.7014

Table 11

Mobility of People Affected by Cross Subsidizing Tuition Across Counterfactual Distributions

College: Fraction of Total Population who Switch from College to High School due to the policy: 0.0473

0.1095	1	0.5473	0.2945	0.1135	0.0316	0.0062	0.0012	0.0000	0.0000	0.0000	0.0000
0.1056	2	0.1076	0.3257	0.2937	0.1789	0.0716	0.0204	0.0016	0.0004	0.0000	0.0000
0.1035	3	0.0180	0.1473	0.2776	0.2657	0.1833	0.0857	0.0200	0.0024	0.0000	0.0000
0.1013	4	0.0004	0.0355	0.1535	0.2349	0.2866	0.1890	0.0847	0.0150	0.0004	0.0000
0.1012	5	0.0000	0.0050	0.0467	0.1503	0.2654	0.2705	0.1903	0.0668	0.0050	0.0000
0.0979	6	0.0000	0.0000	0.0091	0.0513	0.1678	0.2683	0.2972	0.1786	0.0276	0.0000
0.0977	7	0.0000	0.0000	0.0000	0.0087	0.0463	0.1609	0.3071	0.3387	0.1362	0.0022
0.0953	8	0.0000	0.0000	0.0000	0.0004	0.0044	0.0430	0.1560	0.4020	0.3617	0.0324
0.0964	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0127	0.1337	0.5355	0.3173
0.0882	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0034	0.0915	0.9051

Note: Cross subsidy consists in making tuition zero for people with family income below average and making the budget balance by raising tuition for college students with family income above the average. For example, we read from the first panel row 1, column 1 that 6.67% of the people who switch from high school to college come from the lowest decile of the prepolicy high school distribution. Out of those, 82.66% are still in the first decile of the post policy college earnings distribution after the policy is implemented while 1.40% "jump" to the third decile. Panel 2 has the same interpretation but it only looks at people who switch from college to high school.

- Most people stay in their decile or move closely to adjacent ones.
- An advantage of our method is that it allows us to calculate the effect that the policy has on welfare.
- Table 12 shows the result of such an exercise.

Table 12

Voting outcome of proposing cross subsidizing* tuition

Fraction of the total population who switch schooling levels: 0.0932

Average pre-policy lifetime earnings**	920.55
Average post-policy lifetime earnings**	905.96
Fraction of the population who votes	
Yes	0.0716
No	0.6152
Indifferent	0.3132

*Cross subsidy consists in making tuition zero for people with family income below average and making the budget balance by raising tuition for college students with family income above the average.
** In thousands of dollars.