NBER WORKING PAPER SERIES

A NEW FRAMEWORK FOR THE ANALYSIS OF INEQUALITY

Flavio Cunha
James Heckman

Working Paper 12505
http://www.nber.org/papers/w12505

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2006

This research was supported by NIH R01-HD32058-3, NSF SES-024158, and a grant from the American Bar Foundation. A website for supplementary material is available at http://jenni.uchicago.edu/frame-ineq. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

A New Framework for the Analysis of Inequality
Flavio Cunha and James Heckman
NBER Working Paper No. 12505
August 2006
JEL No. D80, I30

ABSTRACT

This paper presents a new framework for analyzing inequality that moves beyond the anonymity postulate. We estimate the determinants of sectoral choice and the joint distributions of outcomes across sectors. We determine which components of realized earnings variability are due to uncertainty and which components are due to components of human diversity that are forcastable by agents. Using our tools, we can determine how policies shift persons across sectors and outcome distributions across sectors.

Flavio Cunha
The University of Chicago
flavio@uchicago.edu

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
jjh@uchicago.edu

1 Introduction

Most studies of income inequality and social mobility are descriptive in nature. Studies of inequality compare differences in the location in the overall distribution of income among groups at a point in time and over time and the evolution of income distributions within groups over time. Studies of mobility measure movements of income within lifetimes or across generations.[1] These exercises present factual summaries of income inequality and income mobility.

For the purposes of policy analysis, and for interpreting facts within a scientific model, it is necessary to move beyond factual description to construct counterfactuals. They can be used to determine what would happen to mobility or inequality if different policies or interventions were tried than the policies historically observed. They can also be used to decompose observed inequality and mobility into components due to genuine uncertainty ("luck" as described by Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns, and Michelson, 1972) and components of heterogeneity and individual differences that are predictable, at least by a certain age or stage of the life cycle.

This paper describes recent methodological advances that enable analysts to construct counterfactual distributions and separate heterogeneity (predictable variability across persons) from uncertainty.[2] In general, the welfare consequences of predictable heterogeneity and unpredictable uncertainty are not the same. Using the tools reviewed here, analysts can determine how much inequality and mobility is forecastable at a given age and how much is due to unforecastable luck.

These methods allow analysts to move beyond aggregate summary measures of inequality that are based on the anonymity postulate to determine which groups in an initial distribution are affected by a policy change and how they are affected. The anonymity postulate treats two aggregate distributions as equally good if, after income is redistributed among persons, the overall distribution is the same. With our methods we can determine, for reforms that are contemplated but have never been implemented, which groups in an initial position benefit or lose, how much they lose, how they would vote in advance of a reform and how they would vote after it is implemented, once the *ex ante* uncertainty surrounding the outcomes of the reform is resolved.

[1] See Fields (2003) for a comparison of inequality and mobility measures.

[2] We draw on results reported in Aakvik, Heckman, and Vytlacil (1999, 2005); Carneiro, Hansen, and Heckman (2003); Cunha, Heckman, and Navarro (2005); Heckman (1992); Heckman, Smith, and Clements (1997), and Heckman, Lochner, and Todd (2006).

We can 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. Thus we can move beyond traditional inequality and social mobility analyses 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. Conventional studies of inequality consider movements of persons among observed (measured) states (see, e.g., Ravallion, 2003).

The plan of the rest of the paper is as follows. In section 2, we present a choice-theoretic framework for constructing counterfactuals and we consider limitations of our approach. Section 3 presents our method for constructing distributions of counterfactuals based on factor models, extending methods developed by Jöreskog and Goldberger (1975) and Jöreskog (1977) to consider the construction of counterfactuals. We illustrate results on identification derived in a number of previous papers by focusing on a simple parametric version of the model. Section 4 shows how we can use information about choices and subsequent realizations to infer how much agents know about future earnings when making their choices about college. This section reviews a method for estimating "luck" and separating it from predictable heterogeneity that is developed in Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005). Section 6 concludes.

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

Social programs such as job training and college tuition subsidies are central features of the modern welfare state. Because different parties may gain or lose from such programs, there is a demand for knowledge about the redistributive effects of government policies. The central problem in the literature on the evaluation of social policies is the construction of counterfactuals.

In the simplest form of the evaluation problem, there are two possible choices: agents may receive a treatment or not. We denote by $S = 1$ the agents who receive the treatment and by $S = 0$ those who don't. To fix ideas, consider the case in which $S = 0$ denotes a worker who chooses to be a high school graduate, and $S = 1$ a worker who chooses to be a college graduate. The treatment is college education.

Associated with each level of education is a potential outcome. Let (Y_0, Y_1) denote potential outcomes in schooling level $S = 0$ and $S = 1$, respectively. The outcomes Y_0 and Y_1 can be expressed in monetary or utility units, but in the discussion that follows we shall think of the rewards as earnings in each sector expressed in dollars. Each person has a (Y_0, Y_1) pair. The gain for an individual who moves from $S = 0$ to $S = 1$ is Δ, where $\Delta \equiv Y_1 - Y_0$.

If Y_1 and Y_0 could be observed for each individual at the same time, the gain Δ would be known for each person. An evaluation problem arises because we do not observe the pair (Y_0, Y_1) for anybody. This is a missing data problem: in calculating the gains to attending college for a particular individual who chooses to be a college graduate, we observe her college earnings (Y_1), but not her high school earnings (Y_0). We can solve this missing data problem by constructing counterfactuals: how much a college graduate would earn if she had chosen to be a high school graduate. To identify these counterfactuals, different approaches in the literature of program evaluation make different assumptions about how the missing data are related to the available data, and what data are available.

One possible approach is to model the decision making of the individuals. The choices made by agents depend on their perception of outcomes Y_0 and Y_1. The choices may also be influenced by the costs associated with schooling, C. These costs C may arise because of pecuniary (e.g., tuition), or nonpecuniary reasons (e.g, heterogeneity in preferences for schooling), but if we assume that costs C can be expressed in monetary units, it is straightforward to define the net utility I as:

$$I = Y_1 - Y_0 - C \tag{1}$$

If agents maximize their utility index I, they attend college if, and only if, they gain from it. This is represented by the following decision rule:

$$S = 1 \;\text{ if }\; I \geq 0;\; S = 0 \text{ otherwise.} \tag{2}$$

It is possible to consider alternative decision rules by embedding this model into a dynamic environment, analyzing different credit market structures, uncertainty, risk aversion, and so on. (See e.g. Heckman and Navarro, 2006). However, for expositional ease we focus on the simple decision rule generated by (1) and

(2) since it is sufficiently rich to serve our purposes.[3]

Assume that (Y_0, Y_1) have finite means and can be expressed in terms of explanatory variables X in the following manner:

$$Y_0 = X\beta_0 + U_0, \tag{3}$$

$$Y_1 = X\beta_1 + U_1, \tag{4}$$

where $E(Y_0 \mid X) = \mu_0(X)$, $E(Y_1 \mid X) = \mu_1(X)$ and $E(U_0 \mid X) = E(U_1 \mid X) = 0$.

Further, assume that costs C can be expressed in terms of explanatory variables Z and unobservables U_C:

$$C = Z\gamma + U_C. \tag{5}$$

The model described by equations (1), (2), (3), (4), and (5) is the Generalized Roy model. If costs are identically zero for all agents, so that $C = \mu_C(Z) = U_C = 0$, equations (1), (2), (3), and (4) describe the simple Roy model as developed in Roy (1951).

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:

$$ATE(X) = E(\Delta | X) = E(Y_1 - Y_0 | X). \tag{6}$$

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

$$TT(X, Z) = E(\Delta \mid X, S = 1) = E(Y_1 - Y_0 \mid X, S = 1). \tag{7}$$

The evaluation of policies has traditionally focused on such mean effects. Nevertheless, the gains Δ are heterogenous across agents even after one controls for X. As a result, mean gains are not necessarily the most interesting parameter to be estimated. Consider an example from the economics of education: we want to evaluate a policy that reduces tuition by one thousand dollars. The marginal cost of such a policy is, therefore, constant. If gains are heterogenous across people, we need to know (a) how many

[3] See Heckman (2001) for a survey of this literature. These models for counterfactuals and potential outcomes are called Roy (1951) or Generalized Roy models and have an ancient lineage in econometrics. The decision-maker may be a parent, and the outcomes may be for a child.

entrants into education will be induced by the tuition policy, and (b) from where in the distribution of gains to schooling (Δ) the new entrants are coming. In this example, it would be optimal to keep offering tuition reductions up to the point that the marginal gain is equal to the marginal cost. If we compute (b), we can monitor the marginal gain as we vary the size of the policy. If we can compute (a) and (b), we can calculate the aggregate gains from the tuition policy. Consequently, if the gains Δ vary across agents, no single number summarizes the distribution of gains for all purposes of policy evaluation. For each specific policy question we want to address, we must carefully define the parameter of interest (see Heckman and Vytlacil, 2005, 2007a,b). In general, the average gain of those who are in the program is not the relevant parameter of interest. For other problems of distributional analysis, it is of interest to determine where in an initial distribution beneficiaries come from and where they end up in the treatment outcome distribution.

If we want to recognize the fact the gains are heterogenous across people we raise an econometric difficulty. Once we have defined the parameter of interest, say the gain to the marginal entrant, how can we estimate it? If we wish to avoid special assumptions like statistical independence between Y_0 and Y_1 or perfect dependence, the solution is to recover the joint distribution of (Y_0, Y_1). Once we know this distribution, it is possible to calculate the distribution of $(Y_1 - Y_0)$ for any group of people we are interested in and obtain its median or any other quantile.

In this paper, we report on recent research that solves the problem of constructing counterfactuals by identifying the joint distribution of (Y_1, Y_0) and the potential outcomes under some policy regime, conditional on S (or I), using a factor structure model. These models generalize the LISREL models of Jöreskog (1977) and the MIMIC model of Jöreskog and Goldberger (1975) to produce counterfactual distributions.

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 $\left(Y_0^A, Y_1^A\right)$ (see, e.g., Card,

1999) or evaluating schooling policies (e.g., Kane, 1994).[4]

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.[5]

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.

This paper analyzes two sets of counterfactuals: (a) $\left(Y_0^A, Y_1^A\right)$ 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. The tradition in the analysis of income inequality is to make comparisons across regimes *i.e.,* to compare the distribution of $Y^A = Y_1^A S^A + Y_0^A(1 - S^A)$ with the distribution of $Y^B = Y_1^B S^B + Y_0^B(1 - S^B)$. When both A and B are observed, such comparisons are straightforward if there are panel data on incomes of the same persons in both states.

With our methods, we can construct counterfactual distributions of $\left(Y_0^A, Y_1^A\right)$ and (Y_0^B, Y_1^B) 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 $Y_1^B S^B + Y_0^B(1 - S^B)$. We can also compare movements from Y_0^A to Y_1^B as well as other comparisons whether or not A and B are observed. This allows us to obtain a much richer understanding of the inequality and social mobility consequences of policy change than are available from inequality measures based on the anonymity axiom, and allows us to go more deeply than panel data analyses that compare movements from the distribution of Y^A to the distribution of Y^B. We can use our analysis to generate counterfactual states, never measured. We now present our methodology.

[4] Heckman, Lochner, and Taber (1998b,c, 1999) develop general equilibrium policy analysis computing distributional consequences of alternative policies. See also Bourguignon and da Silva (2003).

[5] Thus a large scale expansion of the educational system may depress the returns to schooling.

3 Identifying counterfactual distributions using factor models

The Roy model (1951) postulates that individuals choose the option that gives them the highest outcome. There are no costs associated with receipt or nonreceipt of treatment. The nonparametric identification of the joint distribution of outcomes is established in Heckman and Honoré (1990). Consider, for example, the case in which the error terms U_0, U_1 are normally distributed. In this case sectoral outcomes are

$$Y_s = X\beta_s + U_s, \quad s = 0, 1. \tag{8}$$

Under normality:

$$\left[\begin{array}{c} Y_0 \\ Y_1 \end{array}\right] \Bigg| X \sim N\left(\left[\begin{array}{c} X\beta_0 \\ X\beta_1 \end{array}\right], \left[\begin{array}{cc} Var\left(U_0\right) & Cov\left(U_0, U_1\right) \\ Cov\left(U_0, U_1\right) & Var\left(U_1\right) \end{array}\right] \right)$$

We observe Y_0 or Y_1 for each person but not the pair (Y_0, Y_1). The trick is to show that one can identify $Cov\left(Y_0, Y_1\right)$. If we observe I in equation (1) where $C = 0$, then we could compute the $Cov\left(Y_1, I\right)$ and, consequently, identify $Cov\left(Y_0, Y_1\right)$ because from the definition of I and a property of covariances it follows that:

$$Cov\left(Y_1, I\right) = Cov\left(Y_1, Y_1 - Y_0\right) = Var\left(Y_1\right) - Cov\left(Y_1, Y_0\right). \tag{9}$$

Obviously, we don't observe I directly, but we do observe choices, and we know that $S = 1 \Leftrightarrow I \geq 0$ and have I up to scale $\sigma_{Y_1 - Y_0} = \left[Var\left(Y_1 - Y_0\right)\right]^{1/2}$. Thus we can obtain the left hand sides of

$$\left[\frac{Cov\left(Y_1, I\right)}{\sigma_{Y_1 - Y_0}}\right] = \frac{Var\left(Y_1\right)}{\sigma_{Y_1 - Y_0}} - \frac{Cov\left(Y_1, Y_0\right)}{\sigma_{Y_1 - Y_0}}$$

and

$$\left[\frac{Cov\left(Y_0, I\right)}{\sigma_{Y_1 - Y_0}}\right] = \frac{Var\left(Y_0\right)}{\sigma_{Y_1 - Y_0}} - \frac{Cov\left(Y_1, Y_0\right)}{\sigma_{Y_1 - Y_0}}.$$

Since we can obtain $Var\left(Y_1\right)$ and $Var\left(Y_0\right)$, subtracting the second term from the first term we obtain

$$\sigma_{Y_1 - Y_0} = \frac{Var\left(Y_1\right) - Var\left(Y_0\right)}{\left[\frac{Cov\left(Y_1, I\right)}{\sigma_{Y_1 - Y_0}}\right] - \left[\frac{Cov\left(Y_0, I\right)}{\sigma_{Y_1 - Y_0}}\right]}$$

assuming that $\frac{Cov(Y_1,I)}{\sigma_{Y_1-Y_0}} \neq \frac{Cov(Y_0,I)}{\sigma_{Y_1-Y_0}}$. Hence we can identify $Cov\left(Y_1,Y_0\right)$ and $\sigma_{y_1-y_0}$.

The Roy model makes two strong assumptions. First, agents can choose sectors without incurring any costs. In a schooling model it is natural to allow for both pecuniary costs (such as tuition) as well as nonpecuniary costs (e.g., differences in preferences for education). Second, the model does not allow for uncertainty in future rewards. In the schooling model example, agents are assumed to keep working in the same sector for many periods. They are assumed to have full information on the future evolution of demand and supply of skills at the time they are making their educational decisions. Below we will show how we can extend the simple Roy model to allow for uncertainty and costs, and how we can generate restrictions to test and identify the information set of the agent at the time he makes sectoral choices.

The Generalized Roy model We introduce extensions of the Roy model one step at a time. We start by introducing participation costs, but postpone our discussion of uncertainty to the next section. Let C denote the cost of participation in sector "1". We may not observe C directly, because it may be partially determined by heterogeneity in agent preferences. This is the Generalized Roy model which was developed by Heckman (1976) and used by Willis and Rosen (1979). Allowing for direct costs for participation in sector "1" affects the identification of the joint distribution of rewards. As shown by Heckman and Honoré (1990), given data on X, Z,S and $Y = SY_1 + (1-S)Y_0$, one cannot recover the joint distribution $F\left(Y_0,Y_1\right)$. The problem is that if costs are not observed, choices alone do not provide enough information to identify the covariance between outcomes Y_0 and Y_1. We can immediately see why if we consider the case in which I is observed. In the generalized Roy model, $I = Y_1 - Y_0 - C$. Consequently:

$$Cov\left(Y_1,Y_0\right) + Cov\left(Y_1,C\right) = Cov\left(I,Y_1\right) - Var\left(Y_1\right) \tag{10}$$

We have two unknowns in one equation. If we use the information on $Cov\left(I,Y_0\right)$ it is still not enough to identify the model because we pick up a new term: $Cov\left(Y_0,C\right)$. The problem worsens when we account for the fact that I is observed only up to scale. As a result, without further assumptions, the Generalized Roy model is underidentified. Before presenting these assumptions, we embed the Roy model in an uncertain environment.

The Generalized Roy model in an Uncertain Economy We now introduce uncertainty in the generalized Roy model. We still assume a one-period economy in which agents are uncertain about the rewards at the time they are making their choices. The timing in this economy is as follows: At the beginning of the period, agents have an information set $\mathcal{I}$, whose elements may be unobserved by the analyst although they are known by the individual. Given this information set, agents make the decision about working in sector "0" or "1". After the decision is made, all uncertainty is revealed and agents observe their earnings. In this economy with uncertainty we rewrite the index I as:

$$I = E\left(Y_1 - Y_0 - C \middle| \mathcal{I}\right).$$

The decision rule becomes:

$$S = 1 \text{ if } I \geq 0; \; S = 0 \text{ otherwise.} \tag{11}$$

The decision rule (11) states the agents choose sector 1 if, and only if, the *expected* net gains are positive. We are assuming linear utility, implying that agents are risk neutral. More general models with respect to risk aversion can be identified (see the discussion in Cunha, Heckman, and Navarro, 2005 and the review in Heckman, Lochner, and Todd, 2006), but we focus on the case of linear utility for the sake of simplicity. If we assume that $X, Z, U_C \in \mathcal{I}$ we can rewrite the index I in terms of explanatory and unobserved variables as:

$$I = X\left(\beta_1 - \beta_0\right) - Z\gamma + E\left(U_1 - U_0 - U_C \middle| \mathcal{I}\right).$$

For the uncertain economy, the mean earnings for those who work in sector "1" is:

$$E\left[Y_1 \mid X, Z, I \geq 0\right] = X\beta_1 + E\left[U_1 \mid X, Z, E\left(U_1 - U_0 - U_C \middle| \mathcal{I}\right) \geq -X\left(\beta_1 - \beta_0\right) + Z\gamma\right]. \tag{12}$$

To proceed further we need to put more structure on the problem. In particular, we need to separate two distinct unobservable components. The first is the unobserved component (from the point of view of the analyst) that is known and acted on by the individual. This is the component captured by the expectation term $E\left(U_1 - U_0 - U_C \middle| \mathcal{I}\right)$. If this term changes, it will change the utility index I and expected earnings at the same time. The second term is unobserved by the analyst and unknown by the individual. It is captured by $U_s - E(U_s|\mathcal{I})$ for $s = 0, 1$. It does not affect schooling choices, but it affects realized earnings because we can always rewrite equation (8) as

$$Y_s = X\beta_s + E(U_s|\mathcal{I}) + \{U_s - E(U_s|\mathcal{I})\}$$

From equation (12), there is selection on $E(U_s \mid \mathcal{I})$ but not on $U_s - E(U_s \mid \mathcal{I})$. We name the first component "heterogeneity" and the second "uncertainty". The recent work of Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) develop models in which it is possible to identify and separate heterogeneity from uncertainty.

The key idea underlying their work is to impose a factor structure on the error terms of outcome and choice equations. We decompose the unobservable error terms as:

$$U_0 = \alpha_0\theta + \varepsilon_0, \quad U_1 = \alpha_1\theta + \varepsilon_1, \text{ and } \quad U_C = \alpha_C\theta + \varepsilon_C.$$

To fix ideas assume that the terms $\varepsilon_0, \varepsilon_1$ and ε_C are independent normal random variables:

$$\begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_C \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2_{\varepsilon_0} & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_1} & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_C} \end{bmatrix}\right).$$

The error terms in outcomes, $\varepsilon_0, \varepsilon_1$, are a source of uncertainty in future outcomes, so we impose $\varepsilon_0, \varepsilon_1 \notin \mathcal{I}$. Because agents must face the direct costs of choosing sector $S = 1$ at the time of their choice, it is natural to assume that agents know the realization of cost shocks at the time of choice S, so that $\varepsilon_C \in \mathcal{I}$.

All of the dependence among U_0, U_1, and U_C is captured through the factor θ. The factor θ is independent from $\varepsilon_0, \varepsilon_1, \varepsilon_C, X$ and Z. For simplicity, we assume that $\theta \sim N(0, \sigma^2_\theta)$, but this assumption can be relaxed[6]. Because of the loadings α_0, α_1 and α_C, the factor θ can affect U_0, U_1, and U_C differently, so by adopting the factor structure we are not imposing the sign of the covariance between U_0, U_1 and U_C. For example, it is possible that $\alpha_0 > 0$ while $\alpha_1 \leq 0$.

[6] In this paper we use the normality assumption because it is familiar, easily exposited, and easy to grasp the sources of identification. Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) establish that the distributions of $\varepsilon_0, \varepsilon_1, \varepsilon_C$ and the factor θ can be nonparametrically identified.

We do not need to assume $\theta \in \mathcal{I}$. We can test whether $\theta \in \mathcal{I}$ or $\theta \notin \mathcal{I}$ under conditions we specify in this paper. That is, we can test whether the agent knows (and acts on) information contained in θ that is not observed by the analyst. Using this test we can decompose the residuals of outcome equations between heterogeneity and luck (or uncertainty).

When the Generalized Roy model in the uncertain economy is analyzed using the factor structure proposed by Carneiro, Hansen, and Heckman (2003), the joint distribution of the rewards Y_0, Y_1 conditional on X is:

$$\begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} \mid X \sim N\left(\begin{bmatrix} X\beta_0 \\ X\beta_1 \end{bmatrix}, \begin{bmatrix} \alpha_0^2\sigma_\theta^2 + \sigma_{\varepsilon_0}^2 & \alpha_0\alpha_1\sigma_\theta^2 \\ \alpha_0\alpha_1\sigma_\theta^2 & \alpha_1^2\sigma_\theta^2 + \sigma_{\varepsilon_1}^2 \end{bmatrix} \right).$$

Therefore, in this simple normal linear formulation of Carneiro, Hansen, and Heckman (2003), identification of the joint distribution $F(Y_0, Y_1 \mid X)$ reduces to the identification of the parameters $\beta_0, \beta_1, \alpha_0, \alpha_1, \sigma_{\varepsilon_0}^2, \sigma_{\varepsilon_1}^2$ and σ_θ^2. From the observed data and the factor structure it follows that:

$$E\left(Y_1 \middle| X, S=1\right) = X\beta_1 + \alpha_1 E\left[\theta \middle| X, S=1\right] + E\left[\varepsilon_1 \middle| X, S=1\right] \tag{13}$$

which is equivalent to[7]:

$$E\left(Y_1 \middle| X, S=1\right) = X\beta_1 + \alpha_1 \frac{\left(\alpha_1 - \alpha_0 - \alpha_C\right)\sigma_\theta^2}{\sigma_\eta} \frac{\phi\left(\frac{X(\beta_1-\beta_0)-\boldsymbol{Z}\gamma}{\sigma_\eta}\right)}{\Phi\left(\frac{X(\beta_1-\beta_0)-\boldsymbol{Z}\gamma}{\sigma_\eta}\right)}. \tag{14}$$

It is easy to derive a similar expression for mean observed earnings in sector "0":

$$E\left(Y_0 \middle| X, S=0\right) = X\beta_0 - \alpha_0 \frac{\left(\alpha_1 - \alpha_0 - \alpha_C\right)\sigma_\theta^2}{\sigma_\eta} \frac{\phi\left(\frac{X(\beta_1-\beta_0)-\boldsymbol{Z}\gamma}{\sigma_\eta}\right)}{1-\Phi\left(\frac{X(\beta_1-\beta_0)-\boldsymbol{Z}\gamma}{\sigma_\eta}\right)}. \tag{15}$$

We can apply the two-step procedure proposed in Heckman (1976) to identify β_0, β_1, $\alpha_0\frac{(\alpha_1-\alpha_0-\alpha_c)\sigma_\theta^2}{\sigma_\eta^2}$ and $\alpha_1\frac{(\alpha_1-\alpha_0-\alpha_c)\sigma_\theta^2}{\sigma_\eta^2}$. Taking the ratio of the last two terms allow us to identify the ratio $\frac{\alpha_1}{\alpha_0}$.In factor analysis

[7]For the complete derivation see the "Derivation Appendix" on the website for this paper: http://jenni.uchicago.edu/frame-ineq

we always need a normalization in the factor loadings to set the scale, so we normalize $\alpha_0 = 1$.[8] This normalization sets α_1 and the term $\frac{(\alpha_0+\alpha_C-\alpha_1)\sigma_\theta^2}{\sigma_\eta^2}$. It is not helpful to look at the information from the variance of observed earnings to identify σ_η^2, α_C and σ_θ^2 separately. To see why, consider the variance of observed earnings in sector "1":

$$Var\left(Y_1 \mid X\boldsymbol{,} S=1\right) = \alpha_1^2 Var\left(\theta \mid S=1\right) + \sigma_{\varepsilon_1}^2.$$

From the variance we acquire a new term that identify: $\sigma_{\varepsilon_1}^2$. Because of this fact, when it comes to identifying the parameters related to the factors (such as σ_θ^2 and $\alpha's$), we have to restrict ourselves to information from the covariances, and not from the variances.

If there are variables W that are elements of X but not of Z we can identify more parameters. To see why, write:

$$X\left(\beta_1 - \beta_0\right) = W\left(\beta_1^W - \beta_0^W\right) + Z\left(\beta_1^Z - \beta_0^Z\right).$$

From equations (14) and (15) we can identify $\beta_1^W, \beta_0^W, \beta_1^Z$ and β_0^Z. From the probit coefficients we identify $\frac{\beta_1^W}{\sigma_\eta}, \frac{\beta_0^W}{\sigma_\eta}, \frac{\beta_1^Z}{\sigma_\eta}, \frac{\beta_0^Z}{\sigma_\eta}$ and $\frac{\gamma}{\sigma_\eta}$. Consequently, we can obtain σ_η from the choice equation by comparing β_1^W (from 14) to $\frac{\beta_1^W}{\sigma_\eta}$ (from the choice equation). If we know σ_η we can recover $\left(\alpha_0 + \alpha_C - \alpha_1\right)\sigma_\theta^2$ from the coefficient on the final term of the right hand side of (14) and (15). We cannot proceed any further because we know neither α_C nor σ_θ^2. Note that knowledge of σ_θ^2 is crucial for identification of the joint distribution $F\left(Y_0, Y_1\right)$. The entire dependence between Y_0 and Y_1 is through the factor θ.

As just demonstrated, adding uncertainty and imposing the factor structure is not enough to guarantee identification of the joint distribution of rewards $F\left(Y_0, Y_1\right)$. However, there is something we can say about the pattern of dependence between Y_0 and Y_1. The covariance of Y_0 and Y_1 is:

$$Cov\left(Y_0, Y_1\right) = \alpha_0\alpha_1\sigma_\theta^2 = \frac{\alpha_1}{\alpha_0}\alpha_0^2\sigma_\theta^2. \tag{16}$$

[8] It is easy to see why. Suppose that

$$\left[\begin{array}{c} Y_0 \\ Y_1 \end{array}\right] \mid X \sim N\left(\left[\begin{array}{c} X\boldsymbol{\beta}_0 \\ X\boldsymbol{\beta}_1 \end{array}\right], \left[\begin{array}{cc} \alpha_0^2\sigma_\theta^2 + \sigma_{\varepsilon_0}^2 & \alpha_0\alpha_1\sigma_\theta^2 \\ \alpha_0\alpha_1\sigma_\theta^2 & \alpha_1^2\sigma_\theta^2 + \sigma_{\varepsilon_1}^2 \end{array}\right]\right).$$

Then, if we define $\tilde{\alpha}_0 = k\alpha_0$, $\tilde{\sigma}_\theta = \frac{\sigma_\theta}{k}$ and $\tilde{\alpha}_1 = k\alpha_1$, for some $k > 0$, the parameters $\tilde{\alpha}_0, \tilde{\alpha}_1$ and $\tilde{\sigma}_\theta$ generate the same covariance matrix as the parameters α_0, α_1 and σ_θ.

We cannot directly compute this covariance, because we never observe the pair (Y_0, Y_1) at the same time. However, the factor structure implies that this covariance is given by the product of three terms $\frac{\alpha_1}{\alpha_0}$, α_0^2, and σ_θ^2. The terms α_0^2, and σ_θ^2 are always nonnegative, while the term $\frac{\alpha_1}{\alpha_0}$ is identified from equations (14) and (15). Consequently, we know at least the sign of the covariance, so we are able to say, for example, whether agents who choose to work in sector "0" would tend to have earnings above or below the mean in sector "1", if they had chosen to work in sector "1".

There are two other advantages of using factor models as shown here. First, they provide a flexible structure for adding information that can help to identify the joint distribution of rewards $F(Y_0, Y_1)$. Second, if we have observations on outcomes for more than one period, or if we have observations on more than one outcome in one period, the factor model provides a parsimonious way to capture the dependence across outcomes and between outcomes and choices. We now show how supplementing the basic model just presented aids in identification.

Cross-section data with one measurement equation Sometimes the analyst observes proxies for unobserved variables that affect both selection into states as well as outcomes in each sector. These are part of the source of the correlation between choices and outcomes. Because this correlation is captured by the factor, these proxies are potentially informative about the distribution of the factors. We introduce these proxies to obtain identification of our model. In the schooling example, it is reasonable to assume that unobserved ability affects the decision of how much education to obtain. It is also reasonable to assume that higher ability individuals will have higher earnings (even after controlling for the final education level) because higher ability makes an individual more productive. Suppose, that we have a set of test scores M for each person. Suppose that the measurement M is made for all agents regardless of whether $S = 1$ or $S = 0$, so that we do not have to worry about selection issues in the measurement. If we impose a linear specification we obtain:

$$M = X_M \beta_M + \alpha_M \theta + \varepsilon_M \tag{17}$$

where X_M are observable explanatory variables orthogonal to the factor θ that help predict the measurement. The variables X_M need not be the same as the explanatory variables X which predict outcomes Y_0 and Y_1. The disturbance ε_M is assumed to be normally distributed with $E(\varepsilon_M) = 0$, $Var(\varepsilon_M) = \sigma_{\varepsilon_M}^2$, independent from θ and X_M.

The measurement M is observed for everyone regardless of the schooling choice. Because we do not have to worry about selection problems in the measurement M, and because of the independence between X_M, θ, and ε_M, we can identify β_M from a simple OLS regression of (17). Above, we have already shown that by applying the Heckman two-step procedure to equations (14) and (15), we can identify β_1 and β_0. The availability of the measurement M yields the computation of five more covariance equations:

$$Cov\left(Y_0 - X\beta_0, M - X_M\beta_M\right) = \alpha_0\alpha_M\sigma_\theta^2, \tag{18}$$

$$Cov\left(Y_1 - X\beta_1, M - X_M\beta_M\right) = \alpha_1\alpha_M\sigma_\theta^2, \tag{19}$$

$$Cov\left(I - X\beta_1 + X\beta_0 + Z\gamma, M - X_M\beta_M\right) = \left(\alpha_1 - \alpha_0 - \alpha_C\right)\alpha_M\sigma_\theta^2 \tag{20}$$

$$Cov\left(I - X\beta_1 + X\beta_0 + Z\gamma, Y_0 - X\beta_0\right) = \left(\alpha_1 - \alpha_0 - \alpha_C\right)\alpha_0\sigma_\theta^2, \tag{21}$$

and

$$Cov\left(I - X\beta_1 + X\beta_0 + Z\gamma, Y_1 - X\beta_1\right) = \left(\alpha_1 - \alpha_0 - \alpha_C\right)\alpha_1\sigma_\theta^2 \tag{22}$$

Assume now that we make one normalization: $\alpha_M = 1$ instead of normalizing $\alpha_0 = 1$. By taking the ratio of (22) to (20), we can identify α_1. Because $\alpha_M = 1$, we can use (19) to solve for σ_θ^2, which we use in (18) to identify α_0. The last term to be identified is α_C which we can recover from (20). Note that given the normalization $\alpha_M = 1$ we have four parameters ($\alpha_0, \alpha_1, \alpha_C$, and σ_θ^2) and five equations to identify them. We never used equation (21) in our calculations above. Consequently, we can use this source of overidentification to generate a test that can reject whether agents know the factors or not at the time the sectoral choice is made, which we describe in detail in the next section.

Also, note that the covariance between Y_0 and Y_1 can be identified, because we know each of the terms in the right-hand side of (16). Consequently, from a single cross-section on outcomes, choices and a measurement equation, we can recover the joint distribution of Y_0 and Y_1 and use such knowledge to answer counterfactual questions. We now show how access to panel data allows us to identify the model even in the absence of a measurement.

The observation of outcomes for two different periods The availability of a measurement M guarantees identification by providing more information, *i.e.*, more covariance equations which we can use to solve for the factor loadings and the parameters that characterize the distribution of the factor. However, researchers may not observe candidate proxies that could be used as measurements, but they may observe, for example, outcomes for more than one period. To fix ideas, suppose that we observe earnings for at least two periods, $Y_t = (1-S)\,Y_{0,t} + SY_{1,t}$ and $Y_{t+k} = (1-S)\,Y_{0,t+k} + SY_{1,t+k}$ for some $k \neq 0$. For $\tau = t, t+k$, let

$$Y_{0,\tau} = X\beta_{0,\tau} + \theta\alpha_{0,\tau} + \varepsilon_{0,\tau}$$

$$Y_{1,\tau} = X\beta_{1,\tau} + \theta\alpha_{1,\tau} + \varepsilon_{1,\tau}.$$

We assume that $\varepsilon_{s,\tau}$ are independently normally distributed random variables for $s = 0,1$ and $\tau = t, t+k$.We can allow X to vary over time as well, but this would only complicate the notation without adding any insight to our discussion of identification, so we will abstract from time variation for X. Assuming that the interest rate r is zero, so that there is no discounting, the utility index is defined as

$$I = E\left(Y_{1,t+k} + Y_{1,t} - Y_{0,t+k} - Y_{0,t} - C\right|\mathcal{I}\left.\right).$$

If we maintain the assumption that $\varepsilon_{s,\tau} \notin \mathcal{I}$ for $s = 0,1$ and $\tau = t, t+k$ the net utility index I is defined as:

$$I = X\left(\beta_{1,t+k} + \beta_{1,t} - \beta_{0,t+k} - \beta_{0,t}\right) - Z\gamma + \left(\alpha_{1,t+k} + \alpha_{1,t} - \alpha_{0,t+k} - \alpha_{0,t} - \alpha_C\right)\theta - \varepsilon_C.$$

It is easy to show that mean observed earnings for periods $\tau = t, t+k$ can be expressed as[9]:

$$E\left(Y_{0,\tau}\right| S = 0\left.\right) = X\beta_{0,\tau} - \pi_{0,\tau}\frac{\phi\left(\frac{\tilde{\mu}_I(X,Z)}{\sigma_\varphi}\right)}{1 - \Phi\left(\frac{\tilde{\mu}_I(X,Z)}{\sigma_\varphi}\right)}$$

$$E\left(Y_{1,\tau}\right| S = 1\left.\right) = X\beta_{1,\tau} + \pi_{1,\tau}\frac{\phi\left(\frac{\tilde{\mu}_I(X,Z)}{\sigma_\varphi}\right)}{\Phi\left(\frac{\tilde{\mu}_I(X,Z)}{\sigma_\varphi}\right)}.$$

where:

$$\pi_{s,\tau} = \alpha_{s,\tau}\frac{\left(\alpha_{1,t+k} + \alpha_{1,t} - \alpha_{0,t+k} - \alpha_{0,t} - \alpha_C\right)\sigma_\theta^2}{\sigma_\varphi}.$$

[9]See the "Derivation Appendix" on the website for details. http://jenni.uchicago.edu/frame-ineq.

Again, for each time period τ we can apply Heckman (1976) two-step procedure and obtain consistent estimators of $\beta_{s,\tau}$ and $\pi_{s,\tau}$, $s = 0, 1$; $\tau = t, t + k$. Assume we normalize $\alpha_{0,t} = 1$. If we compute the ratio $\frac{\pi_{s,\tau}}{\pi_{0,t}}$ we can identify $\alpha_{s,\tau}$ for $s = 0, 1$ and $\tau = t, t + k$.We can identify the factor variance σ_θ^2 from the covariance:

$$Cov\left(Y_{1,t+k} - X\beta_{1,t+k}, Y_{1,t} - X\beta_{1,t}\right) = \alpha_{1,t}\alpha_{1,t+k}\sigma_\theta^2.$$

At this point, it remains to show identification of α_C and $\sigma_{\varepsilon_C}^2$. We have already discussed identification of $\sigma_{\varepsilon_C}^2$ and concluded that we can recover it as long as we have variables W that are part of the vector X but not part of the vector Z. If this is not the case, then we must normalize $\sigma_{\varepsilon_C}^2 = 1$. In either case, we can then recover α_C from $\pi_{0,t}$ since:

$$\pi_{0,t} = \frac{\left(\alpha_{1,t+k} + \alpha_{1,t} - \alpha_{0,t+k} - \alpha_{0,t} - \alpha_C\right)\sigma_\theta^2}{\sigma\left(\alpha_{1,t+k} + \alpha_{1,t} - \alpha_{0,t+k} - \alpha_{0,t} - \alpha_C\right)^2\sigma_\theta^2 + \sigma_{\varepsilon_C}^2}$$

and the only unknown in this equation is α_C.

It is interesting to remark that the covariance across counterfactuals:

$$Cov\left(Y_{0,\tau}, Y_{1,\tau'}\right) = \alpha_{0,\tau}\alpha_{1,\tau'}\sigma_\theta^2$$

can be computed since all of the elements on the right-hand side have been determined. Note that again we have one more equation than unknown, and we can use the overidentification to test whether $\theta \in \mathcal{I}$ or not.

Multiple Measurements and Panel data on outcomes Some lucky researchers may observe many measurements and may have the availability of panel data on outcomes. In this case, it is possible to identify models with more than one factor. Consider, for example, the empirical analysis of Carneiro, Hansen, and Heckman (2003). In their study, they have a two-sector model (high school and college labor). They observe a set of test scores $M_1, ..., M_K$. They model test score k, $k = 1, 2, ..., K$ as:

$$M_k = X_M\beta_M^k + \theta_1\alpha_M^k + \varepsilon_M^k.$$

They normalize $\alpha_M^1 = 1$. They model (log) earnings $Y_{s,t}$ as:

$$Y_{s,t} = X\beta_{s,t} + \theta_1\alpha_{s,t} + \theta_2\delta_{s,t} + \varepsilon_{s,t}.$$

They normalize $\delta_{1,1} = 1$. Recall that one normalization is needed to set the scale of the factor. Let C denote the costs of choosing college. They model costs C as:

$$C = Z\gamma + \theta_1\alpha_C + \theta_2\delta_C + \varepsilon_C.$$

Agents choose college if, and only if:

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

Note that Carneiro, Hansen, and Heckman (2003) estimate a two-factor model. Because the test scores only depend on the first factor, they can use the test scores to identify the distribution of θ_1. They use the earnings and choice equations to identify the distribution of factor θ_2. We increase the number of factors that can be identified if the length of the panel increases, or if we have measurements associated with other characteristics that affect both choices and earnings. For example, Cunha, Heckman, and Navarro (2005) identify a model with three factors because they have a set of measurements and earnings for five periods. For the general treatment about the limitation on the number of factors we refer the reader to Carneiro, Hansen, and Heckman (2003). They also present an analysis of identification without imposing distributional assumptions on the unobservables. Cunha, Heckman, and Schennach (2006) generalize their analysis.

4 Distinguishing between Heterogeneity and Uncertainty

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

$$y_t = X\beta + \delta S + v_t, \tag{23}$$

where y_t, X, S, v_t denote, respectively, log earnings, observable characteristics (which may vary over time), educational attainment and unobservable characteristics of person i at time t. Often the error term v_t is decomposed into two or more components. For example,

$$v_t = \phi + \varepsilon_t. \tag{24}$$

The term ϕ is a person-specific fixed effect. The error term ε_t is generally assumed to be serially correlated, say $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$ where η_t is an independently and identically distributed innovation with mean zero. It is widely accepted that components of X, ϕ, and ε_t all contribute to measured inequality. However, the literature is silent about the difference between heterogeneity and uncertainty, the unforecastable part of earnings—what Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns, and Michelson (1972) call "luck". On intuitive grounds, the predictable components of v_t have a different effect on welfare than the unpredictable components, especially if people are risk averse and cannot fully insure against the uncertainty. Is uncertainty ϕ? Is it ε_t? Is it $\phi + \varepsilon_t$? Or η_t? Statistical decompositions such as (23) and (24) tell us nothing about which components of (23) are unforecastable by agents.

The methodology summarized in this chapter, and developed more fully in Cunha, Heckman, and Navarro (2005), provides a framework within which it is possible to identify and separate components that are forecastable from ones that are not. The essential idea of their method can be illustrated in the case of educational choice. In order to choose between high school and college, say at age 17, agents forecast future earnings (and other returns) at each schooling level. Using this information from educational choice at age 17, together with the realization of earnings that are observed at later ages, it is possible to estimate and test which part of future earnings are forecast by the agent at age 17.

In this method, we use choice information to extract *ex ante* or forecast earnings to distinguish them from *ex post* or realized earnings. The difference between forecast and realized earnings allows us to identify the components of uncertainty facing agents at the time they make their schooling decisions. With this method, we can distinguish predictable heterogeneity from uncertainty.

We formalize the argument by considering the two-factor, two-period earnings model of Carneiro, Hansen, and Heckman (2003), Let $\mathcal{I}$ denote the information set of the agent at the time the schooling choice must be made. The population decision rule governing sectoral choice is

$$I = \ E\left(\sum_{t=0}^{1}\left(Y_{1,t}-Y_{0,t}\right)-C \mid \mathcal{I}\right).$$

In this economy, the decision rule is quite simple: one attends school if the expected gains from schooling are greater than or equal to the expected costs. Under the assumptions made the choice index I is written as:

$$I = E\left(\sum_{t=0}^{1}\left[X\left(\beta_{1,t}-\beta_{0,t}\right)+\theta_1\left(\alpha_{1,t}-\alpha_{0,t}\right)+\theta_2\left(\delta_{1,t}-\delta_{0,t}\right)+\varepsilon_{1,t}-\varepsilon_{0,t}\right]-\left(Z\gamma+\theta_1\alpha_C+\theta_2\delta_C+\varepsilon_C\right) \,\middle|\, \mathcal{I}\right).$$

The decision rule of agents is:

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

For the sake of simplicity, assume that X, Z and θ_1 are in the information set of the agent, and that the interest rate is zero ($r = 0$). By assumption, the error terms $\varepsilon_{s,t}$ for $s, t = 0, 1$ are not in the information set of the agent. We want to test whether θ_2 is also in the information set of the agent at the time of the schooling choice. If $\theta_2 \in \mathcal{I}$, then it reflects heterogeneity across agents (since agents know and act on it). If $\theta_2 \notin \mathcal{I}$, then it reflects either uncertainty or some information that agents know, but do not act on.

Suppose that our null hypothesis is $\theta_1 \in \mathcal{I}$, but $\theta_2 \notin \mathcal{I}$. Under the null:

$$I = E\left(Y_{1,1}+Y_{1,2}-Y_{0,1}-Y_{0,2}-C|\mathcal{I}\right) = \mu_I(X,Z)+\alpha_I\theta_1.$$

Consequently, we have that:

$$Cov\left(I-\mu_I(X,Z), Y_{1,1}-X\beta_{1,t})\right) = \alpha_I\alpha_{1,1}\sigma^2_{\theta_1}$$

Let Δ_{θ_2} be such that:

$$Cov\left(I-\mu_I(X,Z), Y_{1,1}-\mu_1(X)\right)-\alpha_I\alpha_{1,1}\sigma^2_{\theta_1}-\Delta_{\theta_2}\alpha_I\delta_{1,1}\sigma^2_{\theta_1} = 0.$$

Then, we reject the null, and conclude that agents know and act on the information contained in factor 2,

θ_2, if we reject that $\Delta_{\theta_2} = 0$.

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.[10] Cunha, Heckman, and Navarro (2005) develop this analysis further.

5 Empirical Results

We now apply these models to data. In this section we report estimates of a model of schooling choice, show how one can recover the distribution of θ, α_C and $\alpha_{s,t}$, $s = 0, 1$, $t = 0, \ldots, T$, and put our analysis to use to estimate counterfactual distributions for different policies, to compute their consequences for mobility and inequality and to measure the contributions of "luck" to post-schooling earnings. We answer how much of the post-schooling earnings is predictable at the age schooling decisions are made.

5.1 The Data, Equations, and Estimation

In our empirical analysis, we use a sample of white males from the NLSY data. Following the preceding theoretical analysis, we consider only two schooling choices: high school and college graduation. We assume that credit markets are perfect. By this we mean that restrictions on borrowing against future earnings are assumed not to be important. See Cameron and Taber (2004) and Carneiro and Heckman (2002) for evidence supporting this assumption in the context of schooling choices.[11] Carneiro, Hansen, and Heckman (2003) assume the absence of credit markets and obtain empirical results on the extent of uncertainty similar to the ones presented here, so the issue of whether credit markets function or not does not affect the main conclusions of our analysis.

[10] Identification depends on the length of the panel, the number of measurement equations and the variation in Z and X.

[11] We note, however, that a large literature in macroeconomics based on separable preferences claims to find evidence of substantial departures from complete contingent claims markets using consumption data. See e.g. the survey in Browning, Hansen, and Heckman (1999).

To economize on space, we place many tables on our website for this paper: http://jenni.uchicago.edu/frame-ineq. Web table W1a presents the list of variables used in our empirical analysis. They show that college graduates have higher present value of earnings than high school graduates. College graduates also have higher test scores and come from better family backgrounds. They are more likely to live in a location where a college is present, and where college tuition is lower.

To simplify the empirical analysis, we divide the lifetimes of individuals into two periods. The first period covers ages 17 through 28 and the second goes from 29 through 65. We impute missing wages and project earnings for the ages not observed in the NLSY data using the procedure described in Appendix B of Carneiro, Hansen, and Heckman (2003). In Cunha, Heckman, and Navarro (2005), we consider alternative ways to create full life cycle histories. Combination of data sets is required since the NLSY does not contain information on the full life cycle of earnings. We augment the NLSY data with data from the PSID to estimate the lifetime earnings of the NLSY sample members. For each schooling level s, $s \in \{0, 1\}$, for each period $t \in \{1, 2\}$ we calculate the present value of earnings at age 17, $Y_{s,t}$. We assume that the error term for $Y_{s,t}$ is generated by a two factor model,

$$Y_{s,t} = X\beta_{s,t} + \theta_1\alpha_{s,t,1} + \theta_2\alpha_{s,t,2} + \varepsilon_{s,t}. \tag{25}$$

We omit the "i" subscripts to eliminate notational burden. This model is all that is required to fit our data. Additional factors, when entered, do not contribute to the fit of the model. In web table W1b we list the elements of X used in our empirical analysis. They are listed in the column "PV Earnings." We normalize $\alpha_{0,1,2} = 1$.

For the measurement system for cognitive ability we use five components of the ASVAB test battery: arithmetic reasoning, word knowledge, paragraph composition, math knowledge and coding speed. We link the first factor to this system of ability tests, and exclude the other factor from it. Thus, we adopt two normalizations. First, the loading on the first factor on the arithmetic reasoning test is set to one. Second, the loading on the second factor is set to zero in all test equations. Thus the test scores are devoted to measuring ability. We include family background variables among the covariates X_T in the ASVAB test equations. Web table W1b lists the variables used in X_T. They are listed in the column "Test System." Formally, let T_j denote the test score j :

$$T_j = X_T \omega_j + \theta_1 \alpha_{test_j,1} + \varepsilon_{test_j}. \tag{26}$$

The cost function C is given by:

$$C = Z\gamma + \theta_1 \alpha_{C,1} + \theta_2 \alpha_{C,2} + \varepsilon_C \tag{27}$$

where the Z are variables that affect the costs of going to college and include variables that do not affect outcomes $Y_{s,t}$ (such as local tuition, and distance to college). Web table W1b shows the full set of covariates used in Z under the column "Cost Function."

For the educational choice equation, we assume that agents know X, Z, ε_C and some, but not necessarily all, components of θ. Let the components known to the agent be $\bar{\theta}$. The decision rule for attending college is based on:

$$V = E\left(Y_{1,1} + \frac{Y_{1,2}}{1+r} - Y_{0,1} - \frac{Y_{0,2}}{1+r} \mid X, \bar{\theta}\right) - E\left(C \mid Z, X, \bar{\theta}, \varepsilon_C\right) \tag{28}$$

where future earnings are discounted at interest rate $r = .03$. Individuals go to college if $V > 0$. We test and do not reject the hypothesis that individuals, at the time they make their college decisions, know their cost functions, the factors $\bar{\theta}$, and unobservables in cost ε_C. However, they do not know $\varepsilon_{s,t}$, $s \in \{0,1\}$, $t \in \{1,2\}$ at the time they make their educational choices, and they may not know other components of $\vec{\theta}$. The factor loadings on the θ not in $\bar{\theta}$ are estimated to be zero in the choice equation. See Cunha, Heckman, and Navarro (2005) for further discussion of this test and for extensions of the method.

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. As shown in Ferguson (1983), mixtures of normals with a large number of components approximate any distribution of θ_k arbitrarily well.[12] Even though the $\varepsilon_{s,t}$ are nonparametrically identified, we assume in the empirical work reported

[12]In the ℓ^1 norm.

here that they are normally distributed.[13] We estimate the model using Markov Chain Monte Carlo methods. In web tables W2a–W2c , available at our website, we present estimated coefficients and factor loadings for the model.

5.2 Results

We now present the empirical results for our model.

5.2.1 How the Model Fits the Data

To assess the validity of our estimates, we perform a variety of checks of fit of predictions against the data. We first compare the proportions of people who choose each schooling level. In the NLSY data, 55% choose high school and 45% choose college. The model predicts 56% and 44%, respectively. The model replicates the observed proportions, and formal tests of equality of predicted and actual proportions cannot be rejected.

In web figure W1 we show the densities of the predicted and actual present value of earnings for the overall sample of high school and college graduates. The fit is also good. Web figures W2 and W3, show the same densities restricted to the sample of those who choose high school (W2) and college (W3). The fit is good. When we also perform formal tests of equality of predicted and actual distributions, we cannot reject the hypothesis of equality of the distributions for each schooling choice using a chi-square goodness of fit test at a 5% significance level for all 3 cases (Table 1). We conclude that a two factor model is enough to fit the data. From this analysis, we conclude that earnings innovations $\varepsilon_{s,t}$ relative to a two factor model are not in the agents' information sets at the time they are making schooling decisions. If they were, additional factors would be required to capture the full covariance between educational choices and future earnings, but when we enter additional factors, they do not improve the fit of the model to data and have zero estimated factor loadings in the choice equation.

[13] Models where the $\varepsilon_{s,t}$ are allowed to be mixtures do not change the conclusions of this paper. However, they increase the complexity of the simulation analysis. For this reason, we use the simple normal framework to estimate the uniquenesses. We stress, however, that it is not a requirement, just a matter of convenience.

5.2.2 The Factors: Non-normality and Evidence on Selection

In order to fit the data, one must allow for non-normal factors, as one can see from the evidence summarized in figure 1. To generate figure 1, we compute the variance of the distribution of factor 1, say σ_{θ_1}. Since the factors have mean zero, we can plot the estimated density of factor 1 against that of a normal random variable with mean zero and variance σ_{θ_1}. We proceed in a similar fashion for factor 2. Neither factor is normally distributed. A traditional assumption used in factor analysis (see, e.g., Jöreskog, 1977) is violated. Our approach is more general and does not require normality.

Web figure W4 plots the density of factor 1 conditional on educational choices. Since factor 1 is associated with cognitive tests, we can interpret it as an index of "ability". The agents who choose college have, on average, higher ability. Selection on ability is important in explaining college attendance. A similar analysis of factor 2 reveals that schooling decisions are not very much affected by it. However, factor 2 is important for predicting future earnings, as we show below.

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

In estimating the distribution of earnings in counterfactual schooling states within a policy regime (*e.g.*, the distributions of college earnings for people who actually choose to be high school graduates), the usual approach is to assume that both distributions are the same except for an additive constant — the coefficient of a schooling dummy in an earnings regression. More recently developed methods partially relax this assumption by assuming preservation of ranks across potential outcome distributions, but do not freely specify the two outcome distributions (see Chernozhukov and Hansen, 2005; Heckman, Smith, and Clements, 1997; Vytlacil, 2002). Our approach relaxes this assumption by allowing for arbitrary dependence across potential outcome distributions. Table 2 presents the conditional distribution of *ex post* potential college earnings given *ex post* potential high school earnings decile by decile. The table displays a strong positive dependence between the relative positions of individuals in the two distributions. In particular, for all high school deciles more than 50% of the individuals located at any decile in the high school earnings distribution will be within one decile of their original position in the college earnings distribution. However, the dependence is far from perfect. For example, almost 10% of those who are at the sixth decile of the high school distribution would be in the eighth decile of the college distribution. Observe that this comparison is not being made in terms of positions in the overall distribution of earnings. We can determine where individuals are located in the distribution of population potential high school earnings and the distribution of potential college earnings although in the data we only observe them in either one or the other state. The assumption of perfect dependence across components of counterfactual distributions that is maintained in much of the recent literature (e.g., Juhn, Murphy, and Pierce, 1993) is, however, too strong, at least in this application.

Figures 2 and 3 present the marginal densities of predicted and counterfactual earnings for college (figure 2) and high school participants (figure 3). When we compare the densities of the present value of earnings in the college sector for persons who choose college against the counterfactual densities of high school earnings for college graduates, the density of the present value of earnings for college graduates is to the right of the counterfactual density of the present value of earnings of college graduates if they were high school graduates. Figure 4 reveals that college graduate earnings are higher for high school graduates than high school earnings. The surprising feature of both figures is that the overlap of the distributions is substantial. Many high school graduates would have large earnings as college graduates.

Figure 4 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). College graduates have returns distributed somewhat "to the right" of high school graduates, so the difference is not only a difference for the mean individual but is actually present over the entire distribution. An economic interpretation of figure 4 is that agents who choose a college education are the ones who tend to gain more from it.

Web table W3 reports the fitted and counterfactual present value of earnings for agents who choose high school. The typical high school student would earn \$703.78 thousand dollars over the life cycle. She would earn \$1,021.97 thousand if she had chosen to be a college graduate.[14] This implies a return of 46% to a college education over the whole life cycle (*i.e.,* a monetary gain of 318.19 thousand dollars). In web table W4, we note that the typical college graduate earns \$1,122.69 thousand dollars (above the counterfactual earnings of what a typical high school student would earn in college), and would make only \$756.13 thousand dollars over her lifetime if she chose to be a high school graduate instead. The returns to

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

college education for the typical college graduate (which in the literature on program evaluation is referred to as the effect of Treatment on the Treated) is 50% above that of the return for a high school graduate. In monetary terms we would say that a college graduate has a gain of going to college almost 50,000 dollars higher over her lifetime than does the typical high school graduate.

With our methodology we can also determine returns to the marginal student. Web table W5 reveals that the average individual who is just indifferent between a college education and a high school diploma earns $743.40 thousand dollars as a high school graduate or $1,089.97 thousand dollars as a college graduate. This implies a return of 48%. The returns to people at the margin are above those of the typical high school graduate, but below those for the typical college graduate. Since persons at the margin are more likely to be affected by a policy that encourages college attendance, their returns are the ones that should be used in order to compute the marginal benefit of policies that induce people into schooling.

A major question that emerges from our analysis is, why, if high school graduates have positive returns to attending college, don't all people attend? People do not pick schooling levels based only on monetary returns. Recall that their choice criterion (equation (28)) includes also the pecuniary and non-pecuniary costs of actually attending college. Figure 5 shows the estimated density of the monetary value of this cost both overall and by schooling level. While almost no high school graduate perceives a negative cost (*i.e.,* a benefit) of attending college; around one third of college graduates actually perceive it as a benefit. Web table W6 explores this point in more detail by presenting the mean total cost of attending college and the mean cost that is due to ability. Costs on average are smaller for college graduates. College graduates have higher ability. The average contribution of ability to costs is positive for high school graduates (a true cost). It is negative for college graduates, so it is perceived as a benefit. This is the answer to our puzzle: people do not only (or even mainly) make their schooling decisions by looking at their monetary returns in terms of earnings. Psychic costs play a very important role. Differences in ability are one force behind this result.[15,16]

[15] Furthermore, we know that this result is not sensitive to the specification of the credit market. Carneiro, Hansen, and Heckman (2003) obtain a similar conclusion in a model where people are not allowed to borrow or lend. In our model, on the other hand, there are no constraints to borrowing or lending. Cunha, Heckman, and Navarro (2005) present additional evidence on this issue.

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

5.2.4 Mobility and Heterogeneity versus Uncertainty

In figures 6 through 8 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 6 we start by assuming that the agents do not know their factors; consequently, $\Theta = \varnothing$. If we let the agent learn about factor 1,[17] so that, $\Theta = \{\theta_1\}$, then the reduction in the forecast variance is basically nil. This exercise shows that while factor 1 is important for forecasting educational choices, it does not do a very good job in forecasting earnings. Now, assume that the agent is given knowledge of factor 2, but not factor 1, so that $\Theta = \{\theta_2\}$. Then the agent is able to substantially reduce the forecast variance of earnings in high school. Thus, while factor 2 does not greatly affect college choices, it greatly informs the agent about his future earnings. When the agent is given knowledge of both factors 1 and 2, that is, $\Theta = \{\theta_1, \theta_2\}$, he can forecast earnings marginally better. Factor 1 provides information on ability, but almost none on future earnings. Figure 7 reveals much the same story about college earnings.

Table 3 presents the variance of potential earnings in each state, and returns under different information sets available to the agent. We conduct this exercise for the forecast of period 1, period 2, and lifetime earnings. We report baseline variances and covariances without conditioning and the remaining uncertainty state as a fraction of the baseline no information state when different components are given to the agents. Note that knowledge of factor 2 is fundamentally important in reducing forecast variance for period 2 earnings.

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

It is interesting to note that knowledge of the factors enables agents to make better forecasts. Figure 8 presents an exercise for returns to college $(Y_1 - Y_0)$ similar to that presented in figures 6 and 7 regarding information sets available to the agent. Knowledge of factor 2 also greatly improves the forecastability of returns, 80% of the variability in returns is forecastable at age 17. The levels are even more predictable

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

(94% for high school; 97% for college). Most variability across people is due to heterogeneity and not uncertainty[18].

5.2.5 *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. From our analysis, we conclude that, at the time agents pick their schooling, the ε's in their earnings equations are unknown to them. These are the components that correspond to luck as defined by Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns, and Michelson (1972). It is clear that decision making would be different, at least for some individuals, if the agent knew these chance components when choosing schooling levels since decision rule (1)–(2) would now be

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

where no expectation is taken to calculate V since all terms on the right hand side of the top equation are known with certainty by the agent.

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.

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

As an example of the power of our method to evaluate the consequences of policy on income inequality, we analyze a cross-subsidized tuition policy indexed by family income level. We construct joint distributions of outcomes within policy regimes (treatment and no treatment or schooling and no schooling) and joint

[18] The high predictability of future earnings and returns may be inflated by our imputation procedure. In recent unpublished work Cunha and Heckman (2006) using pooled data and avoiding imputation show that we get even higher predictability for returns to college (i.e., only 14.11% of the variance of unobservable in returns are uncertain to the agent), but higher uncertainty for levels (19% for college and 32% for high school).

distributions of choices ($Y = SY_1 + (1-S)Y_0$) across policy regimes. The policy analyzed is as follows. A prospective student whose family income at age 17 is below the mean is allowed to attend college free of charge. The policy is self financing within each schooling cohort. To pay for this policy, persons attending college with family income above the mean pay a tuition charge equal to the amount required to cover the costs of the students from lower income families as well as their own.

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 9 shows that the marginal distributions of income in both the prepolicy state and the postpolicy state are essentially identical. Under the anonymity postulate 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.

Panel 1 of table 4 presents this analysis by describing how the 9.3% of the people who are affected by the policy move between deciles of the distribution of income. These statistics describe movements from one income distribution in the initial regime to another income distribution associated with the new regime. The policy affects more people at the top deciles than at the lower deciles. Around half of the people affected who start at the first decile remain at the first decile. People in the middle deciles are spread both up and down and a large proportion of people in the upper deciles is moved into a lower position (only sixteen percent of those starting on the top decile remain there after the policy is implemented). 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.

Thus far, we have focused on constructing and interpreting the joint distribution of outcomes across the two policy regimes. If outcomes under both regimes are observed, these comparisons can be made using panel data. No use of counterfactuals is necessary. However, our methods will apply if either or both policy regimes are unobserved but are proposed. Taking advantage of the fact that we can identify not only joint distributions of earnings over policy regimes but also over counterfactual states within regimes we can learn a great deal more about the effects of this policy, whether or not policy regimes are observed.[19]

Table 5 and panels 2 and 3 of table 4 reveal that not only 9.3% 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 (4.8% percent of the population). 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 10 summarizes some of the evidence presented in table 5. The panels 2 and 3 of table 4 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). We can also see that the policy tends to move high school graduates up in the income distribution and moves college graduates down. As another example of the generality of our method and the new insight into income mobility induced by policy that it provides, we can determine where people come from and where they end up at in the counterfactual distributions of earnings. Table 6 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. Most people stay in their decile or move closely to adjacent ones. Given that some people benefit from the policy while others lose, it is not clear whether society as a whole values this policy positively or not. An advantage of our method is that it allows us to calculate the effect that the policy has on welfare. An individual's relative utility is not only given by earnings but also by the monetary value of psychic costs. We can predict how people would vote if the policy analyzed in this section were proposed. Table 7 shows the result of such an exercise. The policy lowers the mean earnings for people affected by it. Most people not indifferent to the policy would vote against it.

5.2.7 Robustness

In this section we discuss the robustness of the results with respect to many of the empirical and modelling assumptions in our work. Carneiro, Hansen, and Heckman (2003) assume that markets are incomplete.

[19] It is implausible that we would have panel data on policy regimes where under one regime a person goes to school and under another he does not.

However, this assumption does not change the qualitative implication of a tuition policy. In particular, they find that the agents at the top percentile of the distribution of high school earnings are more likely to change their schooling status under a tuition reduction policy. What drives the result is the estimated positive correlation between college and high school earnings. A reduction in cost tends to generate greater gains for those individuals who have large income in the high school sector.

Cunha and Heckman (2006) show that the aggregation of earnings into a few periods tends to overestimate the share of total variance that is known by individuals. They aggregate earnings in 16 and 24 periods and find that earnings predictability is around 60% for college and high school graduates, and around 55% for returns. Cunha and Heckman (2006) also test the 3% discount rate that is used to calculate the present value of earnings. They consider different values (2%,4%, and 5%). They show that as discount rates increase, there is more predictability in lifetime earnings. Consequently, temporary shocks tend to vanish as we aggregate earnings in fewer periods and use a higher discount rate to calculate present value of earnings. However, the estimated correlation between college and high school earnings is always positive regardless of the aggregation of life cycle earnings strategy we use or the discount rate we adopt to calculate present value of earnings. Heckman, Lochner, and Todd (2006) present a comprehensive survey of the literature.

6 Summary and Conclusion

This paper summarizes and applies a new body of research on counterfactual analyses of income inequality and mobility. We construct counterfactuals within policy regimes and counterfactuals across policy regimes. Using the methods presented here, it is possible to understand the sources of inequality, and the inequality and mobility consequences of social policies much more deeply than is possible using traditional measures based on the anonymity axiom.

We show how to construct distributions of counterfactuals within policy regimes using factor models. With these same tools, we show how to separate variability into two components: (a) those that are predictable by a certain age (heterogeneity) and (b) those that are not (luck).

We apply these methods to analyze the returns to college education. We find that by age 17, when college decisions are made, prospective students can forecast roughly 80% of the lifetime variance in their returns to schooling. Heterogeneity and not uncertainty drives the variance in earnings both cross-sectionally and over time. Given the relatively small role for uncertainty, it is unlikely that expectational errors about future earnings play a major role in explaining college choices.

We also find that counterfactual outcomes both within and across policy regimes are highly correlated. However there is a lot of slippage in ranks across potential outcomes within one policy regime and the outcomes chosen across policy regimes. Ranks are by no means identical across these counterfactual distributions. One justification for the use of the anonymity postulate in the analysis of data on income distributions is that outcomes are independent across policies. That justification is strongly rejected in our data, as is the polar assumption that ranks are perfectly dependent across policies.

We show how our methods can reveal where in initial and final distributions persons induced to change their education by a tuition subsidy policy will come from, and where they end up. Such analyses can be made both in terms of initial and final outcome distributions and in distributions of potential outcomes associated with each educational choice. We present a much richer analysis of the inequality and mobility consequences of policies than are available from analyses based on panel data or data from social experiments.

This paper has ignored the analysis of general equilibrium effects operating through factor markets even though such effects are empirically important for large scale programs. Substantial changes in educational enrollments will affect the wages of college and high school students. The next step in our research program is to graft the methods in this paper to the general equilibrium framework of Heckman, Lochner, and Taber (1998a,b,c, 1999) to provide a more comprehensive analysis of the effects of policies on inequality.

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (1999). Training effects on employment when the training effects are heterogeneous: An application to Norwegian vocational rehabilitation programs. University of Bergen Working Paper 0599; and University of Chicago.

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

Bourguignon, F. and L. A. P. da Silva (2003). *The Impact of Economic Policies on Poverty and Income Distribution*. Washington, D.C.: The World Bank.

Browning, M., L. P. Hansen, and J. J. Heckman (1999, December). Micro data and general equilibrium models. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1A, Chapter 8, pp. 543–633. Elsevier.

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

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

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

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

Chernozhukov, V. and C. Hansen (2005, January). An iv model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Cunha, F. and J. J. Heckman (2006). The evolution of earnings risk in the us economy. Presented at the 9th World Congress of the Econometric Society, London.

Cunha, F., J. J. Heckman, and S. Navarro (2005, April). Separating uncertainty from heterogeneity in life cycle earnings, the 2004 hicks lecture. *Oxford Economic Papers 57*(2), 191–261.

Cunha, F., J. J. Heckman, and S. M. Schennach (2006). Estimating the elasticity of substitution between early and late investments in the technology of cognitive and noncognitive skill formation. Unpublished manuscript, University of Chicago, Department of Economics. Presented at the Yale Conference on Macro and Labor Economics, May 5–7, 2006.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In H. Chernoff, M. Rizvi, J. Rustagi, and D. Siegmund (Eds.), *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*, pp. 287–302. New York: Academic Press.

Fields, G. S. (2003, March). Economic and social mobility really are multifaceted. Paper presented at the Conference on Frontiers in Social and Economic Mobility, Cornell University, Ithaca, New York.

Heckman, J. J. (1976, December). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. (1992). Randomization and social policy evaluation. In C. Manski and I. Garfinkel (Eds.), *Evaluating Welfare and Training Programs*, pp. 201–230. Cambridge, MA: Harvard University Press.

Heckman, J. J. (2001, August). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy 109*(4), 673–748.

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

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

Heckman, J. J., L. J. Lochner, and C. Taber (1998b, May). General-equilibrium treatment effects: A study of tuition policy. *American Economic Review 88*(2), 381–386.

Heckman, J. J., L. J. Lochner, and C. Taber (1998c, May). Tax policy and human-capital formation. *American Economic Review 88*(2), 293–297.

Heckman, J. J., L. J. Lochner, and C. Taber (1999). General-equilibrium cost-benefit analysis of education and tax policies. In G. Ranis and L. Raut (Eds.), *Trade, Growth and Development: Essays in Honor of T.N. Srinivasan*, Chapter 14, pp. 291–349. Elsevier Science B.V.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings equations and rates of return: The Mincer equation and beyond. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*. Amsterdam: North-Holland. forthcoming.

Heckman, J. J. and S. Navarro (2006). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, forthcoming.

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

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

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume 6*. Amsterdam: Elsevier. Forthcoming.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume 6*. Amsterdam: Elsevier. Forthcoming.

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

Jöreskog, K. G. (1977). Structural equations models in the social sciences: Specification, estimation and testing. In P. Krishnaiah (Ed.), *Applications of Statistics*, pp. 265–287. New York: North-Holland.

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

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

Kane, T. J. (1994, October). College entry by blacks since 1970: The role of college costs, family background, and the returns to education. *Journal of Political Economy 102*(5), 878–911.

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

Ravallion, M. (2003). Assessing the poverty impact of an assigned program. In F. Bourguignon and L. Silva (Eds.), *The Impact of Economic Policies on Poverty and Income Distribution*, pp. 103–122. Washington, D.C.: The World Bank.

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

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

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

Table 1

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

Table 2

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

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

High School	College 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.

Table 3
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 Θ = Ø	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 Θ = Ø	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 Θ = Ø	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).

Table 4

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
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
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 5
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.

Table 6

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
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.

Table 7

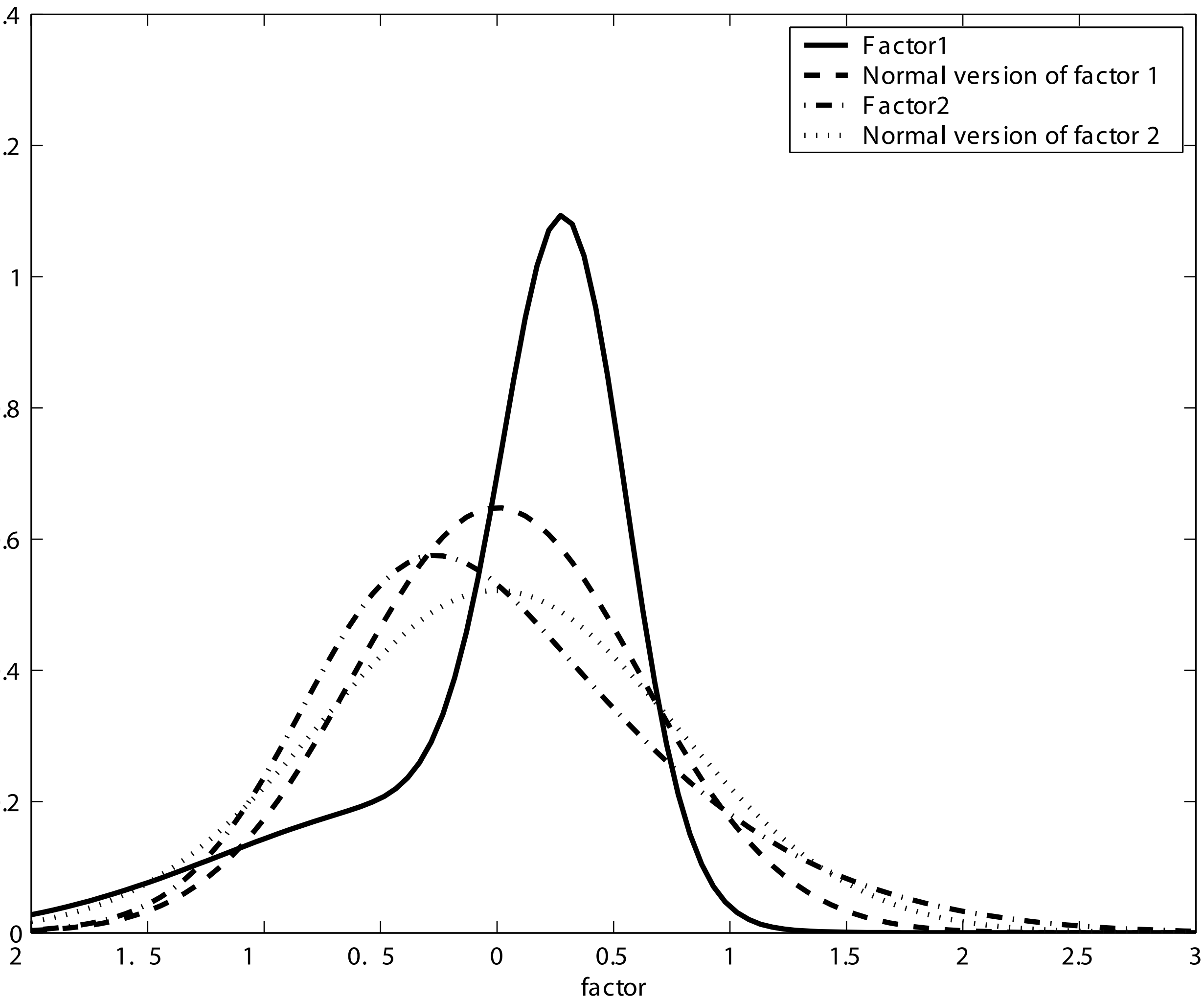
Voting outcome of proposing cross subsidizing* tuition

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.

Figure 1
Densities of factors and their normal equivalents



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

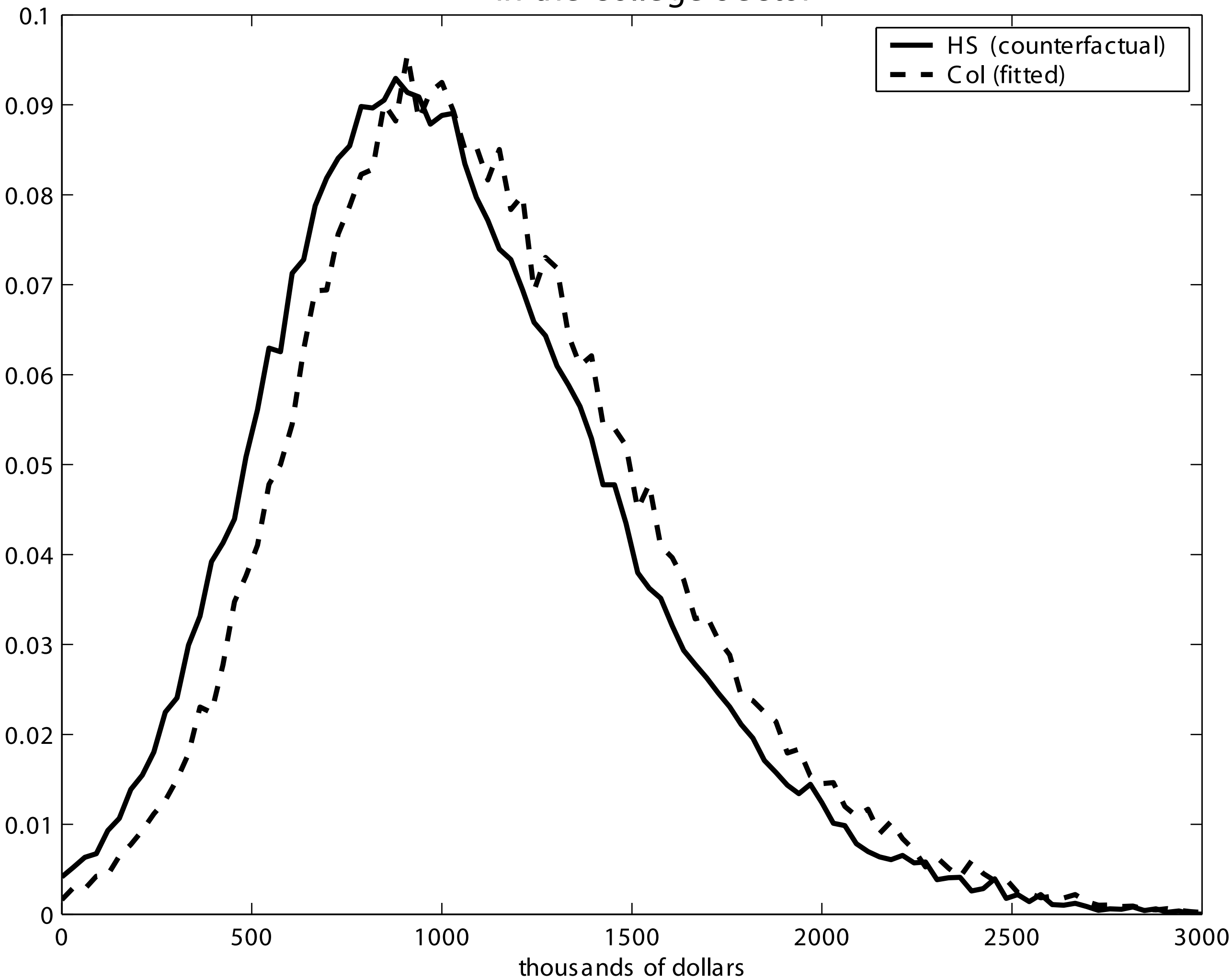
We assume that $f(\theta_1)$ is a mixture of normals. Assume $\mu_1 = E(\theta_1)$, $\sigma_1 = Var(\theta_1)$.

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, $\phi(\mu_1,\sigma_1)$.

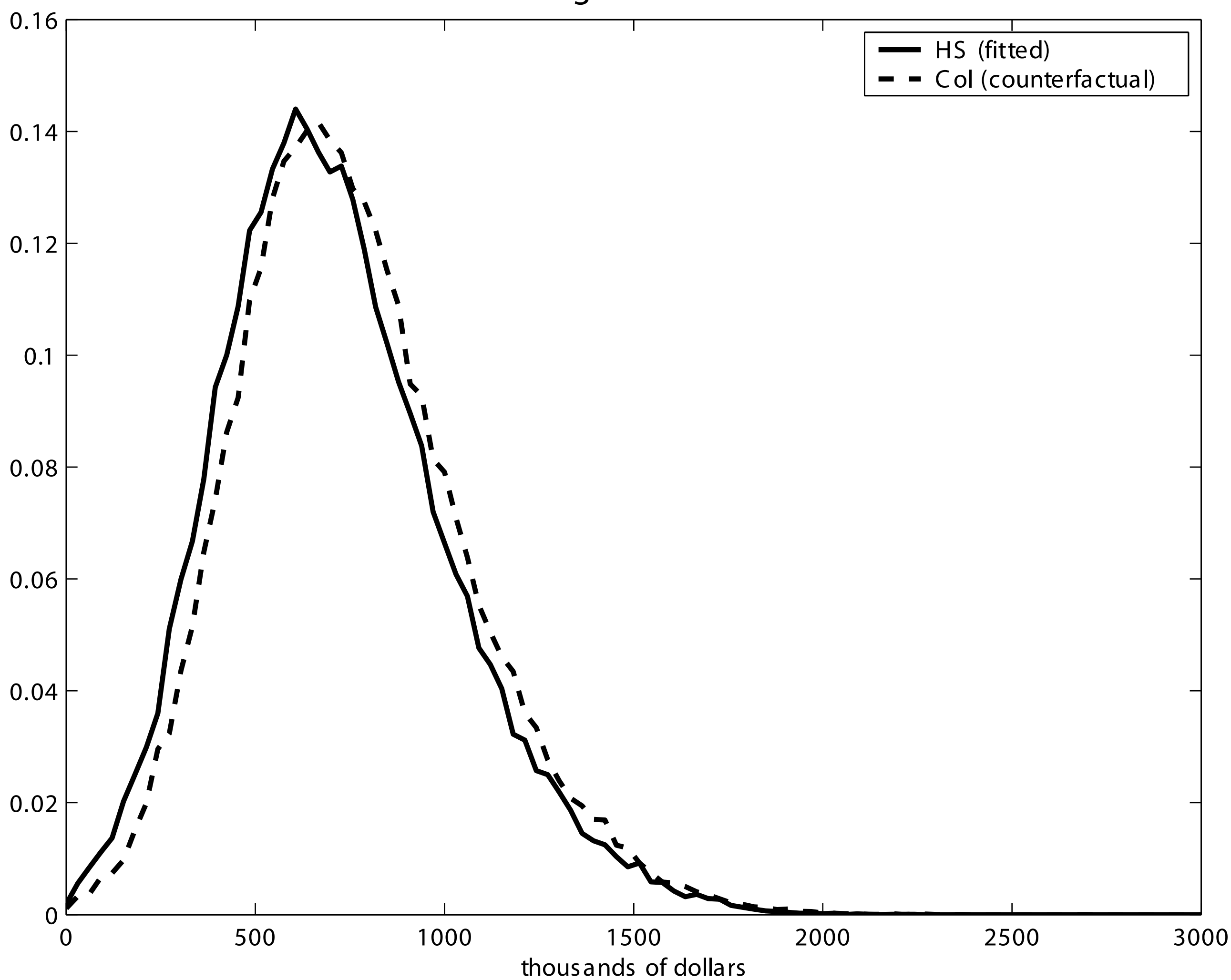
We proceed similarly for factor 2, where the fitted density is plotted with dots and dashes and the normal version is plotted with dots.

Figure 2
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, $f(y_1 \mid S = 0)$.

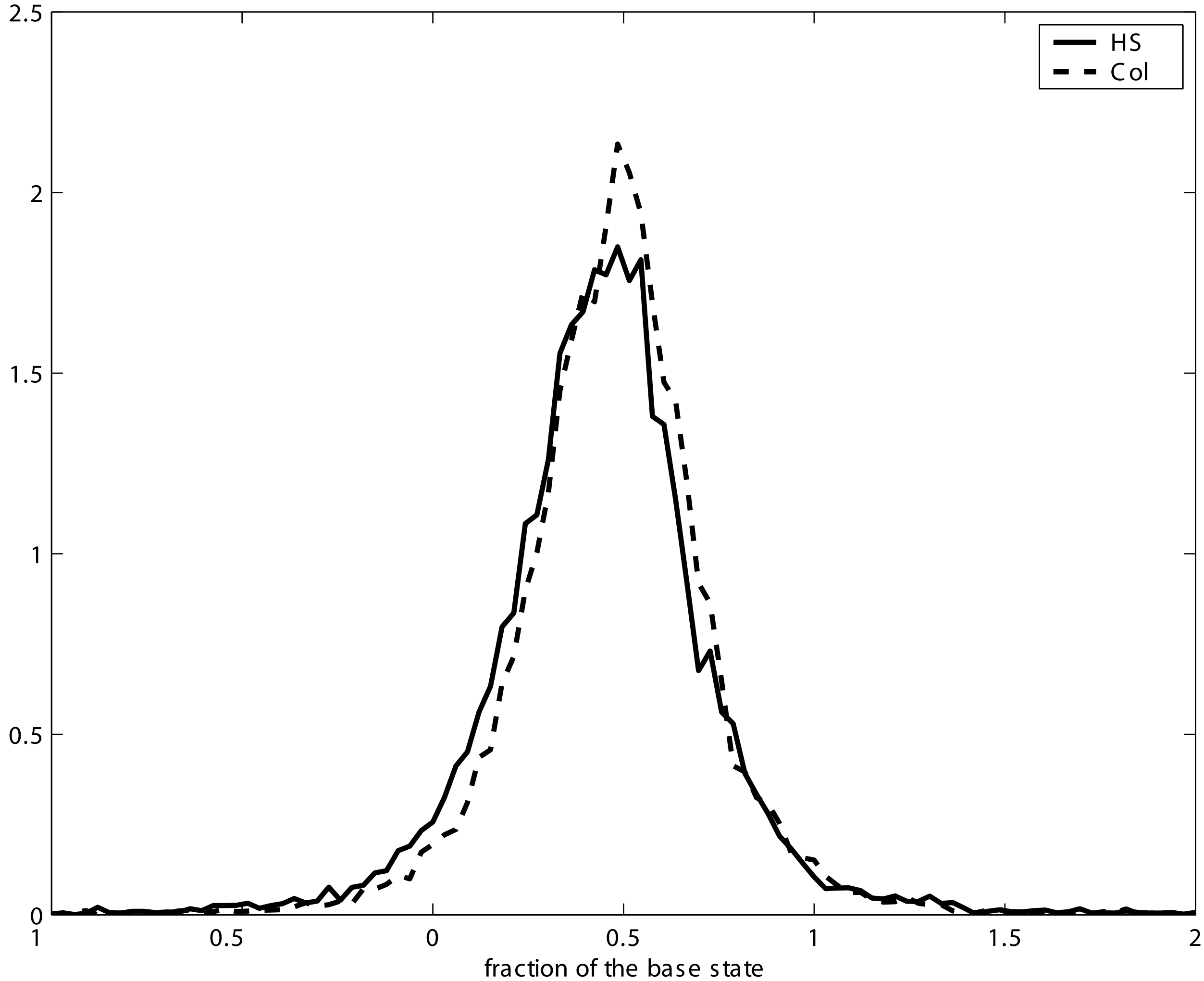
Figure 3
Densities of present value of earnings
in the high school sector



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, $f(y_0 \mid S = 1)$.

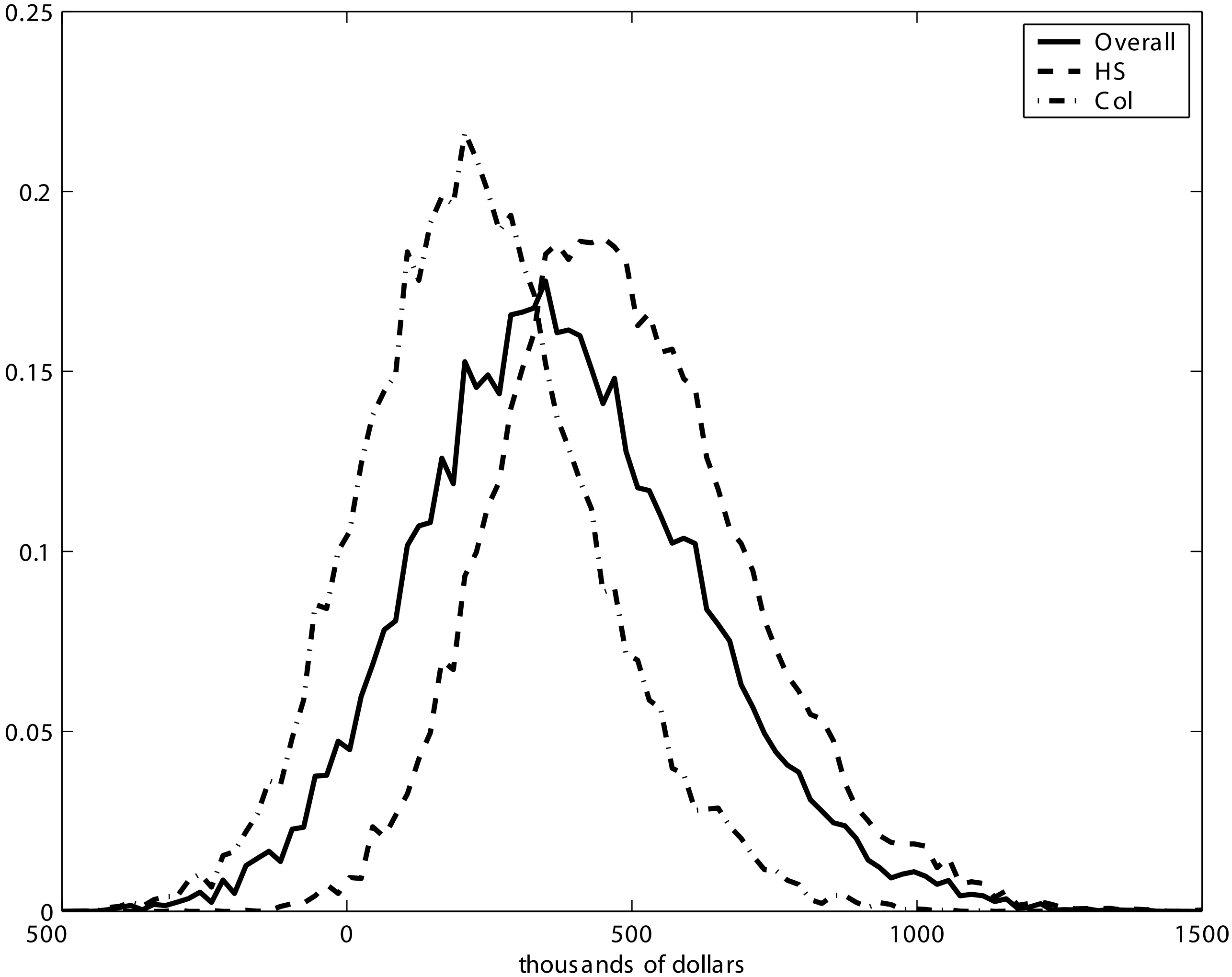
Figure 4

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, $f(r \mid S=1)$.

Figure 5
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.

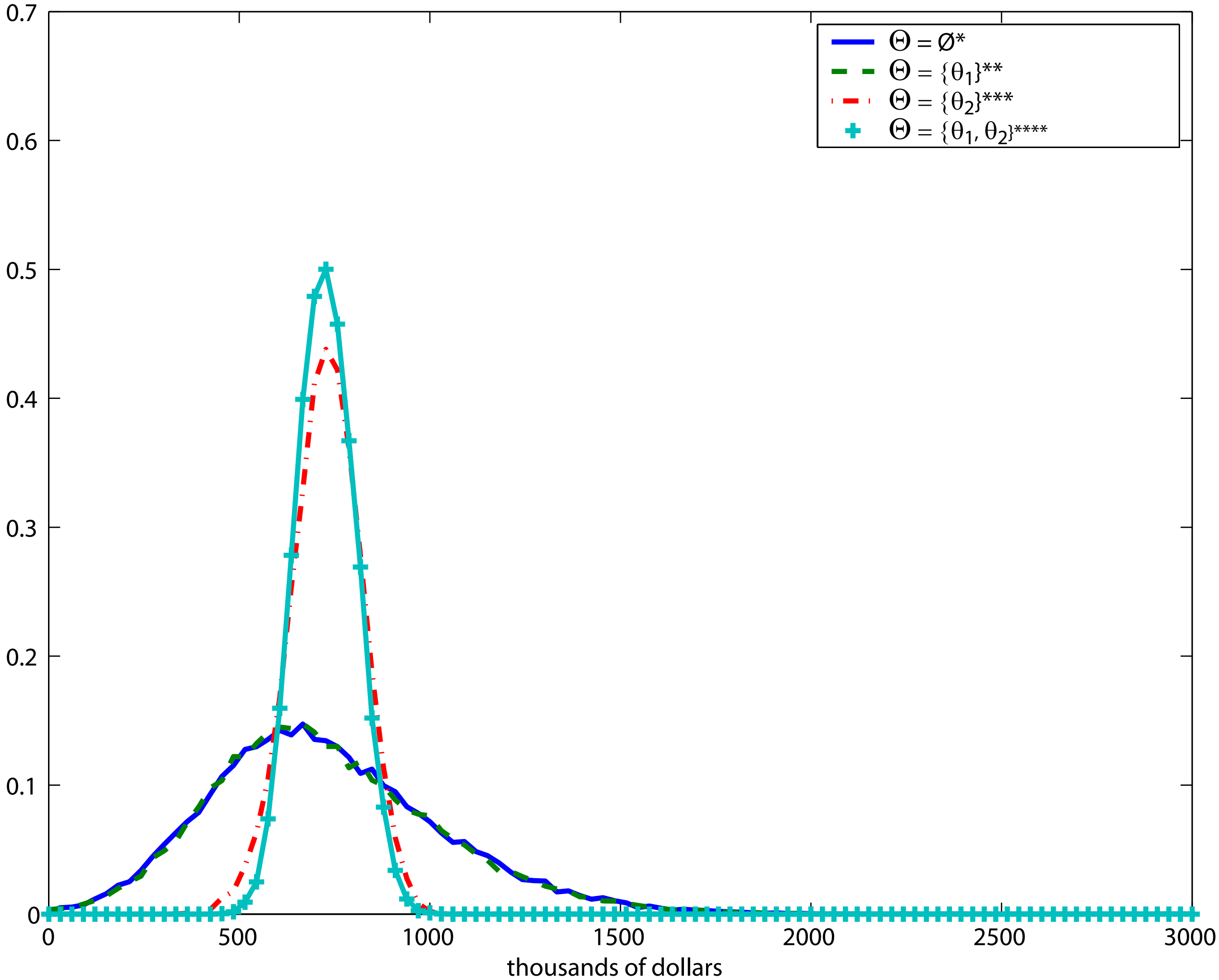
The monetary value of psychic costs is given by:

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

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

Figure 6

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., $\Theta = \emptyset$.

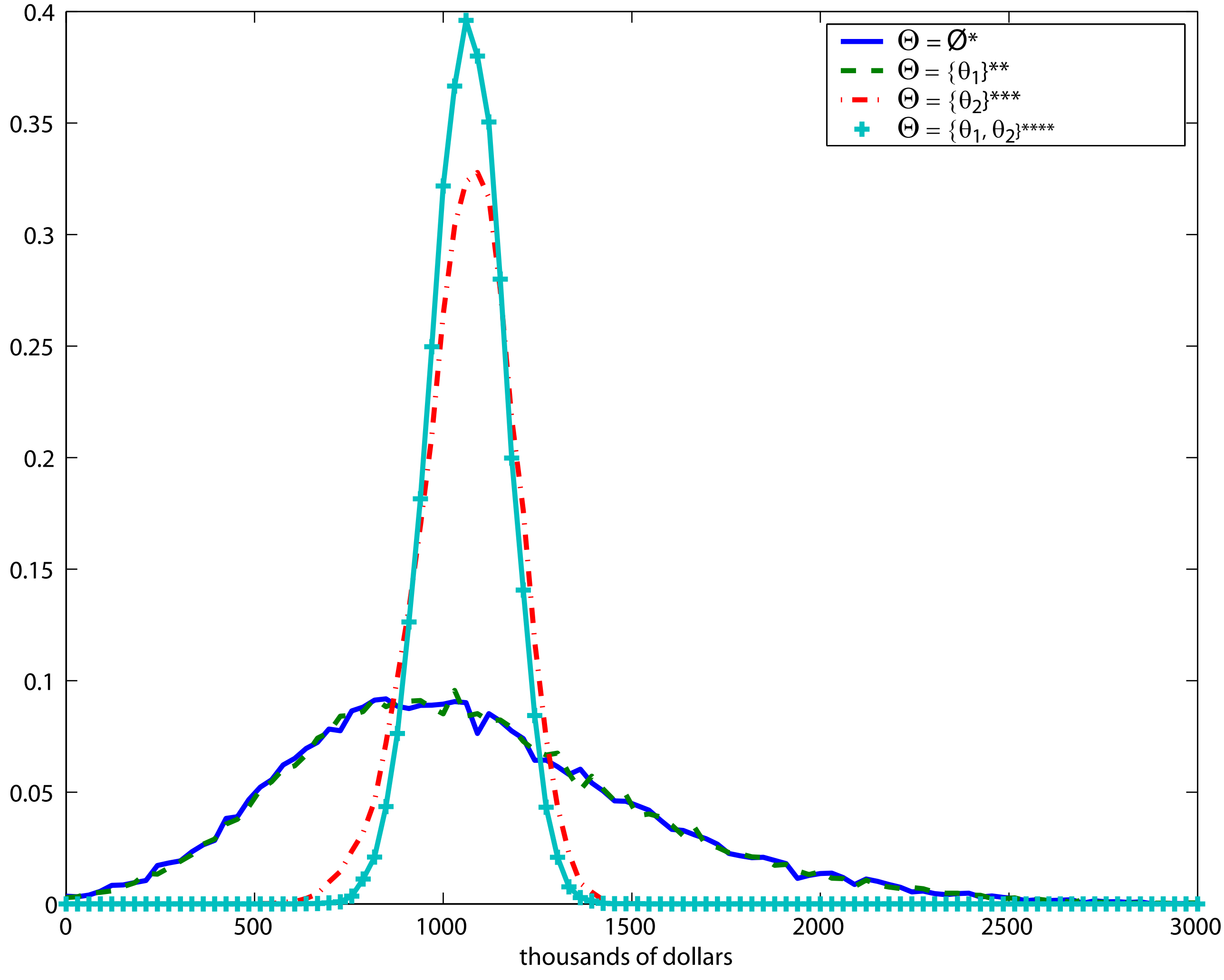
** Plot of $f(y_0 | \mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.

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

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

Figure 7

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., $\Theta = \emptyset$.

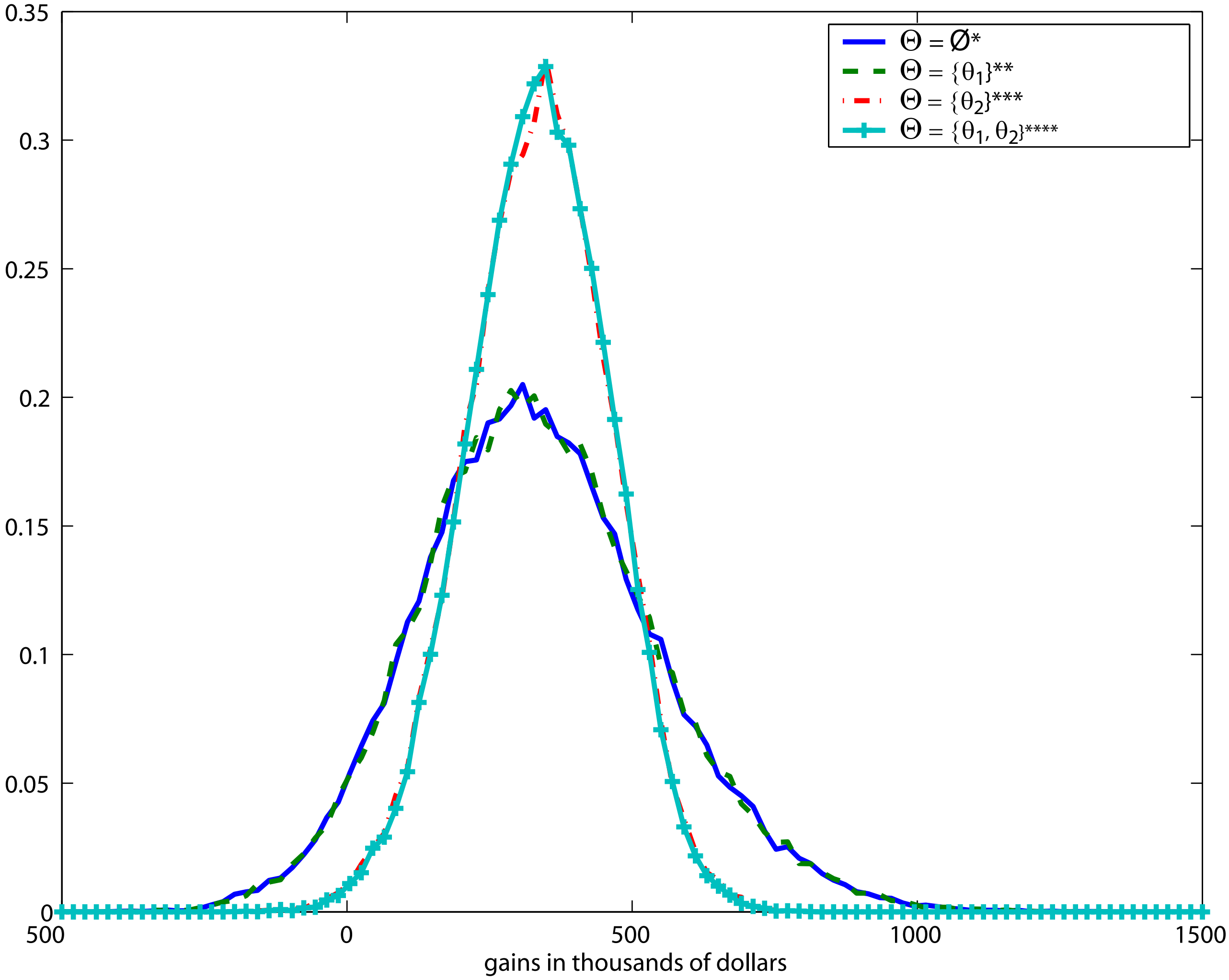
** Plot of $f(y_1 | \mathcal{I})$ when only factor 1 is in the information set, i.e., $\Theta = \{\theta_1\}$.

*** 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., $\Theta = \{\theta_1, \theta_2\}$.

Figure 8

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:

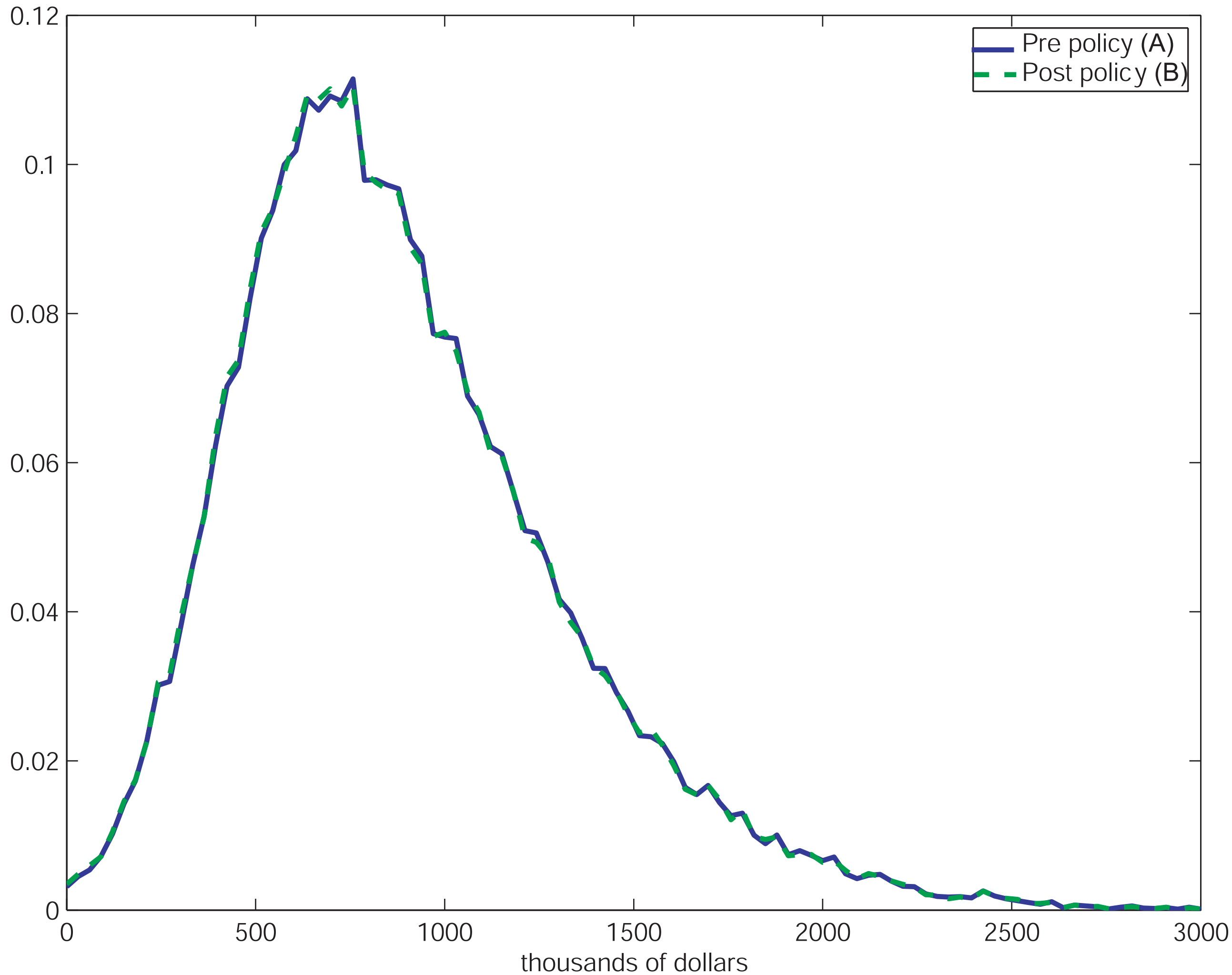
*Plot of $f(\Delta \mid \mathcal{I})$ under no element of θ in the information set, i.e., $\Theta = \varnothing$.

** 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., $\Theta = \{\theta_1, \theta_2\}$.

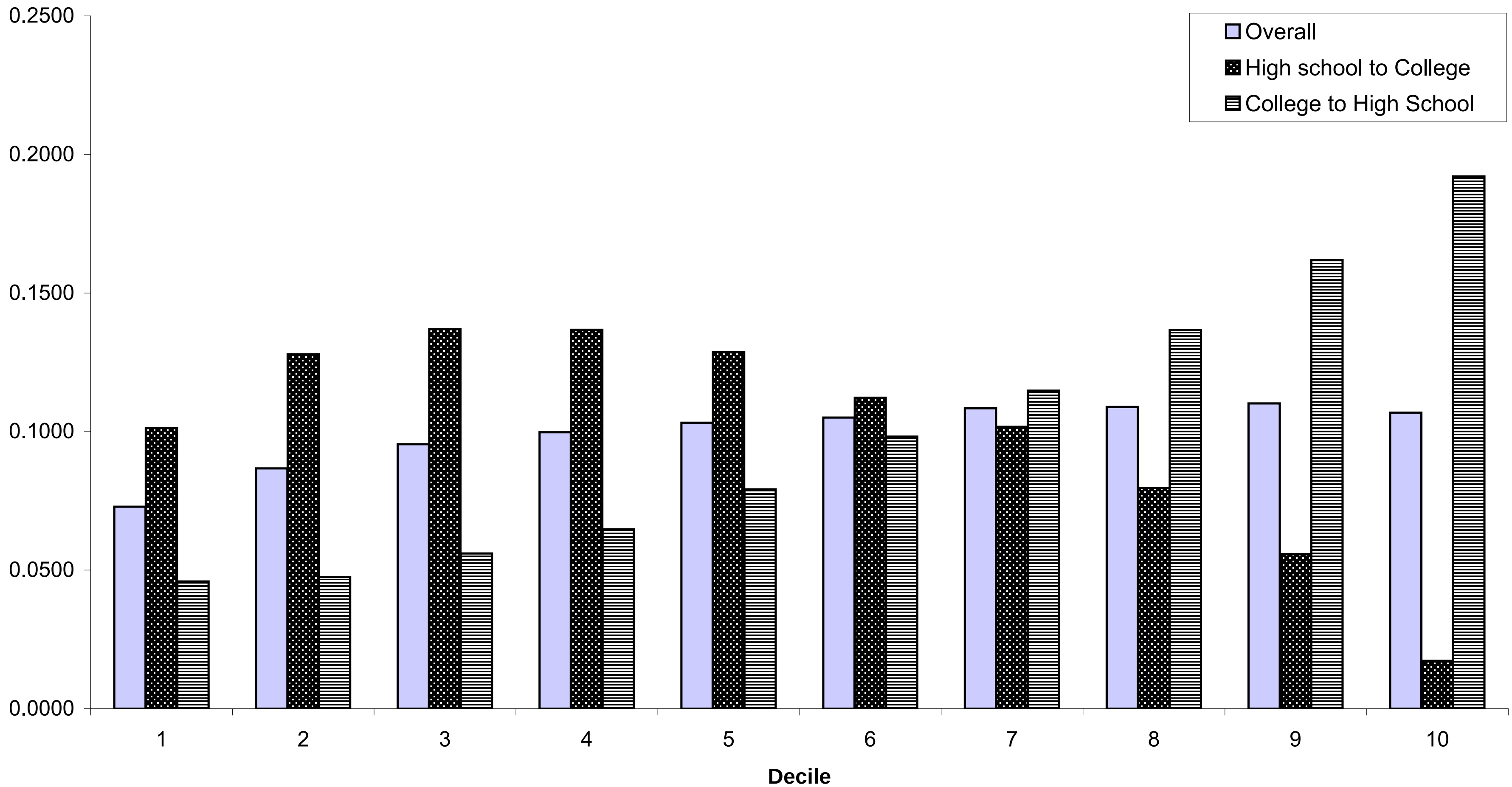
Figure 9
Densities of present value of lifetime earnings
before and after implementing cross subsidy policy



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 $f(y^A)$, $g(y^B)$.

Figure 10

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.