Issues in the Economics and Econometrics of Policy Evaluation*

Explorations Using a Factor Structure Model

Philipp Eisenhauer

University of Mannheim

ZEW Mannheim

IZA Bonn

First draft, September 1, 2007

This draft, November 3, 2012

*For further information or questions and suggestions, please contact us at info@policy-lab.org. We thank James Heckman, Pedro Carneiro, Jing Jing Hsee, Rémi Piatek, Stefano Mosso, Edward L. Sung, Christian Goldammer, Miriam Gensowski, Pia Dovern-Pinger, Martin Nybom, Francois Laisney, and Bernd Fitzenberger for useful comments. We thank George Yates for his advice regarding the computational implementation. We thank the Graduate School of Economic and Social Sciences (GESS) at the University of Mannheim and the Economics Research Center (ERC) at the University of Chicago for funding this research.

Abstract

We combine the generalized Roy model with a factor structure assumption. Together, this allows for a readily accessible discussion of the economics and econometrics of policy evaluation within a unified framework. We explore several issues by an estimation of the returns to college. We report the average returns of a college education. However, we do not stop there. We estimate the whole distribution of returns and document considerable heterogeneity. We establish that agents select their schooling level based on gains unobservable by the econometrician and subsequently show which margin of agents is affected by two alternative policy changes. Finally, we unify the abundance of average effect parameters by using the marginal treatment effect.

Keywords: Econometrics of Policy Evaluation, Potential Outcomes Model, Treatment Effects, Counterfactual Distributions.

JEL Classifications: C15, C21, C31.

Contact Details

Philipp Eisenhauer

Department of Economics

University of Mannheim

L7, 3-5, 68131 Mannheim

Germany

Phone: +49 (621) 181-3428

Email: eisenhauer@policy-lab.org

Contents

1 Introduction

Econometric policy evaluation is important. Policy evaluation informs policy makers and the general public about the relevant economic trade offs between alternative policies. Thus, it contributes to informed policy choices. In its context, the effect of some program on subsequent outcomes is of central importance. For instance, the impacts of social welfare programs, active labor market policies, and the public education system are under high scrutiny as these programs consume a considerable amount of public funds.

Econometric policy evaluation is complicated. A naive comparison of a treated and untreated sample leads to misleading conclusions. Agents that select into treatment are fundamentally different from those that do not (Browning et al., 1999; Heckman, 2001). They make different choices and even experience different outcomes given the same choice. A valid assessment of a policy requires an understanding of the underlying sources of variation. It is this understanding that determines the set of applicable econometric tools (Heckman and Vytlacil, 2005; Heckman et al., 2006b).

Econometric policy evaluation is multifaceted. The effects of policies are summarized by objects of interest, and different objects answer different policy questions (Heckman and Vytlacil, 2007a,b). Often, the focus is on the average effect of assigning a random individual from the population to treatment. However, this does not answer the relevant policy question when policy makers can only affect incentives for voluntary participation. Moreover, a focus on average effects masks potentially important treatment effect heterogeneity. A positive effect on average does not rule out a negative effect for a considerable share of the population. Thus, the whole distribution of effects is of interest and can provide additional insights into the effectiveness of a program (Heckman et al., 1997; Abbring and Heckman, 2007).

In this paper, we combine the generalized Roy model with a factor structure assumption. Together, this allows for a readily accessible discussion of the economics and econometrics of policy evaluation within a unified framework. In doing so, we build on the existing work by Carneiro et al. (2003), Cunha et al. (2005) and Cunha and Heckman (2007). We enrich it with the more recent contributions by Heckman and Vytlacil (2005) and Carneiro et al. (2011).

We explore several issues by an estimation of the returns to college using the National Longitudinal Survey of Youth of 1979 (NLSY79). We report the average returns of a college education. However, we do not stop there. We estimate the whole distribution of returns and document considerable heterogeneity. We establish that agents select their schooling level based on gains unobservable by the econometrician. We show which margin of agents is affected by two alternative policy changes. Finally, we unify the abundance of average effect parameters using the marginal treatment effect (Björklund and Moffitt, 1987; Heckman and Vytlacil, 2007b).

The plan of this paper is as follows. Section 2 introduces our conceptual framework and establishes the required notation. We discuss potential sources of agent heterogeneity and justify the objects of interest. Section 3 presents our empirical illustration. We outline the identification strategy, the dataset, and the estimation approach. There, we also motivate the factor structure assumption. We discuss our results as an informative example for comprehensive econometric policy evaluation. Section 4 concludes.

2 Conceptual Framework

We now present our conceptual framework to discuss the economics and econometrics of policy evaluation. We establish the required notation by introducing the generalized Roy model as a prototypical model of policy evaluation. We discuss possible sources of agent heterogeneity and examine their empirical relevance. We review common objects of interest and motivate them by the policy questions they address. Going beyond the average effects of treatment, we demonstrate the additional information provided by the whole distribution of potential outcomes.

In this section, we focus on the definition and motivation of the concepts and objects of interest. Later, we add the factor structure assumption which allows for their identification and estimation.

2.1 Prototypical Model

We rely on the generalized Roy model (Roy, 1951; Heckman and Vytlacil, 2005) throughout. We restrict the discussion to the static binary treatment case as this is the focus of most of the relevant literature.

Let $\mathbb{I}[\cdot]$ denote an indicator function that is equal to one if the corresponding condition is true and zero otherwise. Then, the generalized Roy model is characterized by the following set of equations.

Potential Outcomes:	Choice:
$Y_1 = \mu_1(X) + U_1$	$D = \mathbb{I}[\, S > 0 \,]$
$Y_0 = \mu_0(X) + U_0$	$S = \mathrm{E}[Y_1 - Y_0 - C \mid \mathcal{I}\,]$
Observed Outcome:	$C = \mu_C(Z) + U_C$
$Y = DY_1 + (1 - D)Y_0$	

(Y_1, Y_0) are objective outcomes associated with each potential treatment state D and realized after the treatment decision. Y_1 refers to the outcome in the treated state and Y_0 in the untreated state. C denotes the subjective cost of treatment participation. Any subjective benefits, e.g. job amenities, are included (as a negative contribution) in the subjective cost of treatment. Agents take up treatment D if they expect the objective benefit to outweigh the subjective cost. In that case, their subjective evaluation, i.e. the expected surplus from participation S, is positive. $\mathcal{I}$ denotes the agent's information set at the time of the participation decision. The observed outcome Y is determined in a switching-regime fashion (Quandt, 1958, 1972). If agents take up treatment, then the observed outcome Y corresponds to the outcome in the presence of treatment Y_1. Otherwise, Y_0 is observed. The unobserved potential outcome is referred to as the counterfactual outcome. We ignore general equilibrium effects and agent interactions in this setup.[1] If costs are identically zero for all agents, there are no observed regressors, and $(U_1, U_0) \sim \mathcal{N}(0, \Sigma)$, then the generalized Roy model corresponds to the original Roy model (Roy, 1951).[2]

From the perspective of the econometrician, (X, Z) are observable while (U_1, U_0, U_C) are not. X are the observed determinants of potential outcomes (Y_1, Y_0), and Z are the observed determinants of the cost of treatment C. Potential outcomes and cost are decomposed into their means $(\mu_1(X), \mu_0(X), \mu_C(Z))$ and their deviations from the mean (U_1, U_0, U_C). (X, Z) might have common elements, and the unobservables might stochastically depend on the observables. Observables and unobservables jointly determine program participation D.

If their ex ante surplus S from participation is positive, then agents select into treatment. Yet, this does not require their expected objective returns to be positive as well. Subjective cost C might be negative such that agents which expect negative returns still participate. Moreover, in the case of imperfect information, an agent's ex ante evaluation of treatment is potentially different from their ex post assessment. Agents regret their educational choice if they expect a positive surplus ex ante but the realization turns out to be negative ex post (or vice versa).

In our empirical illustration, we consider an example from educational choice. There, D takes value one if an agent pursues a higher education and zero otherwise. (Y_1, Y_0) refer to measures

[1] See Heckman et al. (1999) for simulations that assess the magnitude of potential biases from such an approach in the context of tax and tuition policy. Manski (2012) provides results for the identification of treatment responses with social interactions.

[2] Heckman (2008) presents the relationship of the Roy model to other models of potential outcomes. Imbens and Wooldrige (2009) discuss the advantages of the potential outcomes framework over a framework based directly on observed outcomes.

of subsequent labor market success. The subjective cost C of pursuing a higher education does not only involve tuition cost but psychic cost as well. The latter might include expectational error and risk aversion (Cunha et al., 2005). The realizations of (U_1, U_0) contain an agent's unobserved ability but are partly unknown to the agent at the time of the treatment decision. Therefore, agents potentially regret pursuing a higher education.

The evaluation problem arises because either Y_1 or Y_0 is observed. Thus, the effect of treatment cannot be determined on an individual level. If the treatment choice D depends on the potential outcomes, then there is also a selection problem. If that is the case, then the treated and untreated differ not only in their treatment status but in other characteristics as well. A naive comparison of the treated and untreated leads to misleading conclusions. Jointly, the evaluation and selection problem are the two fundamental problems of causal inference (Holland, 1986).

Using the setup of the generalized Roy model, we now highlight several important concepts in the economics and econometrics of policy evaluation. We discuss sources of agent heterogeneity and motivate alternative objects of interest.

2.2 Agent Heterogeneity

What gives rise to variation in choices and outcomes among, from the econometrician's perspective, otherwise observationally identical agents? This is the central question in all econometric policy analyses (Browning et al., 1999; Heckman, 2001).

The individual benefit of treatment is defined as $B = Y_1 - Y_0 = (\mu_1(X) - \mu_0(X)) + (U_1 - U_0)$. From the perspective of the econometrician, differences in benefits are the result of variation in observable X and unobservable characteristics $(U_1 - U_0)$. However, $(U_1 - U_0)$ might be (at least partly) included in the agent's information set $\mathcal{I}$ and thus known to the agent at the time of the treatment decision.

As a result, unobservable treatment effect heterogeneity can be distinguished into private information and uncertainty. Private information is only known to the agent but not the econometrician; uncertainty refers to variability that is unpredictable by both.[3]

In our empirical illustration, agents with the same observable characteristics, including their level of schooling, experience very different labor market outcomes. This variation is in part due to agents' private information about their own level of ability. However, productivity shocks in the labor market, unknown to agent and econometrician at the time of the treatment decision, play a role as well.

Cunha et al. (2005) estimate that about half of all variability in measured lifetime income is due to uncertainty realized after the decision to go to college. Another half is due to predictable components known to agents but not the econometrician. This is in line with Huggett et al. (2011), who estimate a dynamic general equilibrium model. They attribute about 40% of variation in lifetime earnings to shocks and the rest to variation in initial conditions known to the agent at age 23. Based on panel data estimates of the earnings process, Storesletten et al. (2004) assign slightly more than half of the variation to unforeseen shocks. Looking at the evolution of uncertainty over time, Cunha and Heckman (2007) document an increase in the share of earnings volatility explained by uncertainty. Nevertheless, they report considerable differences between skill groups. For less skilled workers, about 60% of the increase in wage variability is due to uncertainty, while for the higher skilled this is only 8%.

The information available to the econometrician and the agent determines the set of valid estimation approaches for the evaluation of a policy. The concept of essential heterogeneity emphasizes this point (Heckman et al., 2006b).

Essential Heterogeneity If agents select their treatment status based on benefits unobserved by the econometrician (selection on unobservables), then there is no unique effect of a treatment or a policy even after conditioning on observable characteristics. Average benefits are different from marginal benefits, and different policies select individuals at different margins. Conventional econometric methods that only account for selection on observables, like matching (Cochran and Rubin, 1973; Rosenbaum and Rubin, 1983; Heckman et al., 1998), are not able to identify any parameter of interest (Heckman and Vytlacil, 2005; Heckman et al., 2006b).

[3]See Meghir and Pistaferri (2011) for a recent overview on decomposition strategies.

Carneiro et al. (2011) present evidence on agents selecting their level of education based on their unobservable gains. They demonstrate the importance of adjusting the estimation strategy to allow for this fact. Heckman et al. (2010) propose a variety of tests for the presence of essential heterogeneity.

In our empirical illustration, we implement an estimation strategy which allows for the presence of essential heterogeneity. We show that agents in fact choose their education level based on their own unobservable returns.

2.3 Objects of Interest

Treatment effect heterogeneity requires to be precise about the effect being discussed. There is no single effect of neither a policy nor a treatment. For each specific policy question, the object of interest must be carefully defined (Heckman and Vytlacil, 2005, 2007a,b). We present several potential objects of interest and discuss what question they are suited to answer. We start with the average effect parameters. However, these neglect possible effect heterogeneity. Therefore, we explore their distributional counterparts as well.

Conventional Average Treatment Effects It is common to summarize the average benefits of treatment for different subsets of the population. In general, the focus is on the average effect in the whole population, the average treatment effect (ATE), or the average effect on the treated (TT) or untreated (TUT).

$$\begin{aligned}
ATE &= \mathrm{E}\left[Y_1 - Y_0\right] \\
TT &= \mathrm{E}\left[Y_1 - Y_0 \mid D = 1\right] \\
TUT &= \mathrm{E}\left[Y_1 - Y_0 \mid D = 0\right]
\end{aligned}$$

The relationship between these parameters depends on the assignment mechanism that matches agents to treatment. If agents select their treatment status based on their own benefits, then agents that take up treatment benefit more than those that do not and thus $TT > TUT$. If agents select their treatment status at random, then all parameters are equal.

The policy relevance of the conventional treatment effect parameters is limited. They are only informative about extreme policy alternatives. The ATE is of interest to policy makers if they weigh the possibility of moving a full economy from a baseline to an alternative state or are able to assign agents to treatment at random. The TT is informative if the complete elimination of a program already in place is considered. Conversely, if the same program is examined for compulsory participation, then the TUT is the policy relevant parameter.

To ensure a tight link between the posed policy question and the parameter of interest, Heckman and Vytlacil (2001b) propose the policy-relevant treatment effect ($PRTE$). They consider policies that do not change potential outcomes, but only affect individual choices. Thus, they account for voluntary program participation.

Policy-Relevant Average Treatment Effects The $PRTE$ captures the average change in outcomes per net person shifted by a change from a baseline state B to an alternative policy A. Let D_B and D_A denote the choice taken under the baseline and the alternative policy regime respectively. Then, observed outcomes are determined as

$$Y_B = D_B Y_1 + (1 - D_B) Y_0$$
$$Y_A = D_A Y_1 + (1 - D_A) Y_0.$$

A policy change induces some agents to change their treatment status ($D_B \neq D_A$), while others are unaffected. More formally, the $PRTE$ is then defined as

$$PRTE = \frac{1}{E[D_A] - E[D_B]} \left(E[Y_A] - \mathrm{E}[Y_B] \right).$$

In our empirical illustration, in which we consider education policies, the lack of policy relevance of the conventional effect parameters is particularly evident. Rather than directly assigning individuals a certain level of education, policy makers can only indirectly affect schooling choices, e.g. by altering tuition cost through subsidies. The individuals drawn into treatment by such a policy will neither be a random sample of the whole population, nor the whole population of the previously (un-)treated. That is why we estimate the policy-relevant effects of alternative education policies and contrast them with the conventional treatment effect parameters. We also show how the $PRTE$ varies for alternative policy proposals as different agents are induced to change their treatment status.

The average effect of a policy and the average effect of a treatment are linked by the marginal treatment effect (MTE). The MTE was introduced into the literature by Björklund and Moffitt (1987) and extended in Heckman and Vytlacil (2001a, 2005, 2007b).

Marginal Treatment Effect The MTE is the treatment effect parameter that conditions on the unobserved desire to select into treatment. Let $V = E[U_C - (U_1 - U_0) \mid \mathcal{I}]$ summarize the expectations about all unobservables determining treatment choice and let $U_S = F_V(V)$. Then, the MTE is defined as

$$MTE(x, u_S) = \mathrm{E}\left[Y_1 - Y_0 \mid X = x, U_S = u_S \right].$$

The MTE is the average benefit for persons with observable characteristics $X = x$ and unobservables $U_S = u_S$. By construction, U_S denotes the different quantiles of V. So, when varying U_S but keeping X fixed, then the MTE shows how the average benefit varies along the distribution of V. For u_S evaluation points close to zero, the MTE is the average effect of treatment for individuals with a value of V that makes them most likely to participate. The opposite is true for high values of u_S.

The MTE provides the underlying structure for all average effect parameters previously discussed. These can be derived as weighted averages of the MTE (Heckman and Vytlacil, 2005). Parameter j, $\Delta_j(x)$, can be written as

$$\Delta_j(x) = \int_0^1 MTE(x, u_S) h_j(x, u_S)\, du_S,$$

where the weights $h_j(x, u_S)$ are specific to parameter j, integrate to one, and can be constructed from data.[4] All parameters are identical only in the absence of essential heterogeneity. Then, the $MTE(x, u_S)$ is constant across the whole distribution of V as agents do not select their treatment status based on their unobservable benefits.

In our empirical illustration, we estimate the MTE of a college education. We show how the

[4]See Table 8 in Appendix B for a selection of the weights.

return varies along the unobservable margin. We also exploit its properties to organize and interpret the multiplicity of average effect parameters.

So far, we have only discussed average effect parameters. However, these conceal possible treatment effect heterogeneity, which provides important information about a treatment. Hence, we now present their distributional counterparts (Aakvik et al., 2005).

Distribution of Potential Outcomes Several interesting aspects of policies cannot be evaluated without knowing the joint distribution of potential outcomes (see Abbring and Heckman (2007) and Heckman et al. (1997)). The joint distribution of (Y_1, Y_0) allows to calculate the whole distribution of benefits. Based on it, the average treatment and policy effects can be constructed just as the median and all other quantiles. In addition, the portion of people that benefit from treatment can be calculated for the overall population $\Pr(Y_1 - Y_0 > 0)$ or among any subgroup of particular interest to policy makers $\Pr(Y_1 - Y_0 > 0 \mid X)$.[5] This is important as a treatment which is beneficial for agents on average can still be harmful for some. The absence of an average effect might be the result of part of the population having a positive effect, which is just offset by a negative effect on the rest of the population. This kind of treatment effect heterogeneity is informative as it provides the starting point for an adaptive research strategy that tries to understand the driving force behind these differences (Horwitz et al., 1996, 1997).

In our empirical illustration, we estimate the whole distribution of the returns to education. We show how a focus on average effects masks considerable heterogeneity in the returns to a college education.

[5]For a comprehensive overview on related work see Abbring and Heckman (2007) and the work they cite. The survey by Fortin et al. (2011) provides an overview about the alternative approaches to the construction of counterfactual observed outcome distributions. See Firpo (2007), Abadie et al. (2002), and Chernozhukov and Hansen (2005) for their studies of quantile treatment effects.

3 Empirical Illustration

We now illustrate the issues and concepts introduced in the previous section with an application to the returns to college. Before presenting our results, we provide a description of our identification strategy, the dataset, and our estimation approach. The choice of all three is motivated by a factor structure assumption, which we discuss first.

3.1 Factor Structure Assumption

The factor structure assumption postulates that a low dimensional vector of latent factors θ is the sole source of dependency among the unobservables of a model. Factor models are widely used to proxy latent measures of ability (see Thurstone (1934) and the large literature that followed). This motivates their use in our application as it addresses the empirical regularity that agents select their education level based on their unobserved ability.

Applied to the case of the generalized Roy model, the unobservable components determining potential outcomes and treatment choice are decomposed as:

$$U_1 = \alpha_1 \theta + \epsilon_1 \qquad U_0 = \alpha_0 \theta + \epsilon_0 \qquad U_C = \alpha_C \theta + \epsilon_C.$$

The factor loadings $(\alpha_1, \alpha_0, \alpha_C)$ may be different and thus, θ may affect choices and outcomes differently. The disturbances $(\epsilon_1, \epsilon_0, \epsilon_C)$ are an additional source of variation and assumed mutually independent and independent of the factor.

The factor structure assumption allows to solve the selection problem and is essential for the estimation of the joint distribution of potential outcomes. All the dependencies between the unobservables of the model are driven by θ and conditioning on it allows to construct the agents' counterfactual state experience. At the same time, θ provides the link between the two marginal outcome distributions $(F_{Y_1|D=1}(\cdot), F_{Y_0|D=0}(\cdot))$, which can be constructed from the observed data. Through this link, the joint distribution of potential outcomes $(F_{Y_1,Y_0}(\cdot))$ can be recovered.

The factor structure assumption permits asymmetries in the information structure between agent and econometrician. Agents select their treatment status based on their expected surplus from treatment given the information available to them at the time of treatment decision $\mathcal{I}$. We allow that (θ, ϵ_C) are private information to the agent while (ϵ_1, ϵ_0) are not. The latter reflect uncertain fluctuations in future labor market outcomes. The econometrician observes neither θ nor $(\epsilon_1, \epsilon_0, \epsilon_C)$.

Discrepancies between an agent's ex ante and ex post evaluation of treatment participation arise due to the realizations of (ϵ_1, ϵ_0). These are unknown to the agent at the time of the treatment decision but affect potential outcomes. Thus, unexpected realizations of (ϵ_1, ϵ_0) might lead agents to regret their treatment choice ex post.

3.2 Identification

Conditions for nonparametric identification of the generalized Roy model are presented in Heckman and Vytlacil (2007b).[6] They rely on the availability of exclusion restrictions, i.e. variables that only affect choices but not potential outcomes, and support conditions. This approach is called "identification at infinity" (e.g. Chamberlain (1986) and Heckman (1990)) and requires the existence of limit sets where the probability of treatment participation is either zero or one. Within these limit sets there is no selection, and thus $F_{U_1}(\cdot)$ and $F_{U_0}(\cdot)$ can be recovered. However, this is not enough to determine the joint distribution of potential outcomes. Imposing a factor structure and adding a set of measurement equations on θ permits identification of the joint dependencies among the unobservables of the model under the conditions outlined in Cunha et al. (2010). The measurements provide a signal about θ but also contain noise due to additional disturbances. Nonetheless, orthogonality conditions allow to separate the noise from the signal and to identify the distribution of θ. With this distribution at hand, the joint distribution of (U_1, U_0) can be recovered.[7]

3.3 Data and Estimation Strategy

We use the National Longitudinal Survey of Youth of 1979 (NLSY79) for our empirical illustration. The NLSY79 is a nationally representative sample for the United States of 12,686 young

[6] French and Taber (2011) provide an instructive discussion about alternative identification approaches to the different versions of the Roy model.

[7] For a review of the alternative identification strategies of the joint distribution of potential outcomes, see Abbring and Heckman (2007).

men and women who were 14 to 22 years of age when first surveyed in 1979. The cohort was interviewed annually through 1994. Since then, the survey has been administered biennially.[8] We restrict our sample to white males only. The NLSY79 has an oversample of poor whites and a military sample. We exclude both from our analysis.

The sample was originally prepared for the analysis in Carneiro et al. (2011). We extend it to fit the data requirements of a factor structure model by adding a measurement system to identify the distribution of ability θ.

We estimate a simplified version of the generalized Roy model to investigate the returns to college. Our estimation strategy exploits a variety of separability, linearity, independence, and distributional assumptions. We now present these in detail.

Potential Outcome Model We use the natural logarithm of hourly wages between 1989 and 1993 (individuals are between 28 and 34 years of age in 1991) to determine the return to a college education. We specify a log-linear model per period $t = 1, .., 5$ for each education group. Y_{1t} denotes the outcome in the treated state in period t and Y_{0t} in the untreated state for the same period. Both outcomes are determined by a vector of observable characteristics X with education group and period-specific parameter vectors $\{\beta_{1t}, \beta_{0t}\}$. In addition, outcomes in both states are affected by cognitive ability θ, but potentially to a different extent as determined by the factor loadings $\{\alpha_{1t}, \alpha_{0t}\}$. The idiosyncratic error terms $\{\epsilon_{1t}, \epsilon_{0t}\}$ follow a normal distribution with mean zero and variances $\{\sigma^2_{\epsilon_{1t}}, \sigma^2_{\epsilon_{0t}}\}$.

$$
\begin{aligned}
Y_{1t} &= X\beta_{1t} + \alpha_{1t}\theta + \epsilon_{1t} \qquad \text{with} \quad \epsilon_{1t} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{1t}}) \\
Y_{0t} &= X\beta_{0t} + \alpha_{0t}\theta + \epsilon_{0t} \qquad \text{with} \quad \epsilon_{0t} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{0t}})
\end{aligned}
$$

The unobservables $(\theta, \{\epsilon_{1t}, \epsilon_{0t}\})$ are independent from the observables X. The idiosyncratic components $\{\epsilon_{1t}, \epsilon_{0t}\}$ are independent within and across time and independent of the factor θ. Conditional on the set of control variables X and the factor θ, the estimation simplifies to a normal linear regression model by treatment status.

[8] See Bureau of Labor Statistics (2001) for a detailed description of the NLSY79.

As determinants of log earnings, we specify linear and squared terms for years of true work experience, mother's years of schooling, number of siblings, as well as a dummy variable indicating urban residence at age 14, cohort dummies, and a factor of cognitive ability. We also include linear terms of current (at the time of the outcome in 1991) local wages and local unemployment as well as their long run averages between 1973 and 2000. In what follows, we refer to the long run averages as permanent local wages and permanent local unemployment.

Educational Choice Model We separate individuals in two groups: D = 0 (high school dropouts and high school graduates) and D = 1 (individuals with some college, college graduates, and post-graduates). We specify a linear-in-parameters binary choice model. The schooling decision D depends on the vector X_I that also affects potential outcomes. X_I is only a subset of X as not all components of X are known to the agent at the time of the treatment decision. A vector Z contains a set of observables that affect the subjective cost of treatment participation. X_I and Z contain common elements. γ_X parameterizes the marginal effects of X_I and γ_Z of Z. In addition, the treatment choice depends on cognitive ability θ with loading γ_θ.

$$D = \mathbb{I}\,[\, X_I\gamma_X - Z\gamma_Z + \theta\gamma_\theta - \epsilon_C > 0\,] \qquad \text{with} \quad \epsilon_C \sim \mathcal{N}(0,1)$$

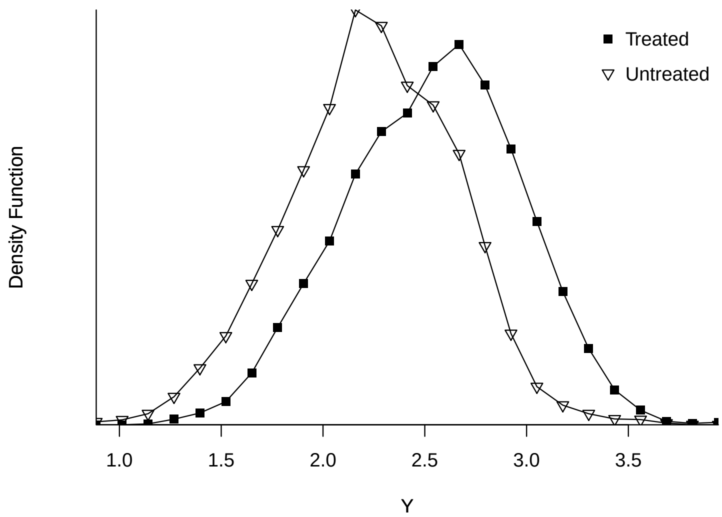
The unobservables (θ, ϵ_c) are independent from the observables (X_I, Z). ϵ_c follows a standard normal distribution and is independent of the factor θ. Conditional on (X_I, Z) and θ, the estimation follows a Probit response model.

We include the covariates that determine potential outcomes (excluding actual work experience and labor market conditions at the time of the outcome realization) in (X_I, Z). We also specify several exclusion restrictions, i.e. variables that only affect subjective cost and are only part of Z. For this purpose, we include local labor market conditions, distance to college, and tuition cost. We use past (at the time of treatment decision at age 17) local wages and local unemployment to capture the local labor market conditions, the presence of a four-year college as a measure of distance to college, and average tuition in public four-year colleges to reflect the direct financial cost of college attendance. All exclusion restrictions are interacted with mother's education and number of siblings.

The validity of the exclusion restrictions hinges on the fact that they are not correlated with the unobservables in the wage equations for the adult years. This is questionable for the local labor market conditions at the time of treatment choice: they might be correlated with the long run economic environment. Following Carneiro et al. (2011), we address this concern as we include measures of permanent local labor market conditions (e.g. average wages and unemployment between 1973 and 2000 for each location of residence at 17). In this setup, only the innovations in the local labor market variables are used as exclusions.

Figure 1 depicts the density function of average log hourly wages over the five year period in our data by treatment status. In our final sample, 45% of the agents pursue a higher education and have on average four more years of schooling. The high-educated earn on average 2.53, while earnings are lower with 2.24 among the low educated. Given the average difference of four years of education between the two groups, this amounts to an annual return of 7.3%. However, this raw difference is not due to schooling alone. If agents select into treatment based on their returns, they differ in other important aspects besides their level of schooling.

Figure 1: Distribution of Observed Outcome



Notes: Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992).

Furthermore, there is considerable heterogeneity in outcomes within each treatment group.

Among the treated, earnings range from 1.97 at the second decile to 3.06 at the eighth decile. For the untreated, earnings at the second decile amount to 1.78 and go up to 2.73 at the eighth decile. In fact, 24% of the untreated earn more than the average among the treated.

Measurement System We take the measurements on θ from the Armed Service Vocational Aptitude Battery (ASVB), which is described in Deparment of Defense (1982). We use the Armed Forces Qualification Test (AFQT), which consists of the following subtests: word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge. The AFQT (sub-) scores are frequently used to account for an individual's ability as a determinant for a variety of economic and social outcomes (Herrnstein and Murray, 1994; Heckman et al., 2006a; Carneiro et al., 2011). The subscores are corrected for the fact that individuals have different amounts of schooling at the time they take the test following the procedure developed in Hansen et al. (2004).

Measurement M_j on θ with $j = 1, ..., 4$ is determined by a set of observable characteristics W and cognitive ability θ. Both translate differently into each measure as parametrized by $\{\psi_j, \delta_j\}$. The unobservables $(\theta, \{\nu_j\})$ are independent from the observables W. The idiosyncratic components $\{\nu_j\}$ are independent of each other and the cognitive factor θ. These independence assumptions allow to extract the noise from the signal. Conditional on W the covariation between measurements is due to the common factor θ only.

$$\begin{aligned}
M_1 &= W\psi_1 + \delta_1\theta + \nu_1 \qquad && \text{with} \quad \nu_1 \sim \mathcal{N}(0, \sigma^2_{\nu_1}) \\
&\vdots \\
M_4 &= W\psi_4 + \delta_4\theta + \nu_4 \qquad && \text{with} \quad \nu_4 \sim \mathcal{N}(0, \sigma^2_{\nu_4})
\end{aligned}$$

The idiosyncratic error terms $\{\nu_j\}$ follow a normal distribution. Conditional on θ and W, the estimation of each measurement equation is carried out as a normal linear regression model. To set the scale of θ, we fix one of the factor loadings to one.

Borghans et al. (2008) emphasize the need to standardize the incentives for and the environment of achievement tests. We follow their advice and control for differences in test-taking behavior by observable characteristics. We model these differences by linear and squared terms in maternal education and the number of siblings.

Table 1: Measurements

Measures	All	Treated	Untreated
Arithmetic Reasoning	0.000	0.355	-0.335
Word Knowledge	0.000	0.287	-0.271
Paragraph Composition	0.000	0.300	-0.284
Math Knowledge	0.000	0.487	-0.460

Notes: Final sample consists of a total of 1,287 white males, where 625 did receive some college education while 662 do not. Measures standardized to mean zero and standard deviation one in the final sample.

Table 1 shows the average value for each measure by treatment status in our sample. They are standardized to mean zero and standard deviation one. The averages of all subscores are at least half a standard deviation higher for the agents with an advanced level of education. This contrast is most pronounced for math knowledge and smallest for word knowledge.

Table 2: Specification

Covariates	Outcomes	Choice	Measures
Years of Experience	X		
Current Local Wages	X		
Current Local Unemployment	X		
Permanent Local Unemployment	X	X	
Permanent Local Wages	X	X	
Mother's Years of Schooling	X	X	X
Number of Siblings	X	X	X
Urban Residence	X	X	
Cohort Dummies	X	X	
Factor of Cognitive Ability	X	X	X
Local Presence of Public College		X	
Local Tuition at Public College		X	
Past Local Wages		X	
Past Local Unemployment		X	

Notes: Specification includes squared terms in experience, number of siblings, mother's education, permanent labor market conditions, and interactions of the exclusion restrictions with number of siblings and mothers' education. Final sample consists of a total of 1,287 white males, where 625 did receive some college education while 662 do not.

Distribution of Skills The distribution of cognitive ability is approximated by a normal finite mixture model (Diebolt and Robert, 1994). Mixtures of normals with a large enough number of components approximate any distribution (Ferguson, 1983) and are frequently used as a flexible semiparametric approach to density estimation (Escobar and West, 1995; Frühwirth-Schnatter, 2006). The unobservable factor θ is distributed as a univariate mixture of K normals with share parameter π_k, mean μ_k, and variance σ_k^2,

$$\theta \sim \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mu_k, \sigma_k^2\right),$$

where $\sum_{k=1}^{K} \pi_k = 1$ and $\sum_{k=1}^{K} \pi_k \mu_k = 0$. We estimate a mixture model for θ with $K = 3$ components.

Table 2 summarizes the covariates used in our specification. Additional descriptive statistics and details about the construction of the dataset are provided in Appendix C. Next, we outline our estimation strategy.

We collect all parameters of the model in Ψ. Conditional on θ and the relevant observables, the observed outcome, choice, and measurement equations are all independent. Thus, the individual likelihood can be written as

$$\begin{aligned}\mathcal{L}(\Psi) = &\int_{\Theta} \prod_{d=0}^{1} \left\{ \Pr(D = d \mid X, Z, \theta; \Psi) \prod_{t=1}^{5} f(Y_{dt} \mid X, \theta; \Psi) \right\}^{\mathbb{I}[D=d]} \\ &\times \prod_{j=1}^{4} f(M_j \mid W, \theta; \Psi) dF_{\theta}(\theta),\end{aligned}$$

where $f(\cdot)$ denotes a density function, and $F_{\theta}(\cdot)$ is the cumulative distribution function of the latent factor θ over the support Θ.

θ needs to be integrated out of the individual likelihood, which leads to a complex nature of the likelihood function. That is why we implement a full Bayesian approach for the estimation of the model and rely on Markov Chain Monte Carlo (MCMC) techniques.[9] The Gibbs sampler, which proceeds by simulating each parameter (or parameter block) from its conditional distribution, is particularly appropriate for this kind of problem (Casella and George, 1992). For the model of educational choice, we rely on the data augmentation approach following Albert and Chib (1993). We run a chain of 1,030,000 iterations. After a burn-in period of 30,000 iterations, we save the draws from every 100^{th} iteration. The resulting 10,000 iterations are used for postestimation inference.

We generate a simulated sample of $100,000$ agents and collect them in the set N. First, we fix all estimated parameters to their posterior means. Second, we draw a set of observable characteristics (X, X_I, Z) with replacement from the original dataset. Third, we simulate the

[9]See Chib (2001) for an overview on MCMC techniques and their use in econometrics and Heckman et al. (2012) for a broad discussion of their use for the estimation of treatment effect in factor models. Piatek (2010) provides the required technical details in the framework of a factor structure model. See Table 9 in Appendix B for the specification of the priors.

unobservables of the model $(\theta, \{\epsilon_{1t}, \epsilon_{0t}\}, \epsilon_c)$. Together, this allows us construct potential outcomes $\{Y_{1t}, Y_{0t}\}$, surplus S_B and treatment choice D_B in the baseline state, and the individual effect of treatment $\{B_t\}$.

We will also consider two policy alternatives $j = 1, 2$. We construct the counterfactual surplus and choice $\{S_{Aj}, D_{Aj}\}$ by modifying Z to Z_{Aj} for each policy alternative.

We present our results as the average over the five time periods to reduce the impact of transitory earnings fluctuations. Thus, we drop the t subscript. In addition, we annualize our estimates of the returns to college by dividing our results by four. This is the average difference in years of schooling between the treated and untreated.

We end up with the following simulated sample:

$$\{Y_{1i}, Y_{0i}, B_i, X_i, X_{Ii}, Z_i, \{Z_{Aji}\}, \{S_{Aji}, D_{Aji}\}, D_{Bi}, \theta_i, \epsilon_{1i}, \epsilon_{0i}, \epsilon_{Ci}\} \qquad \forall \quad i \in N.$$

3.4 Results

We now turn to the presentation and discussion of our results. We start by showing the quality of our model. Then, we establish that agents select their educational attainment based on returns, which are at least partly unobservable by the econometrician. We report the conventional average treatment effects and contrast them with the policy-relevant average treatment effects. We exploit the fact that both these parameters can be expressed as weighted averages of the marginal treatment effect to interpret their differences. Finally, we go beyond the average effects of a treatment and a policy by presenting their whole distribution.

3.4.1 Model Quality

Model Fit Table 3 compares the distribution of actual earnings to its simulated counterpart. Mean and median of the two samples are nearly identical. The standard deviation of the simulated sample is slightly smaller compared to the actual sample. This is due to the thinner tails of the simulated distribution. Overall, the model fits the observed data quite well.

Table 3: Model Fit

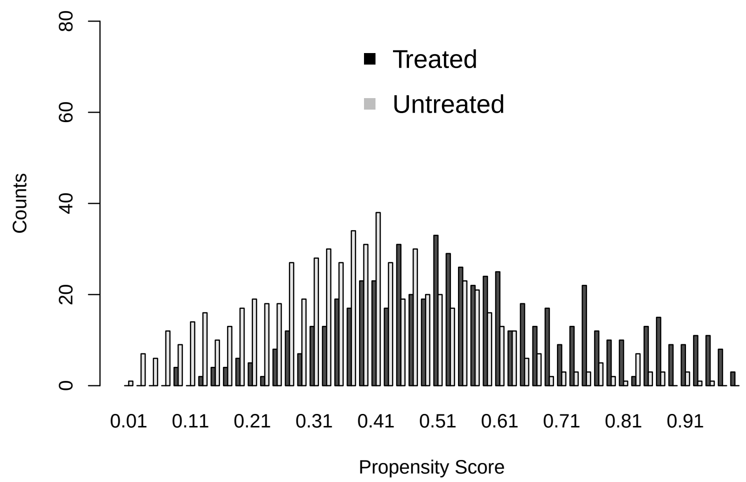
Outcome					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.385	0.434	2.013	2.388	2.753
Model	2.393	0.336	2.110	2.393	2.676

Notes: based on 100,000 simulated agents and 1,287 actual agents, Sd. = Standard Deviation.

Our estimation strategy imposed a variety of functional form and distributional assumptions. However, as we included several exclusion restrictions in our specification, a much more flexible model is still identified. The range of common support plays a central role in this context.

Common Support Nonparametric identification of the model relies on the existence of exclusion restrictions and "identification at infinity" arguments. Empirically, the latter requires that some agents select treatment with probability one or zero. Figure 2 displays the support of the estimated probability of treatment participation, i.e. the propensity score, in the actual sample. Among the treated, the support ranges from 0.07 to 0.99. For the untreated, the range of support is slightly shifted to the left. It starts at 0.01 and extends up to 0.95. Thus, the range of common support is close to the full unit interval. An "identification at infinity" strategy is valid in our data.

Figure 2: Common Support



Notes: Counts of the choice probabilities in the actual sample.

3.4.2 Selection on Unobservables

The factor structure assumption allows for an explicit exposition of selection on returns that are unobservable by the econometrician. The unobservable returns $(U_1 - U_0)$ and the unobservable dislike for treatment participation V can be decomposed as

$$\begin{aligned} U_1 - U_0 &= (\alpha_1 - \alpha_0)\theta + (\epsilon_1 - \epsilon_0) \\ V &= -\gamma_\theta \theta + \epsilon_C. \end{aligned}$$

Private information θ and uncertainty $(\epsilon_1 - \epsilon_0)$ jointly generate unobservable variability in the returns to college. The agent's private information about θ induces a dependency between V and $(U_1 - U_0)$ and creates selection on unobservables.

We estimate $\gamma_\theta > 0$, so that the likelihood of obtaining a higher education increases with ability. Thus, the treated and untreated differ systematically in their realizations of θ. Figure 3 shows the density function of the simulated distribution of ability by treatment status. On average, unobserved ability is higher among the treated. Yet, there is considerable heterogeneity within each treatment group. Among the untreated, about 23% have a higher level of cognitive ability than the average treated individual.

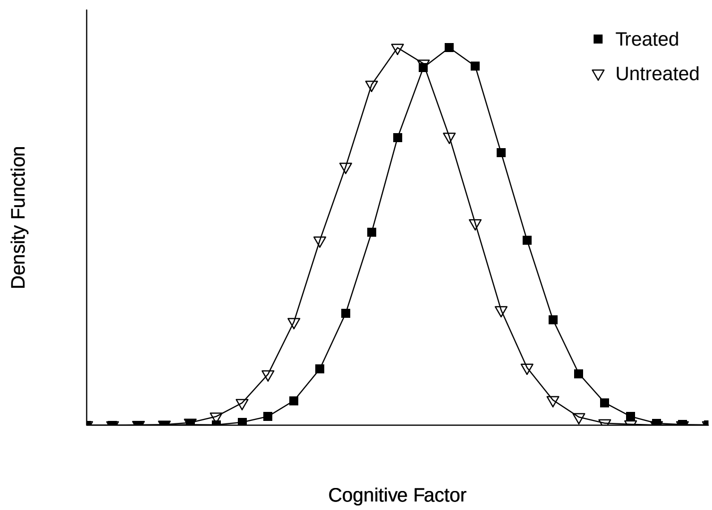
Figure 4 plots the marginal effect (ME) of cognitive ability on average wages and the probability of a college education along the quantiles q_θ of the distribution of θ. These are computed based on the simulation as follows

$$\begin{aligned} ME_{Y_1}(X = \bar{x}, q_\theta = q) &= \bar{x}\beta_1 + \alpha_1 F_\theta^{-1}(q) \\ ME_{Y_0}(X = \bar{x}, q_\theta = q) &= \bar{x}\beta_0 + \alpha_0 F_\theta^{-1}(q) \\ ME_D(X_I = \bar{x}_I, Z = \bar{z}, q_\theta = q) &= \Phi\left(\bar{x}_I\gamma_X - \bar{z}\gamma_Z + \gamma_\theta F_\theta^{-1}(q)\right), \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and $F_\theta^{-1}(\cdot)$ denotes the quantile function of the distribution of θ. Throughout, the observable characteristics (X, X_I, Z) are fixed at their mean values $(\bar{x}, \bar{x}_I, \bar{z})$.

Cognitive ability affects educational choice and wages in both potential outcome states. Two patterns emerge. First, the effect of ability on college choice is positive $(\gamma_\theta > 0)$ and quite

Figure 3: Distribution of Cognitive Skills by Treatment Status

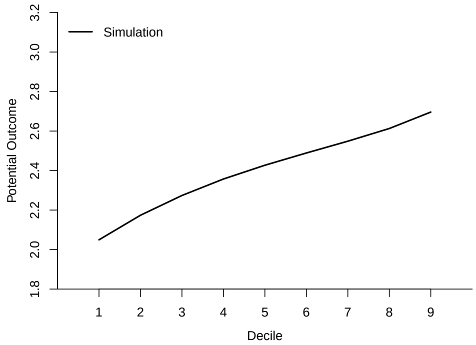


Notes: Sample based on 100,000 simulated agents. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992).

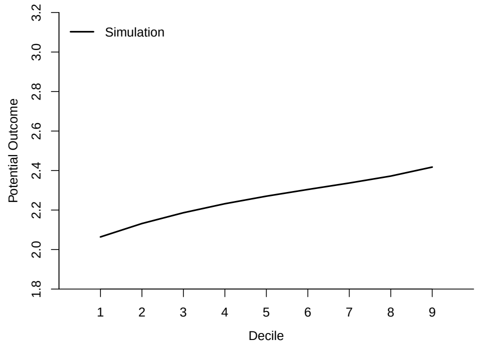
strong. The probability of a college education increases from 10% to 82% when moving an individual from the bottom to the top decile of the ability distribution. Second, the effect of ability on earnings differs between the two potential outcome states ($\alpha_1 \neq \alpha_0$). The returns to ability are higher in the treated state compared to the untreated state ($\alpha_1 - \alpha_0 > 0$). In the treated state, an increase in ability from the bottom to the top decile yields an increase in wages by 64%. In the untreated state, wages only increase by 35%. Furthermore, earnings in the treated state are usually higher when moving along the ability distribution. This is not the case in the first decile. An individual with such a low level of ability has, on average, higher earnings in the untreated state.

θ is unobservable by the econometrician and affects returns and treatment choice. As a result, agents are selecting their treatment status based on unobservable returns. Thus, essential heterogeneity is present in our data. Any estimation strategy that does not take this into account results in biased estimates.

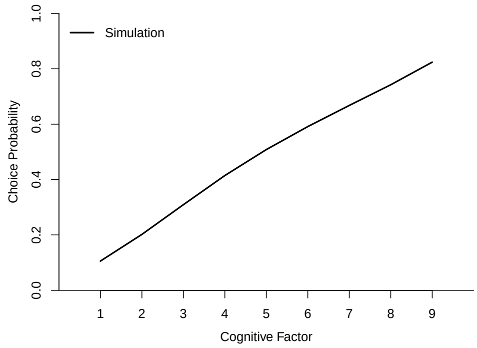
Figure 4: Marginal Effects of Ability



(a) Treated State



(b) Untreated State



(c) Choice Model

Notes: Sample based on 100,000 simulated agents.

3.4.3 Average Effects

Our factor structure implementation of the generalized Roy model allows for observable and unobservable treatment effect heterogeneity. Thus, different treatment effect parameters answer different policy questions. At first, we present the conventional effects of treatment and contrast them to the policy relevant effects. Afterwards, we use the unifying properties of the marginal treatment effect to reconcile their differences.

Conventional Average Treatment Effects Table 4 presents the conventional average treatment effects. Based on the simulated sample, we can calculate the average treatment effect as the mean difference in potential outcomes in the full sample.

$$ATE = \frac{1}{|N|} \sum_{i \in N} (Y_{1i} - Y_{0i})$$

The TT and the TUT are determined by separate calculations among the group of the treated ($D_B = 1$) and untreated ($D_B = 0$) respectively.

Table 4:
Conventional
Average Treatment
Effects

Population	Effect
All	0.035
Treated	0.047
Untreated	0.026

Notes: Sample based on 100,000 simulated agents.

On average, the return to education is 3.5% for each additional year of schooling. Among the treated, returns are higher than average and amount to 4.7%. For the untreated, returns are considerably lower with only 2.6% on average. Thus, the agents who pursue a higher education have the most to gain.

Nevertheless, returns for the untreated are positive. But still, they do not pursue a higher education. Their subjective cost must be so high that the positive returns are not high enough. As a result, their expected surplus S_B remains negative.

The average return is less than half of the 7.3% raw difference in outcomes. So, systematic differences in observable and unobservable characteristics drive the difference in raw returns.

As discussed in Section 2, the conventional treatment effects are only informative about extreme policy alternatives. That is why we turn to the policy-relevant treatment effects next.

Policy-Relevant Average Treatment Effects We consider two generic policy alternatives:[10]

- **Policy Alternative A:** Building public colleges in all counties which do not yet provide one.

[10]For the purposes of this paper, we abstract from balanced budget considerations.

- **Policy Alternative B:** Equalization of tuition fees in all existing public colleges to their mean value.

Table 5 presents the $PRTE$, i.e. the change in the average outcomes per net person shifted, for each of the two policy alternatives. Let $\mathcal{P}_j$ denote the set of agents which are induced to change their treatment status, i.e. $D_B \neq D_{Aj}$, due to policy j. Then, we can calculate the overall $PRTE_j$ for each policy alternative as the average difference in potential outcomes among the agents in $\mathcal{P}_j$.

$$PRTE_j = \frac{1}{|\mathcal{P}_j|} \sum_{i \in \mathcal{P}_j} (Y_{1i} - Y_{0i})$$

Along the same line, we can separately determine the average effect among the agents that enter or withdraw from treatment.

Table 5: Policy-Relevant Average Treatment Effects

Population	Policy A	Policy B
All	0.032	-0.002
Entering	0.032	0.034
Withdrawing	—	-0.036

Notes: Sample based on 100,000 simulated agents.

Policy A only affects agents living in counties that do not yet provide a public college and makes college attendance more likely for this group. Among the whole population, 3.5% of agents revise their treatment decision and now intend to pursue a higher education. On average, these agents realize a return of 3.2%.

Policy B has more heterogeneous impacts. Agents who face a college with costs higher than average will experience a reduction of tuition fees compared to the baseline state. However, tuition fees at cheaper colleges will rise. Overall, the impact of Policy B is less pronounced. Only 1.7% of the population alter their treatment choice. Among those, about 0.9% enter treatment while another 0.8% withdraw. Both groups experience very similar returns of about 3.5% on average. Thus, the overall effect on observed outcomes is negligible as their realized returns just cancel out.

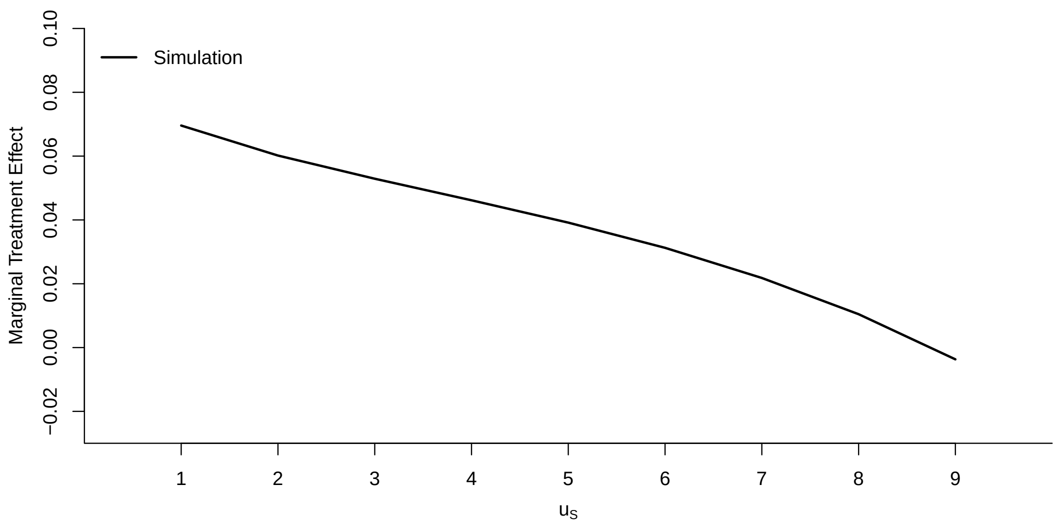
Both, the conventional and policy relevant treatment effects capture average effects of treatment and policies. And yet, they differ. The marginal treatment effect (MTE) allows to investigate these differences further. We start with a discussion of the MTE itself and then use its unifying properties to explain the differences between the average effect parameters.

Marginal Treatment Effect Figure 5 presents the MTE, which shows the average benefit of treatment along the distribution of V for fixed $X = \bar{x}$. Recalling that $V = -\gamma_\theta \theta + \epsilon_C$ and the definition of $U_S = F_V(V)$, the MTE can be calculated as

$$MTE(X = \bar{x}, U_S = u_S) = \bar{x}(\beta_1 - \beta_0) - \frac{(\alpha_1 - \alpha_0)}{\gamma_\theta} F_\theta^{-1}(u_s).$$

Thus, the MTE allows to examine how the returns to college vary for different margins of V.

Figure 5: Marginal Treatment Effect



Notes: Sample based on 100,000 simulated agents.

Benefits range from 6.8% at the bottom decile of V to 0.0% at the top decile. Agents who are most likely to take up treatment, i.e. those with a low level of V, have the most to gain. Their private information about their relatively high level of ability θ results in higher expected returns.

The MTE allows to rationalize the differences between the numerous average effect parameters (Heckman and Vytlacil, 2005, 2001b). All parameters are weighted averages of the MTE, but each weighs parts of the distribution of V differently.

Table 6 compares the multiple average effects of treatment. The $PRTE_A$ for Policy A is very close to the ATE. For Policy B, the $PRTE_B$ is zero as the effects for those agents induced to change their treatment status cancel out. The average effects among the affected subgroups by either policy are less pronounced than the TUT or TT.

Table 6: Comparing the Effects of Treatment

Conventional		Policy-Relevant		
Population	Estimate	Population	Policy A	Policy B
All	0.035	All	0.032	-0.002
Treated	0.047	Entering	0.032	0.034
Untreated	0.026	Withdrawing	—	-0.036

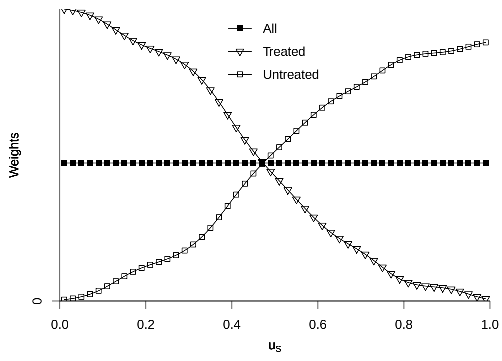
Notes: Sample based on 100,000 simulated agents.

Figure 6 shows the empirical weights for the average effect parameters.[11] First, we discuss the weights for the conventional parameters. The ATE samples evenly across the whole distribution of V, whereas the TT oversamples agents with a high probability of treatment participation. The opposite is true for the TUT, which puts larger weight on individuals with high values of V. This makes them unlikely to take up treatment. Second, we turn to the weights for the policy-relevant parameters. Policy A accentuates the tails of the distribution of V, while Policy B stresses the middle part. Accordingly, different policies affect different margins of V. Notably, the weights for the policy-relevant treatment effects are not necessarily positive. For both policies, some parts of the distribution of V receive a negative weight.

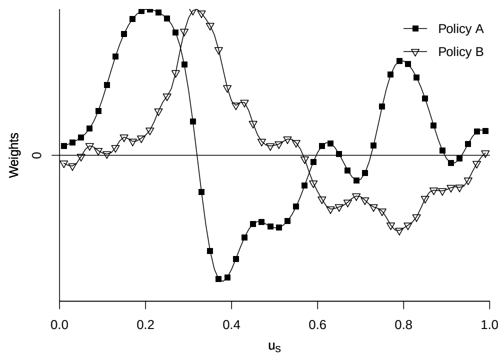
So far, the focus of the discussion has been on average effect parameters. However, these mask considerable treatment effect heterogeneity. That is why we discuss their distributional

[11] The weights vary for different realizations of X and integrate to one by construction. Since $\bar{x}$ is a high dimensional vector, it is not computationally feasible to condition on it. Instead, as an approximation, we condition on the index $\bar{x}(\beta_1 - \beta_0)$.

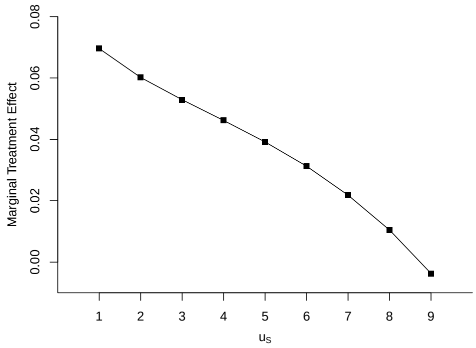
Figure 6: Weights



(a) Conventional



(b) Policy-Relevant



(c) Marginal Treatment Effect

Notes: (a) and **(b)** depict conditional density estimates using he method of Hall et al. (2004) based on a sample of 100,000 simulated agents. The weights are scaled to fit the picture. **(c)** based on simulation from the estimates of the model. Observable characteristics X set to their mean values in the sample.

counterparts next.

3.4.4 Distributional Effects

We are able to recover the joint distribution of potential outcomes due to the factor structure assumption. With this joint distribution at hand, we can calculate the marginal distribution of benefits, the joint distribution of benefits and surplus, and the marginal distribution of policy effects. We discuss each in turn.

Joint Distribution of Potential Outcomes Figure 7 presents the results for the joint distribution of potential outcomes $F_{Y_1,Y_0}(y_1, y_0)$ in a contour plot.

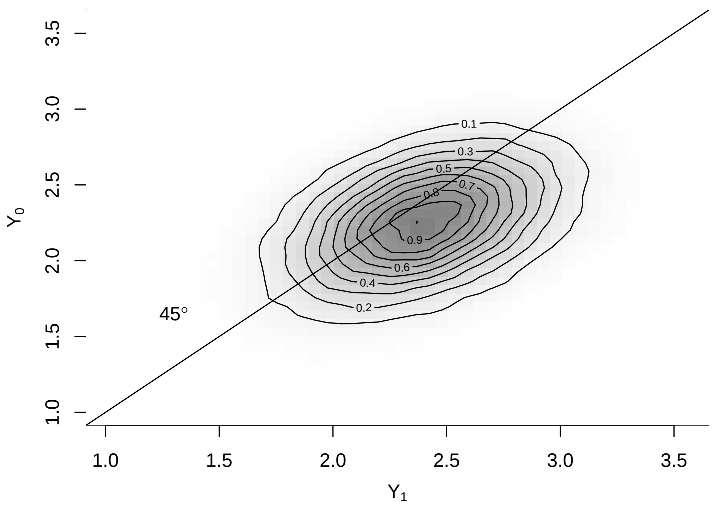
The surface of the plot is directed at the top right corner. Thus, potential outcomes are positively correlated. Agents who do well in one of the education groups also tend to do well in the other. The 45° degree line separates the agents with positive and negative returns. Below the straight line, benefits are positive as the potential outcome in the treated state Y_1 is higher than in the untreated state Y_0. Above, the opposite is true. It becomes clear that there is significant treatment effect heterogeneity and a considerable share of agents has negative returns to education.

Next, we investigate this in more detail by looking at the marginal distribution of benefits directly.

Marginal Distribution of Benefits Figure 8 shows the marginal distribution of benefits $F_B(b)$. Benefits range from -7% at the first decile to +14% at the ninth decile of the distribution. Mean and median benefits are very similar with 3.5% each. Roughly 34% of agents exhibit negative returns to education. Plotting the conditional distributions by treatment status reveals only a slight shift between the two. Nevertheless, the share of agents with a negative return is considerably smaller among the treated (30%) than the untreated (38%).

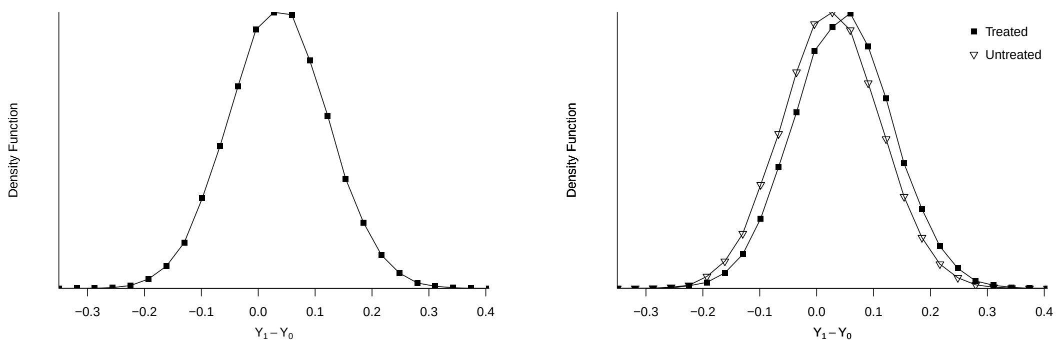
Still, even among the treated, a quite considerable share of agents exhibits negative returns. Among them, there are two groups. First, some agents expected negative returns but have negative subjective cost of education so that they pursue a higher education anyway. Second, some agents expected positive returns ex ante but realized unfavorable draws of (ϵ_1, ϵ_0) ex post. There is a tight link to the conventional average treatment effects reported in Table 4. They

Figure 7: Joint Distribution of Potential Outcomes



Notes: Sample based on 100,000 simulated agents. Two-dimensional kernel density estimation with an axis-aligned bivariate normal kernel, evaluated on a square grid (Venables and Ripley, 2002).

Figure 8: Distribution of Gains



(a) Overall **(b)** Treatment Status

Notes: Sample based on 100,000 simulated agents. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992).

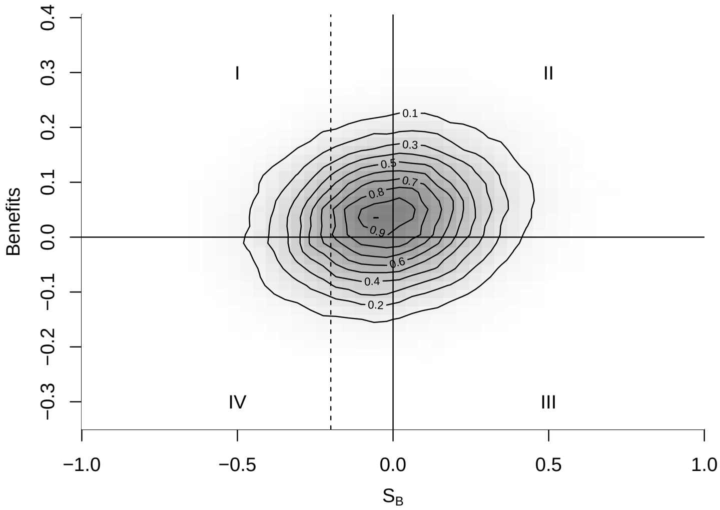
correspond to the mean values of the respective distribution. All the heterogeneity in returns remains unnoticed by a focus on average effects only. But this heterogeneity requires to be precise about the effect of a treatment and the effect of a policy. Different policies select agents at different margins with different benefits from program participation. In this regard, the joint distribution of benefits and surplus offers some illuminating insights.

Joint Distribution of Benefits and Surplus Recall that the surplus S_B from treatment participation is derived as:

$$S_B = X_I\gamma_X - Z\gamma_Z + \gamma_\theta\theta - \epsilon_C.$$

Figure 9 presents a contour plot of the joint distribution $F_{B,S_B}(b,s)$ of benefits and the surplus in the baseline state (up to the scale normalization).

Figure 9: Joint Distribution of Surplus and Benefits



Notes: Sample based on 100,000 simulated agents. Two-dimensional kernel density estimation with an axis-aligned bivariate normal kernel, evaluated on a square grid (Venables and Ripley, 2002).

The graph is separated into four distinct quadrants by the two solid lines. Agents with a positive surplus (II + III) take up treatment while those with a negative surplus (I + IV) do not. Among both groups, some show negative returns (III for the treated and IV for the untreated). Again, there is a direct link to the conventional average treatment effects. The TT is the mean return among those agents where $S_B > 0$, the TUT corresponds to the average return where $S_B < 0$.

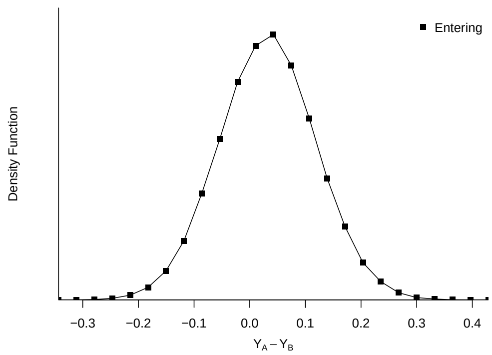
However, Figure 9 is most informative on how different policies affect different margins. In the baseline state, agents with $S_B > 0$ select into treatment while those with $S_B < 0$ do not. An alternative policy regime affects the surplus calculation. Let us consider Policy A as an example. Under the new policy regime, the subjective cost of treatment is reduced for agents who previously lived in a county without a public college. Among those, agents with $S_B > 0$ in the baseline state will not change their treatment choice. However, agents for which S_B was only slightly negative might. In this example, the agents located between the dashed and solid vertical line will change their treatment status. They do so as under the alternative policy regime their expected benefits outweigh the reduced subjective cost. The $PRTE_A$ reflects the average returns for this subset of agents. But, by exploiting the factor structure assumption, we can determine the whole distribution of policy effects.

Marginal Distribution of Policy Effects Figure 10 shows the distribution of benefits among the agents that are affected by the two policies $F_{B|\mathcal{P}_j}(b)$.

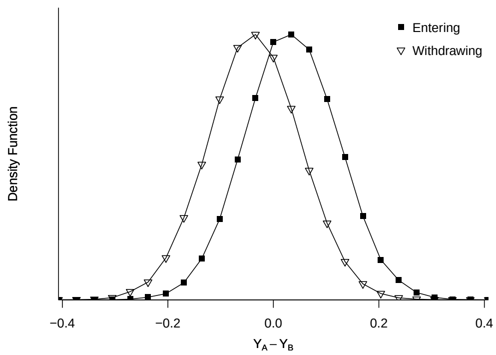
For Policy A, the overall average effect was very similar to the ATE. The same is true for the whole distribution of policy effects. The overall distribution (Figure 8) and the distribution of benefits realized due to the policy change are very much alike. For Policy B this is also the case. There, the shift between the two distributions results from the switch in sign for agents withdrawing from treatment.

The average values of the distributions yield the $PRTE_j$'s as reported in Table 5. A positive average effect is still in line with some agents experiencing negative returns (Policy A). A negligible overall average effect does not rule out considerable heterogeneity in the effects of a policy (Policy B).

Figure 10: Distribution of Policy Effects



(a) Policy A



(b) Policy B

Notes: Sample based on 100,000 simulated agents. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992).

4 Conclusion

The combination of the generalized Roy model with a factor structure assumption allowed for an instructive discussion of the economics and econometrics of policy evaluation. We explored sources of agent heterogeneity, examined resulting treatment effect heterogeneity, and clarified the distinction between the effects of a treatment and a policy.

We used an application to the returns to college as an empirical illustration. We reported average returns but also estimated their whole distribution. We found that agents select their treatment status based on returns unobservable by the econometrician. We also showed how different parameters answer different policy questions.

However, we only provided a discussion *within* the framework of a factor structure model. Yet, this is just one element in the econometrician's toolkit for policy evaluation. Alternative methods (matching, instrumental variables, regression discontinuity design, etc.) differ in their data requirements, assumptions about the sources of agent heterogeneity, simplifications required for their empirical feasibility, and policy questions they are suited to answer. Further research should focus on a comparison *between* these alternatives for a given policy questions. Ultimately, what matters is that empirical researchers are aware of the trade offs involved and how these affect their conclusions.

References

Aakvik, A., Heckman, J. J., and Vytlacil, E. J. (2005). Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs. *Journal of Econometrics*, 125(1-2):15–51.

Abadie, A., Angrist, J. D., and Imbens, G. W. (2002). Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica*, 70(1):91–117.

Abbring, J. and Heckman, J. J. (2007). Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B, pages 5145–5303. Elsevier Science, Amsterdam, Netherlands.

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomos Response Data. *Journal of the American Statistical Association*, 88(422):669–679.

Björklund, A. and Moffitt, R. (1987). The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *The Review of Economics and Statistics*, 69(1):42–29.

Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008). The Economics and Psychology of Cognitive and Non-Cognitive Traits. *Journal of Human Resources*, 43(4):815–858.

Browning, M., Heckman, J. J., and Hansen, L. P. (1999). Micro Data and General Equilibrium Models. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1A, pages 543–633. Elsevier Science, Amsterdam, Netherlands.

Bureau of Labor Statistics (2001). *NLS Hanbook 2001: The National Longitudinal Surveys*. U.S. Department of Labor, Washington, DC.

Carneiro, P., Hansen, K., and Heckman, J. J. (2003). Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review*, 44(2):361–422.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–2781.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistican*, 46(3):167–174.

Chamberlain, G. (1986). Asymptotic Efficiency in Semi-Parametric Models with Censoring. *Journal of Econometrics*, 32(2):189–218.

Chernozhukov, V. and Hansen, C. B. (2005). An IV Model of Quantile Treatment Effects. *Econometrica*, 75(1):1243–1284.

Chib, S. (2001). Markov Chain Monte Carlo Methods: Computation and Inference. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 5, pages 3569–3649. Elsevier Science, Amsterdam, Netherlands.

Cochran, W. G. and Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. *Sankhya: The Indian Journal of Statistics Series A*, 34(4):417–446.

Cunha, F. and Heckman, J. J. (2007). The Evolution of Inequality, Heterogeneity and Uncertainty in Labor Earnings in the U.S. Economy. *IZA Discussion Paper*, No. 3115.

Cunha, F., Heckman, J. J., and Navarro, S. (2005). Separating Uncertainty from Heterogeneity in Life Cycle Earnings. *Oxford Economic Papers*, 57(2):191–261.

Cunha, F., Heckman, J. J., and Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931.

Deparment of Defense (1982). *Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery*. Office of the Assistant Secretary of Defense for Manpower and Reserve Affairs, Washington, DC.

Diebolt, J. and Robert, C. P. (1994). Estimation of Finite Mixture Distributions Through Bayesian Sampling. *Journal of the Royal Statistical Society B*, 56(2):363–375.

Escobar, M. D. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, 90(460):577–588.

Ferguson, T. S. (1983). Bayesian Density Estimation by Mixtures of Normal Distributions. In Rizvi, M. H. and Chernoff, H., editors, *Recent Advances in Statistics*, pages 287–302. Academic Press, New York, NY.

Firpo, S. (2007). Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica*, 75(1):703–745.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, pages 1–102. Elsevier Science, Amsterdam, Netherlands.

French, E. and Taber, C. (2011). Identification of Models of the Labor Market. In Ashenfelter, O. and Card, D., editors, *Hanbook of Labor Economics*, volume 4A, pages 537–617. Elsevier Science, Amsterdam, Netherlands.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York, NY.

Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the Estimation of Conditional Probability Densities. *Journal of the American Statistical Association*, 99(468):1015–1026.

Hansen, K., Heckman, J. J., and Mullen, K. (2004). The Effect of Schooling and Ability on Achievement Test Scores. *Journal of Econometrics*, 121(1-2):39–89.

Heckman, J. J. (1990). Varieties of Selection Bias. *American Economic Review*, 80(2):313–318.

Heckman, J. J. (2001). Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy*, 109(4):673–748.

Heckman, J. J. (2008). Econometric Causality. *International Statistical Review*, 76(1):1–27.

Heckman, J. J., Ichimura, H., Smith, J., and Todd, P. E. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66(5):1017–1098.

Heckman, J. J., Lochner, L., and Taber, C. (1999). Human Capital Formation and General Equilibirum Treatment Effects: A Study of Tax and Tuition Policy. *Fiscal Studies*, 20(1):25–40.

Heckman, J. J., Lopes, H., and Piatek, R. (2012). Treatment Effects: A Bayesian Perspective. *Econometric Reviews*, forthcoming.

Heckman, J. J., Schmierer, D., and Urzua, S. (2010). Testing the Correlated Random Coefficient Model. *Journal of Econometrics*, 158(2):177–203.

Heckman, J. J., Smith, J., and Clements, N. (1997). Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies*, 64(4):487–535.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006a). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Heckman, J. J., Urzua, S., and Vytlacil, E. J. (2006b). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Heckman, J. J. and Vytlacil, E. J. (2001a). Local Instrumental Variables. In Hsiao, C., Morimune, K., and Powell, J., editors, *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pages 1–46. Cambridge University Press, Cambridge, United Kingdom.

Heckman, J. J. and Vytlacil, E. J. (2001b). Policy-Relvant Treatment Effects. *American Economic Review*, 91(2):107–111.

Heckman, J. J. and Vytlacil, E. J. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–739.

Heckman, J. J. and Vytlacil, E. J. (2007a). Econometric Evaluation of Social Programs, Part I: Causal Effects, Structural Models and Econometric Policy Evaluation. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B, pages 4779–4874. Elsevier Science, Amsterdam, Netherlands.

Heckman, J. J. and Vytlacil, E. J. (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 Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B, pages 4875–5144. Elsevier Science, Amsterdam, Netherlands.

Herrnstein, R. J. and Murray, C. A. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press, New York, NY.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960.

Horwitz, R. I., Singer, B., Makuch, R. W., and Viscoli, C. M. (1996). Can Treatment that is Helpful on Average be Harmful to Some Patients? A Study of Conflicting Information Needs of Clinical Inquiry and Drug Regulation. *Journal of Clinical Epidemiology*, 49(4):395–400.

Horwitz, R. I., Singer, B., Makuch, R. W., and Viscoli, C. M. (1997). Reaching the Tunnel at the End of the Light. *Journal of Clinical Epidemiology*, 50(7):753–755.

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of Lifetime Inequality. *American Economic Review*, 101(2):2923–2954.

Imbens, G. W. and Wooldrige, J. M. (2009). Recent Developments in the Econometrics of Policy Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Kling, J. R. (2001). Interpreting Instrumental Variables Estimates of the Returns to Schooling. *Journal of Business and Economic Statistics*, 19(3):358–364.

Manski, C. (2012). Identification of Treatment Response with Social Interactions. *Journal of Econometrics*, forthcoming.

Meghir, C. and Pistaferri, L. (2011). Earnings, Consumption and Life Cycle Choices. In Ashenfelter, O. and Card, D., editors, *Hanbook of Labor Economics*, volume 4B, pages 773–854. Elsevier Science, Amsterdam, Netherlands.

Piatek, R. (2010). Bayesian Inference for Factor Structure Models via Gibbs Sampling. *Unpublished Technical Report*.

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying two Separate Regimes. *Journal of the American Statistical Association*, 53(284):873–880.

Quandt, R. E. (1972). A New Approach to Estimating Switching Regressions. *Journal of the American Statistical Association*, 67(338):306–310.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2):135–146.

Scott, D. W. (1992). *Multivariate Density Estimation. Theory, Practice and Visualization.* Wiley, New York, NY.

Silverman, B. W. (1986). *Density Estimation.* Chapman & Hall, London, United Kingdom.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Consumption and Risk Sharing Over the Life Cycle. *Journal of Monetary Economics*, 51(3):609–633.

Thurstone, L. L. (1934). The Vectors of Mind. *Psychological Review*, 41(1):1–32.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S.* Springer, New York, NY.

A Estimation Material

Table 7: Model Fit

Period 1					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.360	0.476	1.975	2.400	2.740
Model	2.376	0.543	1.921	2.377	2.834
Period 2					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.358	0.483	1.975	2.370	2.740
Model	2.371	0.564	1.895	2.375	2.846
Period 3					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.374	0.502	2.006	2.400	2.775
Model	2.392	0.581	1.903	2.391	2.882
Period 4					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.399	0.489	2.035	2.405	2.809
Model	2.385	0.556	1.918	2.383	2.853
Period 5					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.437	0.523	2.026	2.435	2.839
Model	2.438	0.601	1.931	2.439	2.945
Average					
Source	Mean	Sd.	2. Decile	5. Decile	8. Decile
Data	2.385	0.434	2.013	2.388	2.753
Model	2.393	0.336	2.110	2.393	2.676

Notes: based on 100,000 simulated agents and 1,287 actual agents, Sd. = Standard Deviation.

B Supplementary Material

Table 8: MTE Weights

$h_{ATE}(x, u_S)$	$=$	1
$h_{TT}(x, u_S)$	$=$	$\left[\int_{u_S}^{1} f(P \mid X = x)dp\right] \frac{1}{E[P \mid X=x]}$
$h_{TUT}(x, u_S)$	$=$	$\left[\int_{0}^{u_S} f(P \mid X = x)dp\right] \frac{1}{E[(1-P) \mid X=x]}$
$h_{PRTE}(x, u_S)$	$=$	$\left[\frac{F_{P*,X}(u_S) - F_{P,X}(u_S)}{\Delta P}\right]$

Source: Heckman and Vytlacil (2005).

Table 9: Prior Specifications

Potential Outcome, Educational Choice, and Measurement Model		
Group	Parameters	Priors
slope parameters	$\{\beta_{1t}, \beta_{0t}\}, \{\gamma_X, \gamma_Z\}, \{\psi_j\}$	uninformative
factor loadings	$\{\alpha_{1t}, \alpha_{0t}\}, \gamma_\theta, \{\delta_j\}$	uninformative
variances	$\{\sigma^2_{\epsilon_{1t}}, \sigma^2_{\epsilon_{0t}}\}, \{\sigma^2_{\nu_j}\}$	$\mathcal{G}^{-1}(2, 1)$

Normal Finite Mixture Model		
Group	Parameters	Priors
shares	$\{\pi_k\}$	$Dir(10)$
means	$\{\mu_k\}$	$\mathcal{N}(0, 10)$
variances	$\{\sigma_k\}$	$\mathcal{W}^{-1}(2, 1)$

Notes: $\mathcal{N}$ = Normal distribution, Dir = Dirichlet Distriubution, $\mathcal{W}$ = Wishart Distribution.

C NLSY Data

The dataset used in the analysis was originally prepared for Carneiro et al. (2011). By adding the AFQT subscores, the dataset was extended to fit the needs of a factor structure model.

The dataset is based on the National Longitudinal Survey of Youth of 1979 (NLSY79).[12] The NLSY79 is a nationally representative sample for the United States of 12,686 young men and women who were 14 to 22 years of age when first surveyed in 1979. The cohort was interviewed annually through 1994. Since 1994, the survey has been administered biennially. The sample is restricted to white males only. The oversample of poor whites and the military sample are excluded. The raw data contains 2,439 observations before addressing missing values and reporting errors. We present details on the construction of the variables and the resulting descriptive statistics below.

Individual Characteristics The data includes mothers's years of education, number of siblings, dummy variables indicating urban residence at age 14, dummies for year of birth, and labor market experience. Labor market experience is actual work experience in weeks (divided by 52 to express it as a fraction of a year) accumulated from 1979 to 1991 (annual weeks worked are imputed to be zero if they are missing in any given year).

Labor Market Conditions Current (time of outcome measure), past (time of educational choice) as well as permanent labor market conditions are part of the dataset. For the current economic environment, the local average wages in the county of residence in 1991, and the average unemployment rate in the state of residence in 1991 are included. Reflecting past economic circumstances, local average wages in the county of residence at 17 and local unemployment rate in state of residence at age 17 are available. To account for long-run economic conditions, measures of permanent local labor market conditions, i.e. average wages and unemployment between 1973 and 2000 for each location of residence at 17, are included.

Educational Opportunities The presence of a four-year college in the county of residence at age 14, average tuition in public four-year colleges in the county of residence at age 17 (deflated to 1993) are part of the dataset.

[12]See Bureau of Labor Statistics (2001) for a detailed description of the NLSY79.

Educational Choice Individuals are separated into two groups. The first comprises high school dropouts and high school graduates, while the second is made up of individuals with some college education, college graduates and post-graduates. Schooling is measured in 1991 (individuals are between 28 and 34 years of age in 1991). Those with a higher level of educational attainment have on average four more years of education.

Table 10: Covariates

Individual Characteristics	All	Treated	Untreated
Years of Experience	7.963	6.468	9.308
Mother's Years of Schooling	12.042	12.848	11.258
Number of Siblings	2.983	2.637	3.295
AFQT Score	0.393	0.918	-0.095
Urban Residence	0.744	0.787	0.705
Labor Market Conditions	All	Treated	Untreated
Current Local Wages	10.291	10.317	10.267
Current Local Unemployment	6.869	6.873	6.865
Past Local Wages	10.280	10.279	10.282
Past Local Unemployment	7.140	7.144	7.136
Permanent Local Wages	10.286	10.301	10.272
Permanent Local Unemployment	6.272	6.222	6.316
Educational Opportunities	All	Treated	Untreated
Local Presence of Public College	0.521	0.576	0.471
Local Tuition at Public College	19.745	19.360	20.090
Educational Choice	All	Treated	Untreated
Treatment	0.473	1.000	0.000

Notes: based on nonmissing values in the raw data.

Measurements on Cognitive Ability The measurements are taken from the Armed Service Vocational Aptitude Battery (ASVB), which are described in Deparment of Defense (1982). It includes the Armed Forces Qualification Test (AFQT), which consists of the subtests word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge. These subscores are corrected for the fact that different individuals have different amounts of schooling at the time they take the test following the procedure developed in Hansen et al. (2004).

Table 11: Measurements

	All	Treated	Untreated
Arithmetic Reasoning	0.000	0.364	-0.340
Word Knowledge	0.000	0.268	-0.251
Paragraph Composition	0.000	0.285	-0.266
Math Knowledge	0.000	0.466	-0.436

Notes: based on nonmissing values in the raw data; measures standardized to mean zero and standard deviation one in the whole overall sample.

Outcome The wage variable that is included are hourly wages reported in 1989, 1990, 1991, 1992, and 1993. All wage observations that are below 1 or above 100 are deleted.

Table 12: Outcome

Period	All	Treated	Untreated
1	11.884	13.666	10.180
2	11.754	13.726	9.915
3	11.903	13.803	10.111
4	12.603	14.836	10.399
5	13.409	16.110	10.735

Notes: based on nonmissing values in the raw data.

Additional Data Sources and Local Averages Local wages and unemployment rates are averages across all individuals in the population residing in a given area (county for wages, state for unemployment), independent of age, gender, race, and skill level. For each location, permanent local wages and unemployment are based on the average of each variable between 1973 and 2000 are computed by location of residence at 17 (county for wages, state for unemployment). County wages correspond to the average wage per job in the county, constructed using data from the Bureau of Economic Analysis (BES) and deflated to 2000. The state unemployment rate data come from the Bureau of Labor Statistics (BLS) website. However, from the BLS website it is not possible to get state unemployment data for all states in all years. Data are available for all states from 1976 on, and for 29 states for 1973, 1974 and 1975. Therefore, for some of the individuals the unemployment rate in the state of residence in 1976 (which will correspond to age 19 for those born in 1957 and age 18 for those born in 1958) is assigned.

Annual records on tuition, enrollment, and location of all public four-year colleges in the United States were constructed from the Department of Education's annual "Higher Education General Information Survey (HEGIS)" and Integrated Postsecondary Education Data System's "Institutional Characteristics Surveys (IPES)". By matching location with county of residence, the presence of four-year colleges is determined. The distance variable used is the one used in Kling (2001), available at the Journal of Business and Economics Statistics website. Tuition measures are taken as enrollment weighted averages of all public four-year colleges in a person's county of residence (if available) or at the state level if no college is available. County and state of residence at 17 are not available for everyone in the NLSY, but only for the cohorts born in 1962, 1963, and 1964 (age 17 in 1979, 1980, and 1981). However, county and state of residence at age 14 are available for most respondents. Therefore, location at 17 to be equal to location at 14 for cohorts born between 1957 and 1962 is imputed unless location at 14 is missing, in which case location in 1979 is used for the imputation. Many individuals report having obtained a bachelors degree or more and, at the same time, having attended only 15 years of schooling (or less). Years of schooling for these individuals are recoded to be 16.