# Essays in the Econometrics of Policy Evaluation

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Philipp Eisenhauer

November 2012

| | |
|---|---|
| Abteilungssprecher: | Prof. Dr. Martin Peitz |
| Referent: | Prof. Dr. Dr. h.c. mult. Wolfgang Franz |
| Korreferent: | Prof. James J. Heckman, Ph.D. |

Tag der Verteidigung: 10. Dezember 2012

# Contents

# List of Tables

# List of Figures

# Introduction

My research focuses on the evaluation and design of social programs. I aim to inform policy makers about the benefits, costs, and relevant trade offs of their considered policy alternatives. This requires the ex post evaluation and the ex ante design of policies. Iterating between the two allows to accumulate evidence on the effects of policies and their underlying mechanisms. It is this understanding that permits informed policy choices.

Policies target individuals, whose response to the changes in their circumstances decides upon the success or failure of a policy measure. Effective policy making has to take this reaction to changes in the incentives and constraints that individuals face into account. Otherwise, the policy's objective might not be fulfilled or even reversed due to unanticipated and unintended effects. With its focus on the individual and its actions, economics provides a useful framework for the analysis of a variety of social programs. Confronting economic theory with data allows to verify, validate, and quantify its predictions by means of econometric analysis. Again, it is the underlying economics that determines the choice of the appropriate econometric tools that allows for a valid policy assessment. In particular, differences in the amount of information available to the individuals and the policy maker require careful consideration.

By combining economics and econometrics with suitable data sources, I hope to collect robust evidence on the performance of public policies.

My thesis is a first step in this direction.

In Chapter 1, I explore several issues in the economics and econometrics of policy evaluation by an estimation of the returns to college using the National Longitudinal Survey of Youth of 1979. I report the average returns of a college education, but also estimate the whole distribution of returns and document considerable heterogeneity. My results indicate that agents select their schooling level based on gains unobservable by the econometrician. In addition, I clarify the difference between the effect of a treatment and the effect of a policy by showing which margin of agents is affected by two alternative policy changes. Finally, I unify the abundance of average effect parameters using the fact that all can be expressed as weighted averages of the marginal treatment effect.

This essay allowed me to get an overview regarding the open issues in the economics and econometrics of policy evaluation. After this general perspective, I focused on two selected topics in this research area.

In Chapter 2, I review the marginal benefit of treatment parameter and develop the dual cost and surplus parameters. This project is joint work with Prof. James Heckman (University of Chicago) and Prof. Edward Vytlacil (Yale University). We derive and apply a nonparametric identification analysis of benefits, costs, and surpluses of treatment participation. We show how the overall average effect parameters can be expressed as weighted averages of their marginal counterparts. We illustrate the empirical content of our analysis with an application to educational choice. We find that agents select into college based on their idiosyncratic benefits and perceived costs. The variability in subjective benefits drives college attendance more than the variation in costs.

In the first two essays I focused on the presence and consequences of observable and unobservable treatment effect heterogeneity. In both, I presented econometric strategies that deal with the arising difficulties. Next, I investigated the possibility to exploit these heterogeneities for an optimal policy design.

In Chapter 3, which is joint work with Edward Sung (University of Chicago), I explore how alternative assignment mechanisms affect the performance of social programs. We present a framework that links reallocation policies, i.e. policies that fix the amount of resources and only change their allocation, to the econometrics of policy evaluation. We argue that this connection is particularly policy relevant for active labor market programs and illustrate the importance of our contribution with an application to the National Evaluation of Welfare-to-Work Strategies dataset. Our results indicate that the assignment rule has a profound impact on the scale of a program, its overall performance, and the relative effectiveness of alternative treatments. Thus, our results point to the choice of the assignment mechanism as an important component of an optimal policy design.

# 1 Issues in the Econometrics of Policy Evaluation

## 1.1 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).

This chapter profited greatly from comments by 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. I thank George Yates for his advice regarding the computational implementation.

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 chapter, 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 chapter is as follows. Section 1.2 introduces our conceptual framework and establishes the required notation. We discuss potential sources of agent heterogeneity and justify the objects of interest. Section 1.3 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 1.4 concludes.

# 1.2 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.

## 1.2.1 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 $\mathrm{I}[\cdot]$ 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: |
|---|---|
| $Y_1 = \mu_1(X) + U_1$ | $D = \mathrm{I}[\, S > 0\,]$ |
| $Y_0 = \mu_0(X) + U_0$ | $S = \mathrm{E}[Y_1 - Y_0 - C \mid \mathcal{I}\,]$ |
| 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 $(U_1, U_0) \sim \text{N}(0, \Sigma)$, 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. $(Y_1, Y_0)$

[1] See Heckman et al. (1999b) 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.

refer to measures 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.

### 1.2.2 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

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

any parameter of interest (Heckman and Vytlacil, 2005; Heckman et al., 2006b).

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.

### 1.2.3 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 (2001c) 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

$$
\begin{aligned}
Y_B &= D_B Y_1 + (1 - D_B) Y_0 \\
Y_A &= D_A Y_1 + (1 - D_A) Y_0.
\end{aligned}
$$

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

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

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 (2001b, 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

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

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.

[4] See Table D.1 in Appendix D for a selection of the weights.

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

## 1.3 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.

### 1.3.1 Factor Structure Assumption

The factor structure assumption postulates that a low dimensional vector of latent factors $\theta$ 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, $\theta$ 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 $\theta$ and conditioning on it allows to construct the agents' counterfactual state experience. At the same time, $\theta$ 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 $(\theta, \epsilon_C)$ are private information to the agent while $(\epsilon_1, \epsilon_0)$ are not. The latter reflect uncertain fluctuations in future labor market outcomes. The econometrician observes neither $\theta$ 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 $(\epsilon_1, \epsilon_0)$. These are unknown to the agent at the time of the treatment decision but affect potential outcomes. Thus, unexpected realizations of $(\epsilon_1, \epsilon_0)$ might lead agents to regret their treatment choice ex post.

### 1.3.2 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 $\theta$ 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 $\theta$ 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 $\theta$. With this distribution at hand, the joint distribution of $(U_1, U_0)$ can be recovered.[7]

[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).

### 1.3.3 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 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 $\theta$.

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 $\theta$, 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 $\{\sigma^2_{\epsilon_{1t}}, \sigma^2_{\epsilon_{0t}}\}$.

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

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

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 $\theta$. Conditional on the set of control variables $X$ and the factor $\theta$, the estimation simplifies to a normal linear regression model by treatment status.

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. $\gamma_X$ parameterizes the marginal effects of $X_I$ and $\gamma_Z$ of $Z$. In addition, the treatment choice depends on cognitive ability $\theta$ with loading $\gamma_\theta$.

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

The unobservables $(\theta, \epsilon_c)$ are independent from the observables $(X_I, Z)$. $\epsilon_c$ follows a standard normal distribution and is independent of the factor $\theta$. Conditional on $(X_I, Z)$ and $\theta$, 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.

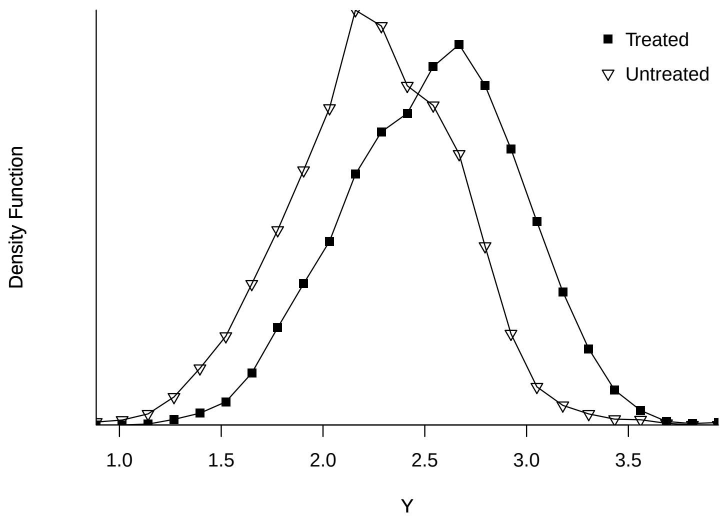
The validity of the exclusion restrictions hinges on the fact that they are not correlated with the unobservables in the wage equations for the adult years. This is questionable for the local labor market conditions at the time of treatment choice: they might be correlated with the long run economic 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 1.1 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.

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 $\theta$ 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

**Figure 1.1:** 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).

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 $\theta$ with $j = 1, ..., 4$ is determined by a set of observable characteristics $W$ and cognitive ability $\theta$. 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 $\theta$. These independence assumptions allow to extract the noise from the signal. Conditional on $W$ the covariation between measurements is due to the common factor $\theta$ only.

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

The idiosyncratic error terms $\{\nu_j\}$ follow a normal distribution. Conditional on $\theta$ and $W$, the estimation of each measurement equation is carried out as a normal linear regression model. To set the scale of $\theta$, 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 1.1:** 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 1.1 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 1.2:** 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 $\theta$ is distributed as a univariate mixture of $K$ normals with share parameter $\pi_k$, mean $\mu_k$, and variance $\sigma_k^2$,

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

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

Table 1.2 summarizes the covariates used in our specification. Additional descriptive statistics and details about the construction of the dataset are provided in Appendix B. Next, we outline our estimation strategy.

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

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

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

$\theta$ 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

[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.

of observable characteristics $(X, X_I, Z)$ with replacement from the original dataset. Third, we simulate the 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:

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

### 1.3.4 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.

#### Model Quality

**Model Fit** Table 1.3 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 1.3:** 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**: Samples 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 1.2 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.

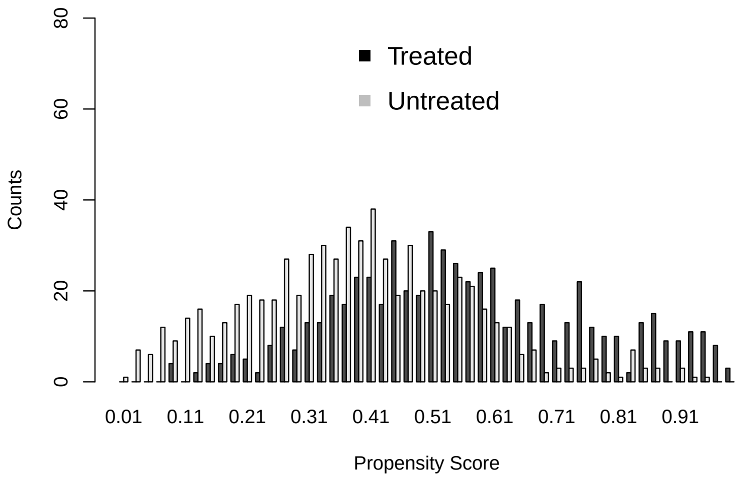
### Selection on Unobservables

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 $\theta$ and uncertainty $(\epsilon_1 - \epsilon_0)$ jointly generate unobservable variability in the returns to college. The agent's private information about $\theta$ induces a

**Figure 1.2:** Common Support




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

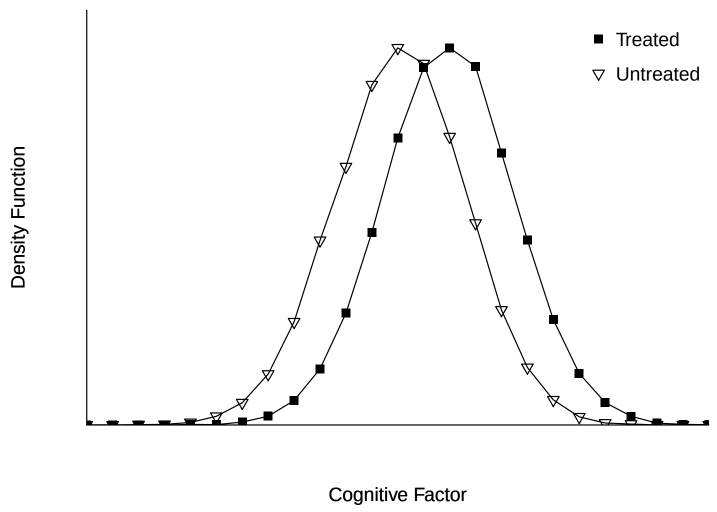
dependency between $V$ and $(U_1 - U_0)$ and creates selection on unobservables.

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 $\theta$. Figure 1.3 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 1.4 plots the marginal effect ($ME$) of cognitive ability on average wages and the probability of a college education along the quantiles $q_\theta$ of the distribution of $\theta$. These are computed based on the simulation as follows

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

**Figure 1.3:** Distribution of Cognitive Skills by Treatment Status

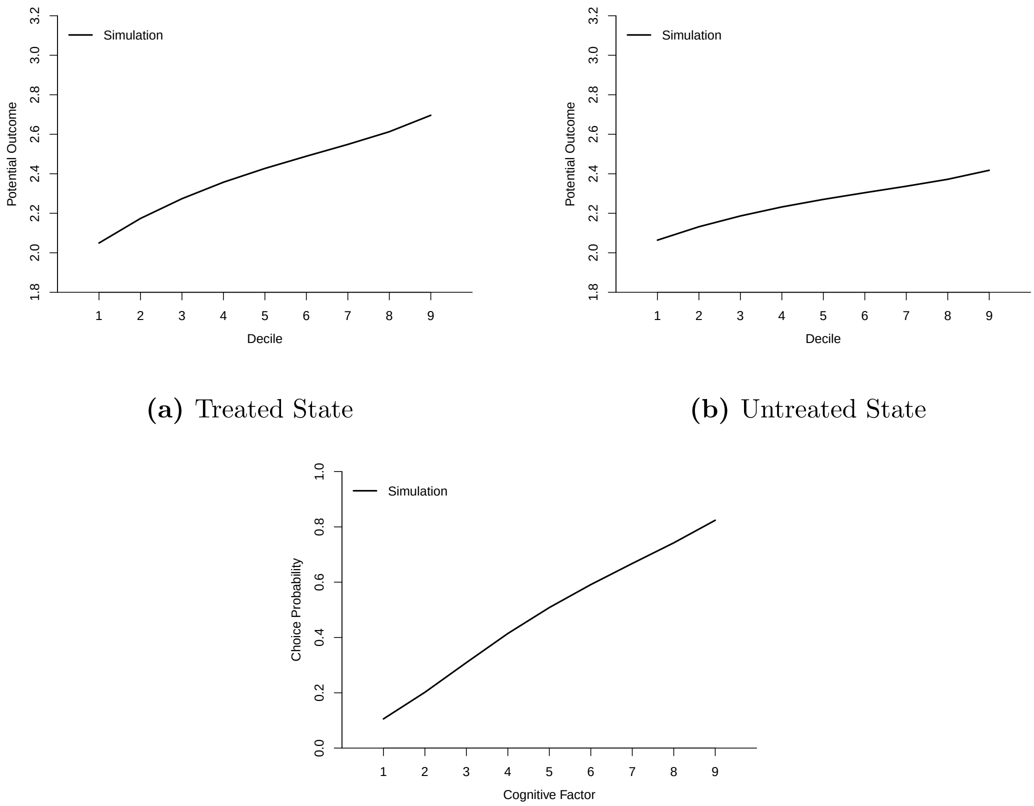



**Notes:** Sample based on 100,000 simulated agents. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992).

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 $\theta$. 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 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.

**Figure 1.4:** Marginal Effects of Ability




**(a)** Treated State

**(b)** Untreated State

**(c)** Choice Model

**Notes:** Sample based on 100,000 simulated agents.

$\theta$ 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.

### 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 1.4 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 1.4:**
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 1.2, 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.
- **Policy Alternative B:** Equalization of tuition fees in all existing public colleges to their mean value.

Table 1.5 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 1.5:** 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

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

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 1.5 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

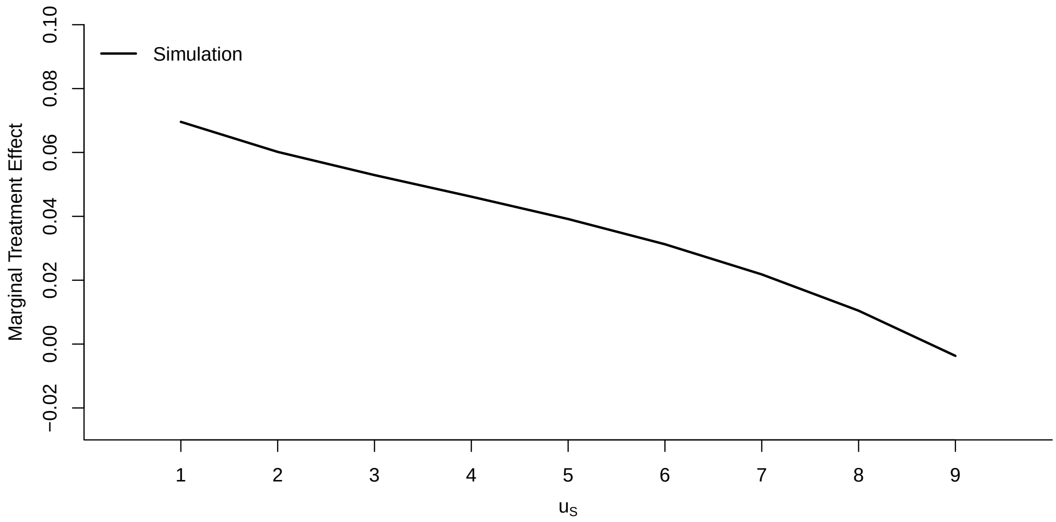
$$MTE(X = \bar{x}, U_S = u_S) = \bar{x}(\beta_1 - \beta_0) - \frac{(\alpha_1 - \alpha_0)}{\gamma_\theta} F_\theta^{-1}(u_s).$$

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

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 $\theta$ results in higher expected returns.

The $MTE$ allows to rationalize the differences between the numerous average effect parameters (Heckman and Vytlacil, 2005, 2001c). All parameters are weighted av-

**Figure 1.5:** Marginal Treatment Effect




**Notes:** Sample based on 100,000 simulated agents.

erages of the $MTE$, but each weighs parts of the distribution of $V$ differently.

Table 1.6 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 1.6:** Comparing the Effects of Treatment

| Conventional | | Policy-Relevant | | |
|---|---|---|---|---|
| Population | Estimand | 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 |

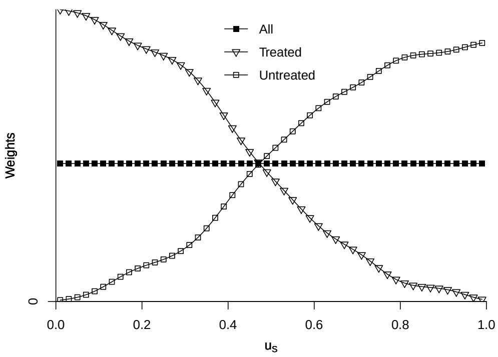
**Notes**: Sample based on 100,000 simulated agents.

Figure 1.6 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.

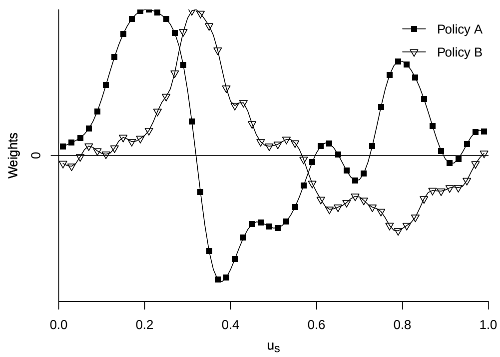
So far, the focus of the discussion has been on average effect parameters. However, these mask considerable treatment effect heterogeneity. That is why we discuss their distributional counterparts next.

[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 $\bar{x}(\beta_1 - \beta_0)$.

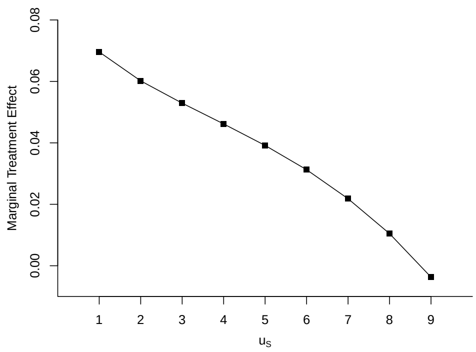
**Figure 1.6:** Weights




**(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. The observable characteristics $X$ are set to their mean values in the sample.

### 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 1.7 presents the results for the joint distribution of potential outcomes $F_{Y_1,Y_0}(y_1, y_0)$ in a contour plot.

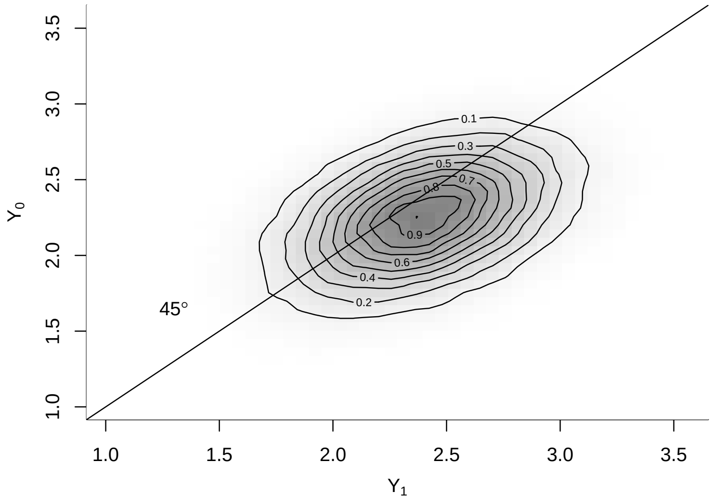
The surface of the plot is directed at the top right corner. Thus, potential outcomes are positively correlated. Agents who do well in one of the education groups also tend to do well in the other. The 45° degree line separates the agents with positive and negative returns. Below the straight line, benefits 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 1.8 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%).

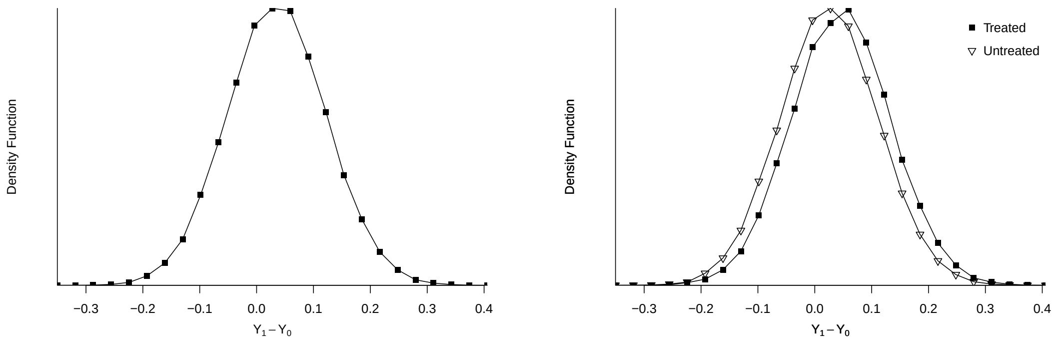
Still, even among the treated, a quite considerable share of agents exhibits negative returns. Among them, there are two groups. First, some agents expected negative returns but have negative subjective cost of education so that they pursue a higher education anyway. Second, some agents expected positive returns ex ante but realized unfavorable draws of $(\epsilon_1, \epsilon_0)$ ex post.

**Figure 1.7:** 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 1.8:** Distribution of Benefits




**(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).

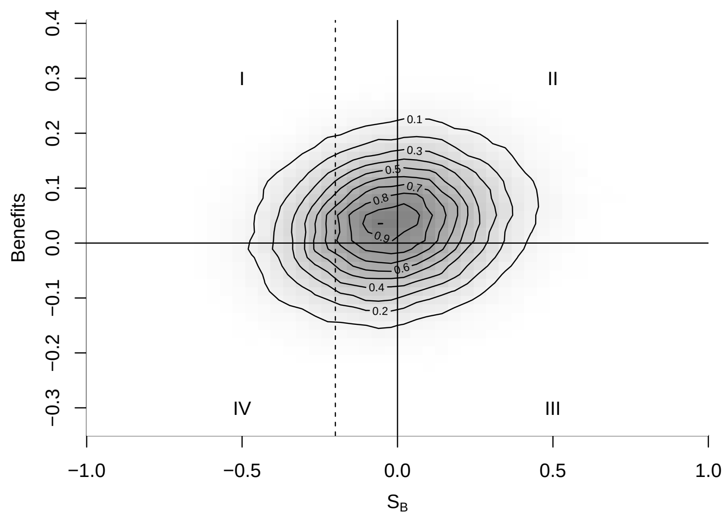
There is a tight link to the conventional average treatment effects reported in Table 1.4. They correspond to the mean values of the respective distribution. All the heterogeneity in returns remains unnoticed by a focus on average effects only. But this heterogeneity requires to be precise about the 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 1.9 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 1.9:** Joint Distribution of Benefits and Surplus




**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 $S_B < 0$.

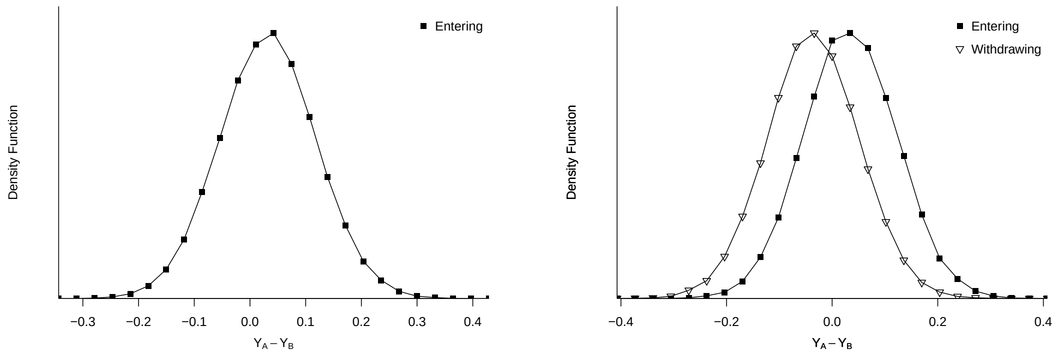
However, Figure 1.9 is most informative on how different policies affect different margins. In the baseline state, agents with $S_B > 0$ select into treatment while those with $S_B < 0$ do not. An alternative policy regime affects the surplus calculation. Let us consider Policy $A$ as an example. Under 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 1.10 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 1.8) 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 1.5. 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 1.10:** 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).

## 1.4 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.

# 2 Cost-Benefit Analysis of Social Programs

## 2.1 Introduction

The classical approach to the evaluation of public policy compares the benefits and costs of policies and forms net measures of surplus to determine whether policies should be undertaken (see Tinbergen (1956), Harberger and Jenkins (2002), and Chetty (2009)). The recent literature on program evaluation or "treatment effects" focuses on gross benefits of policies at unidentified margins of choice and does not consider the marginal costs associated with the programs being evaluated.[1]

In a fundamental paper, Björklund and Moffitt (1987) estimate marginal gains and surpluses for policies within a parametric normal generalized Roy model. They use structural econometric methods to identify the components of the cost and benefit functions. This chapter extends their analysis to a more general setting. It develops and applies a nonparametric identification analysis of benefits, costs, and surpluses without the need to identify all of the ingredients of a fully specified structural model. This approach implements Marschak's Maxim (Heckman, 2010) by directly estimating the cost, benefit, and surplus parameters rather than estimating the full structural model from which the parameters can be constructed. We present ex ante and ex post analyses of benefits and consider alternative specifications of agent information sets. Applying our methods to the data on gross benefits analyzed by Carneiro et al. (2011), we find that within a semiparametric model variability in

[1]See the discussion in Heckman and Vytlacil (2007b) and Heckman (2010).

subjective benefits drives much of the observed variability in the net returns to college and thus the decision to attend college.

In the generalized Roy model, the agent chooses treatment if the perceived benefit exceeds the perceived subjective cost. This creates a simple relationship between the cost and benefit parameters that we exploit for identifying the cost and surplus parameters. Our main analytic result is that cost and surplus parameters in the generalized Roy model can be identified without direct information on the costs of treatment. This is valuable for the analysis of the choice of education, where subjective costs have been estimated to be substantially greater than tuition costs (Cunha et al., 2005). Our analysis complements and extends the analysis of Björklund and Moffitt (1987) who first noted the duality between cost and benefit parameters in the generalized Roy model.

Our analysis is reminiscent of the Heckman (1974) model of female labor supply. In that analysis, the econometrician observes the offered wage only for the agents who choose to work, and the economist never observes the reservation wage of any agent. Yet, the analysis identifies the parameters of both the offered wage equation and the reservation wage equation by using the information that the agents' decision to work is governed by the relationship that the offered wage exceeds the reservation wage.[2] In our analysis, we only observe program outcomes for the agents who select into treatment, and we observe the no treatment outcome only for the agents who do not select into treatment. We do not observe the cost of treatment for any agent. Yet, using the economics of the model, we are able to identify the average benefit and average cost of treatment parameters by exploiting the agent's decision rule of selecting into treatment if the benefit exceeds the cost.

Our analysis is very different from an analysis using a randomized experiment to infer treatment effects. In many common applications of randomization, it is not possible to identify the choice probability (Heckman, 1992; Heckman and Smith, 1995). Instead of using randomization to bypass problems of self-selection, we use the information that agents self-select into treatment to infer information on the cost

[2]The same methodology applies to search theory (Flinn and Heckman, 1982).

of the treatment that could not have been recovered by a randomized experiment.

Our identification analysis applies classical exclusion restrictions that shift costs and benefits from treatment. We use cost shifters to identify the benefit of the treatment, and benefit shifters to identify the cost of the treatment.

This chapter unfolds in the following way. Section 2.2 introduces the generalized Roy model. Section 2.2.2 reviews the average benefit of treatment parameters from Heckman and Vytlacil (1999, 2005, 2007b), and develops and analyzes the dual cost parameters that match the benefit parameters. Section 2.3 presents an identification analysis of the cost and surplus parameters, based on first identifying the marginal benefit of treatment parameter using Local Instrumental Variables (LIV) and then exploiting the duality between cost and benefit parameters in the generalized Roy model to identify marginal cost and surplus parameters. Section 2.3.2 extends our analysis to allow agents to have imperfect foresight of future outcomes. In Section 2.4 we apply the analysis to the study of college-going to infer cost and benefit parameters. Section 2.5 summarizes.

## 2.2 Conceptual Framework

### *2.2.1 Prototypical Model*

Suppose there are two potential outcomes $(Y_0, Y_1)$, and a choice indicator $D$ with $D = 1$ if the agent selects into treatment so that $Y_1$ is observed and $D = 0$ if the agent does not select into treatment so that $Y_0$ is observed. Anticipating our empirical analysis, $Y_1$ is the annualized flow of income from college, and $Y_0$ is the annualized flow of income from high school. The observed outcome $Y$ can be written in switching regression form (Quandt, 1958, 1972)

(2.2.1) $Y = DY_1 + (1 - D)Y_0,$

where $E(Y_j \mid X) = \mu_j$ and

(2.2.2) $Y_j = \mu_j(X) + U_j$

for $j = 0, 1$. $X$ is a vector of regressors observed by the observing economist while $(U_0, U_1)$ are not. Combining equations (2.2.1) and (2.2.2),

$$Y = \mu_0(X) + \{[\mu_1(X) - \mu_0(X)] + U_1 - U_0\}D + U_0.$$

The individual gross benefit of treatment associated with moving an otherwise identical person from zero to one is $B = Y_1 - Y_0$ and is defined as the causal effect on $Y$ of a ceteris paribus move from zero to one. Defining $E(C \mid Z) = \mu_C(Z)$, the subjective cost of choosing treatment as perceived by the agent is

$$\text{(2.2.3)} \quad C = \mu_C(Z) + U_C,$$

where $Z$ is an observed random vector of cost shifters and $U_C$ is an unobserved random variable. Individuals choose treatment if the benefit from treatment is greater than the subjective cost:

$$\text{(2.2.4)} \quad D = 1 \quad \text{if} \quad S \geq 0\,; \qquad D = 0 \quad \text{otherwise},$$

where $S$ is the surplus, i.e. the net gain, from treatment:

$$\begin{aligned} S &= (Y_1 - Y_0) - C \\ &= \{[\mu_1(X) - \mu_0(X)] - \mu_C(Z)\} - [U_C - (U_1 - U_0)] \\ &= \mu_S(X, Z) - V \end{aligned}$$

with $\mu_S(X, Z) = [\mu_1(X) - \mu_0(X)] - \mu_C(Z)$ and $V = U_C - (U_1 - U_0)$. We do not assume any particular functional form for the functions $\mu_0, \mu_1$ and $\mu_C$, and we do not assume that the distribution of $U_0$, $U_1$, or $U_C$ is a known parametric form. We maintain equations (2.2.1) – (2.2.4) throughout this chapter.

The original Roy (1951) model assumes that there are no observed $X$ regressors, that the cost of treatment is identically zero (i.e. $\mu_C = 0, U_C = 0$), and that $(U_0, U_1) \sim N(0, \Sigma)$. Heckman and Honore (1990) develop a nonparametric version of the Roy model using variation in regressors and making no parametric assumption on the distribution of $(U_0, U_1)$. Their version of the Roy model also imposes that the cost of treatment is identically zero. In contrast, we allow non-zero cost of treatment. For our identification analysis we require nondegenerate cost of treatment and observed cost-shifters.

From the point of view of the economist $(X, Z)$ is observed and $(U_1, U_0, U_C)$ is unobserved. This model supposes that agents know the true gross benefit, $B = Y_1 - Y_0$, of the treatment. We extend our results to a broader class of models in which the agents participate in the program if the expected benefits given the information available to them is greater than their cost of treatment in Section 2.3.2. This model also supposes that there is no other aspect of the benefit of the treatment other than $Y_1 - Y_0$. Implicitly, any subjective benefits of the program are incorporated into the costs of treatment, i.e. the cost function includes the subjective benefits of the treatment. For example, if job training allows the individual to work in a job with preferred amenities, this is modeled as a (negative) contribution to the subjective cost of treatment. To simplify the exposition, we suppose that $Z$ and $X$ do not contain any common elements. At greater notational cost, all of the analysis of this chapter can be seen as implicitly conditioning on all common elements of $X$ and $Z$.

We make the following technical assumptions:

(A-1) *$(U_0, U_1, U_C)$ is independent of $(X, Z)$.*

(A-2) *The distribution of $\mu_C(Z)$ conditional on $X$ is absolutely continuous with respect to Lebesgue measure.*

(A-3) *The distribution of $V = U_C - (U_1 - U_0)$ is absolutely continuous with respect to Lebesgue measure and has a cumulative distribution function that is strictly increasing.*

(A-4) *The means of $E|Y_1|$, $E|Y_0|$ and $E|C|$ are finite.*

(A-1) assumes that $(U_0, U_1, U_C)$ is independent of $(X, Z)$. Thus, $D$ is endogenous but other regressors in both the treatment equation and the outcome equation are exogenous. We implicitly condition on any regressors that enter both the outcome equations and the cost equation. Thus, this condition should be interpreted as an independence assumption of the error terms from the unique elements of $X$ and $Z$ conditional on the regressors that enter both equations. (A-2) requires that there exists at least one continuous component of $Z$ conditional on $X$. This assumption will only be required for our identification analysis, and is not needed for our definitions or analysis of the cost and surplus parameters. (A-3) is a regularity condition. It allows for the possibility that $U_C$ is degenerate (costs do not vary conditional on $Z$) or that $U_1 - U_0$ is degenerate (treatment effects do not vary conditional on $X$), though not both. Assumption (A-4) is needed to satisfy standard integration conditions. It guarantees that the mean benefit and cost parameters are well defined. An implication of our model with Assumptions (A-1) and (A-3) is that $0 < \Pr(D = 1 \mid X, Z) < 1$ with probability one, so that there is a control group for almost all $(X, Z)$. Note that this restriction still allows the support of $\Pr(D = 1|X, Z)$ to be the full unit interval.

Let $P(X, Z)$ denote the probability of selecting into treatment given $(X, Z)$, or the "propensity score" $P(X, Z) \equiv \Pr(D = 1 \mid X, Z) = F_V(\mu_S(X, Z))$, where $F_V(\cdot)$ denotes the distribution of $V$.[3] We sometimes denote $P(X, Z)$ by $P$, suppressing the $(X, Z)$ argument. We also work with $U_S$, a uniform random variable ($U_S \sim \text{Unif}[0, 1]$) defined by $U_S = F_V(V)$. Thus different values of $U_S$ denote different quantiles of $V$. Given our previous assumptions, $F_V$ is strictly increasing, and $P(X, Z)$ is a continuous random variable conditional on $X$.

The generalized Roy model presented in this chapter is a special case of the model of Heckman and Vytlacil (1999, 2005). The model of equations (2.2.1)–(2.2.4) under assumptions (A-1)–(A-4) imply the model and assumptions of Heckman and Vytlacil (1999, 2005). From analysis of Vytlacil (2002), the more general model is equivalent to the conditions that justify the Local Average Treatment Effect (LATE) model of Imbens and Angrist (1994). We impose more restrictions here. In particular, we impose the generalized Roy model and the corresponding assumptions that will allow us to exploit the generalized Roy model for identification of subjective cost parameters. As in the conventional Roy model (Heckman and Sedlacek, 1985), we assume additive separability in the outcome equations (2.2.2). This additive separability is not required in Heckman and Vytlacil (1999, 2005), but is required by our analysis to make additive separability in the latent index equation (2.2.4)

[3] We will refer to the cumulative distribution function of a random vector $A$ by $F_A(\cdot)$ and to the cumulative distribution function of a random vector $A$ conditional on random vector $B$ by $F_{A|B}(\cdot)$. We will write the cumulative distribution function of $A$ conditional on $B = b$ by $F_{A|B}(\cdot \mid b)$.

consistent with the generalized Roy model.[4] We also assume conditions on $X$ that are not required by Heckman and Vytlacil (1999, 2005) to identify the gross benefit parameters. In their analysis, they fully condition on $X$, and thus do not need to assume that $X$ is independent of the error vector. In contrast, in order to use the generalized Roy model to recover subjective cost parameters, we require that the unique elements $X$ are independent of the error vector.[5] We are implicitly fully conditioning on any common elements of $X$ and $Z$, and no independence condition is required for the common elements.

### 2.2.2 Objects of Interest

This section defines and analyzes the benefit, cost, and surplus parameters. We maintain the model of equations (2.2.1)–(2.2.4), and invoke assumptions (A-1) and (A-3)–(A-4). We do not require assumption (A-2) for the definition or analysis of the parameters, but use this assumption in the next section in our identification analysis.

Standard treatment effect analyses identify averaged parameters of the gross benefit of treatment, $B = Y_1 - Y_0$. The most commonly invoked treatment effect parameter is the average benefit of treatment $B^{ATE}(x) \equiv E(Y_1 - Y_0|X = x) = \mu_1(x) - \mu_0(x)$. This is the effect of assigning treatment randomly to everyone of type $X = x$ assuming full compliance, and ignoring any general equilibrium effects. Another commonly invoked parameter is the average benefit of treatment on persons who actually take the treatment, referred to as the benefit of treatment on the treated: $B^{TT}(x) \equiv E(Y_1 - Y_0 \mid X = x, D = 1) = \mu_1(x) - \mu_0(x) + E(U_1 - U_0|X = x, D = 1)$. Heckman and Vytlacil (1999, 2005) unify a broad class of treatment effect parameters including the $B^{ATE}(x)$ and $B^{TT}(x)$ through the marginal benefit of treatment, defined as $B^{MTE}(x, u_S) \equiv E(Y_1 - Y_0|X = x, U_S = u_S) = \mu_1(x) - \mu_0(x) + E(U_1 - U_0|U_S = u_S)$. $B^{MTE}$ is the treatment effect parameter that conditions on the unobserved desire to select into treatment.

[4] Recall again that we are implicitly conditioning on all common elements of $(X, Z)$, so that these need not be additively separable from the error term.

[5] In this respect, our analysis is broadly analogous to the identification strategies and conditions of Yildiz and Vytlacil (2007) and Shaikh and Vytlacil (2011), who also require that there be exogenous regressors in the outcome equation and exploit variation in such regressors for identification.

The conventional treatment analysis does not define, identify, or estimate any aspect of the cost of the treatment. We define a set of cost parameters parallel to the benefit of treatment parameters, where cost is the subjective cost as perceived by the agent. Thus, we define the average cost of treatment, the average cost of treatment on those treated, and the marginal cost of treatment as follows:

$$\begin{aligned} C^{ATE}(z) &= E(C|Z=z) &&= \mu_C(z) \\ C^{TT}(z) &= E(C|Z=z, D=1) &&= \mu_C(z) + E(U_C|Z=z, D=1) \\ C^{MTE}(z, u_S) &= E(C|Z=z, U_S=u_S) &&= \mu_C(z) + E(U_C|U_S=u_S). \end{aligned}$$

Recalling that $S = B - C = \mu_S(X,Z) - V$, where $\mu_S(X,Z) = [\mu_1(X) - \mu_0(X)] - \mu_C(Z)$ and $V = U_C - (U_1 - U_0)$, we now define the corresponding surplus parameters:

$$\begin{aligned} S^{ATE}(x,z) &= E(S|X=x, Z=z) &&= \mu_S(x,z) \\ S^{TT}(x,z) &= E(S|X=x, Z=z, D=1) && \\ &= \mu_S(x,z) - E(V|X=x, Z=z, D=1) && \\ {}^{MTE}(x,z,u_S) &= E(S|X=x, Z=z, U_S=u_S) &&= \mu_S(x,z) - E(V|U_S=u_S) \end{aligned}$$

With these parameters, we can now ask questions not only about the outcome change from the treatment but also the subjective cost of the treatment and the net surplus.

Following Heckman and Vytlacil (1999, 2005), we can represent the average treatment effects and treatment on the treated as averaged versions of the marginal effects of treatment:

(2.2.5)

$$\begin{aligned} B^{ATE}(x) &= \int_0^1 B^{MTE}(x, u_S) du_S \\ B^{TT}(x) &= \int_0^1 B^{MTE}(x, u_S) \frac{1 - F_{P|X}(u_S|x)}{\int_0^1 (1 - F_{P|X}(t|x)) dt} du_S. \end{aligned}$$

Following the same line of argument as used by Heckman and Vytlacil (1999, 2005),

(2.2.6)

$$\begin{aligned} C^{ATE}(z) &= \int_0^1 C^{MTE}(z, u_S) du_S \\ C^{TT}(z) &= \int_0^1 C^{MTE}(z, u_S) \frac{1 - F_{P|Z}(u_S|z)}{\int_0^1 (1 - F_{P|Z}(t|z)) dt} du_S, \end{aligned}$$

and

(2.2.7)

$$\begin{array}{lcl} S^{ATE}(x,z) & = & \int_0^1 S^{MTE}(x,z,u_S)du_S \\ S^{TT}(x,z) & = & \frac{1}{P(x,z)}\int_0^{P(x,z)} S^{MTE}(x,z,u_S)du_S. \end{array}$$

We now establish some relationships among these marginal effects. First, consider the marginal surplus parameter. Defining $U_S = F_V(V)$ with $F_V$ strictly increasing, we have that $U_S = u_S$ is equivalent to $V = F_V^{-1}(u_S)$, and thus

$$S^{MTE}(x,z,u_S) = \mu_S(x,z) - E\left(V|V = F_V^{-1}(u_S)\right) = \mu_S(x,z) - F_V^{-1}(u_S).$$

$F_V^{-1}$ is strictly increasing, and thus $S^{MTE}(x,z,u_S)$ is strictly decreasing in $u_S$. Individuals with low $u_S$ want to enter the program the most and are those with the highest surplus from the program, while individuals with high $u_S$ want to enter the program the least and have the smallest surplus from the program. Again using the fact that $F_V$ is strictly increasing and that $P(X,Z) = F_V(\mu_S(X,Z))$, conditioning on $U_S = P(x,z)$ is equivalent to conditioning on $V = \mu_S(x,z)$, and thus

$$S^{MTE}(x,z,P(x,z)) = \mu_S(x,z) - E\left(V|V = \mu_S(x,z)\right) = 0.$$

An individual with $U_S = P(x,z)$ is an individual who is indifferent towards treatment if assigned $X = x, Z = z$. Since $S^{MTE}(x,z,u_S)$ is strictly decreasing in $u_S$, we have $S^{MTE}(x,z,u_S)$ is positive for $u_S < P(x,z)$, $= 0$ at $u_S = P(x,z)$, and is negative for $u_S > P(x,z)$. If we instead consider holding the $u_S$ evaluation point fixed and consider how $S^{MTE}(x,z,u_S)$ varies with $(x,z)$, we have that $S^{MTE}(x,z,u_S)$ will be positive for all $(x,z)$ such that $P(x,z) > u_S$ and will be negative for all $(x,z)$ such that $P(x,z) < u_S$.

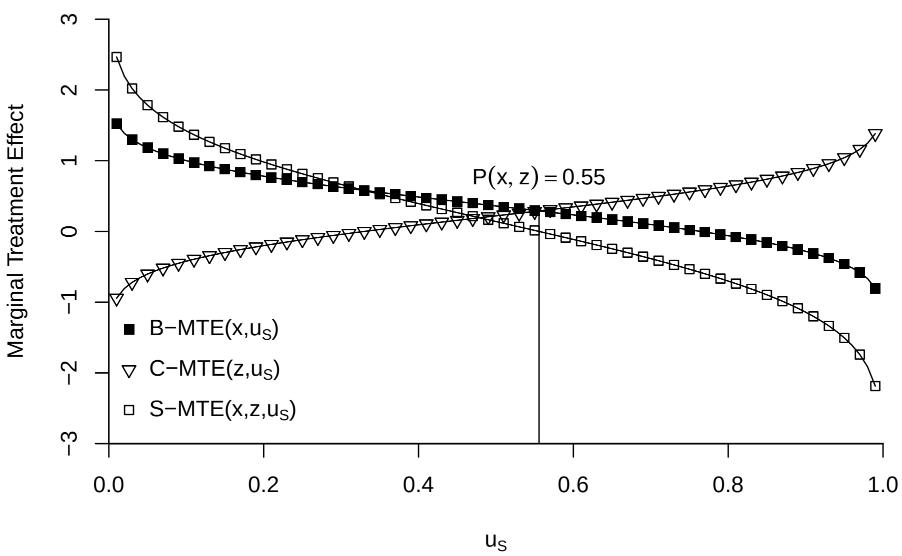
We have thus far discussed only the marginal surplus function. Using the relationship $S^{MTE}(x,z,u_S) = B^{MTE}(x,u) - C^{MTE}(z,u_S)$, we can translate these statements into relative statements about the marginal benefit and marginal cost functions:

$$\begin{array}{lclcl} B^{MTE}(x,u_S) & > & C^{MTE}(z,u_S) & \forall & (x,z,u_S) \text{ s.t. } P(x,z) < u_S \\ B^{MTE}(x,u_S) & = & C^{MTE}(z,u_S) & \forall & (x,z,u_S) \text{ s.t. } P(x,z) = u_S \\ B^{MTE}(x,u_S) & < & C^{MTE}(z,u_S) & \forall & (x,z,u_S) \text{ s.t. } P(x,z) > u_S. \end{array}$$

The benefit and cost parameters coincide when evaluated at $u_S = P(x,z)$, because in this case marginal cost equals the marginal benefit. The equality between marginal benefit and marginal cost for people at the margin is used in the next session to secure identification of cost parameters.

To fix ideas, we show the full set of marginal effects of treatment for a numerical example in Figure 2.1. Evaluated at fixed values of the observables $(X,Z)$, agents select their treatment status based on gains that are not observed by the econometrican. This numerical example plots the marginal effect curves for fixed $(x,z)$ such that $P(x,z) = 0.55$. Individuals with $(X,Z) = (x,z)$ and $u_S = P(x,z)$ have the benefits of treatment just offset by the subjective cost, and so are just indifferent as to whether they receive treatment or not. Thus, at $u_S = P(x,z) = 0.55$, the marginal benefit and marginal cost curves intersect, and the marginal surplus of treatment equals zero. For $u_S < P(x,z)$, the marginal benefit curve lies above the marginal cost curve and thus the marginal surplus curve is strictly positive, while the opposite is true for $u_S > P(x,z)$. In this example, at $u_S = P(x,z)$, the benefit of treatment is still positive, but so is the cost of treatment.

**Figure 2.1:** Marginal Effects of Treatment



This example is constructed to have intuitive properties, with the marginal benefit of treatment $B^{MTE}(x, u_S)$ decreasing in $u_S$ and the marginal cost of treatment $C^{MTE}(z, u_S)$ increasing in $u_S$. Agents with the greatest unobserved desire to select into treatment not only have higher benefits, but also have lower costs. These properties, while intuitive, need not hold in general – individuals with lower values of $u_S$ (and thus greater unobserved desire to take treatment) must necessarily have higher net surplus than those with higher values of $U_S$, but they need not have higher benefits and lower costs. It is possible, for example, that benefits and costs are positively correlated. In that case, those with the greatest unobserved desire to participate have the least benefit but also the least costs with costs so low as to offset their low benefit. Alternatively, they have the highest costs but also the highest benefits with benefits so high as to offset their high costs. However, we show below that the intuitive properties that $B^{MTE}(x, u_S)$ is decreasing in $u_S$ and $C^{MTE}(z, u_S)$ is increasing in $u_S$ will hold under additional conditions.

**Remark 1.** Consider some special cases. If benefits do not vary across individuals conditional on $X$, i.e. if $U_1 - U_0$ is degenerate, than $B^{MTE}(x, u_S) = B^{ATE}(x) = B^{TT}(x)$. In addition, if $U_1 - U_0$ is degenerate, this implies that $V = U_C$ and $U_S = F_{U_C}(U_C)$ so that

$$\begin{aligned} C^{MTE}(z, u_S) &= \mu_C(z) + E(U_C | U_S = u_S) \\ &= \mu_C(z) + E(U_C | U_C = F_{U_C}^{-1}(u_S)) = \mu_C(z) + F_{U_C}^{-1}(u_S) \end{aligned}$$

which is increasing in $u_S$. In this case, variation in unobserved costs drives selection conditional on $(X, Z)$ and those who most want to enter the program (have the lowest $U_S$) have the least cost of treatment. Using equation (2.2.6) and the fact that $C^{MTE}(z, u_S)$ is increasing in $u_S$, it follows that $C^{TT}(z) < C^{ATE}(z)$ so that, conditional on $Z$, those who chose treatment have a lower cost of treatment than those who did not select into treatment. Symmetrically, if costs do not vary across individuals conditional on $Z$, i.e. if $U_C$ is degenerate, then heterogeneity in the benefits of treatment drive selection and (1) $C^{MTE}(x, u_S) = C^{ATE}(x) = C^{TT}(x)$; (2) $B^{MTE}(x, u_S)$ is decreasing in $u_S$; and (3) $B^{TT}(x) > B^{ATE}(x)$.

The marginal surplus parameter is highest for those who most want to participate in the program. Using equation (2.2.7) we thus have that the average surplus among the treated is higher than the unconditional average surplus of treatment. As discussed in Remark 1, degeneracy of either $U_1 - U_0$ or of $U_C$ implies that treatment parameters and cost parameters will have intuitive properties, such as highest benefit or lowest cost for those who most want to participate in the program. We now show a more general set of conditions under which these properties of the treatment effect parameters will hold.

**Theorem 1.** *Assume that equations (2.2.1)–(2.2.4) and our assumptions (A-1)–(A-4) hold.*

1. $S^{TT}(x) > S^{ATE}(x)$, *and* $S^{MTE}(x, u_S)$ *is monotonically decreasing in* $u_S$.
2. *Suppose that* $U_C \perp\!\!\!\perp U_1 - U_0$. *Then* $C^{TT}(z) \leq C^{ATE}(z)$, $B^{TT}(x) \geq B^{ATE}(x)$.
3. *Suppose that* $U_C \perp\!\!\!\perp U_1 - U_0$, *and that* $U_C$ *and* $U_1 - U_0$ *have log concave densities. Then* $C^{MTE}(z, u_S)$ *is monotonically increasing in* $u_S$ *and* $B^{MTE}(x, u_S)$ *is monotonically decreasing in* $u_S$. $\square$

For proof, see Appendix A

The theorem provides intuitive results. If the unobservables related to the cost and benefit are independent, then the average benefit among those who select into treatment is larger than the unconditional average benefit. At the same time, the average cost among those who select into treatment is lower than the unconditional average cost. In other words, under independence of the unobservables related to benefits and costs, it is the high benefit and low cost individuals who select into treatment in the generalized Roy model. Parts (2) and (3) of the theorem state that, under a regularity condition, the expected gain is decreasing while the expected cost is increasing in $U_S$. Note that the normal density as well as many other standard densities are log concave.[6]

The numerical example previously introduced in Figure 2.1 is based on unobserved variables drawn from a normal distribution with unobservable benefits $(U_1 - U_0)$

[6] Heckman and Honore (1990) exploit the restriction of log-concave density functions for the disturbance terms in a Roy model with zero costs. See Bagnoli and Bergstrom (2005) for a review of log concave densities and economic applications.

independent of the observable component of cost $U_C$. The marginal effects of treatment exhibit the shape predicted by Theorem 1, part (3). The $B^{MTE}(x, u_S)$ is decreasing in $u_S$, while the opposite is true for $C^{MTE}(z, u_S)$.

## 2.3 Identification Analysis

### 2.3.1 General Case

Heckman and Vytlacil (1999, 2005) establish that local instrumental variables (LIV) identify the marginal benefit of treatment:

$$(2.3.1)\quad \frac{\partial}{\partial p} E(Y|X = x, P = p) = B^{MTE}(x, p).$$

We can identify $E(Y|X = x, P = p)$ and its derivative for all $(x, p) \in \text{Supp}(X, P)$, where $\text{Supp}(X, P)$ denotes the support of $(X, P(X, Z))$.[7] We thus have identification of $B^{MTE}(x, u_S)$ for all values of $(x, u_S) \in \text{Supp}(X, P)$. For a fixed $x$, we can identify $B^{MTE}(x, u_S)$ for $u_S \in \text{Supp}(P|X = x)$. The more variation in propensity scores $p$ conditional on $X = x$, the larger the set of evaluation points $u_S$ for which we identify $B^{MTE}(x, u_S)$. Variation in propensity scores conditional on $X$ is driven by variation in $Z$, the cost shifters. Thus, if we observe regressors that produce large variations in costs, we will be able to identify $B^{MTE}(x, u_S)$ on a larger set.

We can identify $B^{ATE}(x)$ and $B^{TT}(x)$ by identifying $B^{MTE}(x, u_S)$ over the appropriate support and then integrating the latter with the appropriate weights. By equation (2.2.5), we identify $B^{ATE}(x)$ if $\text{Supp}(P|X = x) = [0, 1]$. For fixed $X = x$, this requires that there be enough variation in the cost shifters $Z$ to drive the probabilities $P(x, Z)$ all the way to zero and to one. In other words, holding fixed the regressors that enter the outcome equation, we must observe costs shifters such that conditional on some values of those cost shifters, the cost to the agent is so low that the agent will select into treatment with probability one, and conditional on other values of the cost shifters, the cost to the agent is so high that the agent

[7] For any random vectors $A$ and $B$, we will write the support of the distribution of $A$ as $\text{Supp}(A)$, and the support of distribution of $A$ conditional on $B = b$ as $\text{Supp}(A|B = b)$.

will select into treatment with probability zero. Likewise, we identify $B^{TT}(x)$ if $\text{Supp}(P|X = x) = [0, p_x^{max}]$ where $p_x^{max}$ is the supremum of $\text{Supp}(P|X = x)$. This support requirement in turn requires that, for fixed $X = x$, there be enough variation in the cost shifters $Z$ to drive the selection probability to zero.[8]

Using equation (2.3.1) and the relationship for people on the margin of choice that $B^{MTE}(x, P(x, z)) = C^{MTE}(z, P(x, z))$, we have

$$\frac{\partial}{\partial p} E(Y|X = x, P = p)\Big|_{p=P(x,z)} = C^{MTE}(z, P(x, z)).$$

Using this relationship, we identify $C^{MTE}(z, u_S)$ for all values of $(z, u_S) \in \text{Supp}(Z, P)$. We thus identify the marginal cost parameter without direct information on the cost of treatment by using the structure of the Roy model and by identifying the marginal benefit of treatment for individuals at the margin of participation. For a fixed $z$, we identify $C^{MTE}(z, u_S)$ for $u_S \in \text{Supp}(P|Z = z)$. The more variation in propensity scores conditional on $Z = z$, the larger the set of evaluation points for which we identify $C^{MTE}(z, u_S)$. Variation in propensity scores conditional on $Z = z$ is driven by variation in $X$, the regressors that drive the outcome. Thus, if we observe regressors that cause large variations in benefits, we will be able to identify $C^{MTE}(z, u_S)$ at a larger set of $u_S$ evaluation points. In contrast, if there are no $X$ regressors, then $P$ only depends on $Z$ and we can only identify $C^{MTE}(z, u_S)$ for $u_S = P(z)$.

From equation (2.2.6), we can identify $C^{ATE}(x)$ if $\text{Supp}(P|Z = z) = [0, 1]$. This requires, for fixed $Z = z$, for there to be enough variation in the outcome shifters $X$ to drive the probabilities $P(X, Z)$ all the way to zero and to one. In other words, holding fixed the regressors that enter the cost equation, we must observe outcome shifters such that conditional on some values of those outcome shifters, the benefit to the agent is so high that the agent will select into treatment with probability one; conditional on other values of the outcome shifters, the benefit to the agent is so low that the agent will select into treatment with probability zero. Likewise, we identify $C^{TT}(x)$ if $\text{Supp}(P|Z = z) = [0, p_z^{max}]$ where $p_z^{max}$ is the supremum of

[8] As shown by Heckman and Vytlacil (2001a), we can identify $B^{ATE}(x)$ and $B^{TT}(x)$ under weaker conditions than those required to follow this strategy of first identifying $B^{MTE}(x, u)$ over the appropriate support.

Supp($P|Z = z$). This support requirement in turn requires that, for fixed $Z = z$, there is enough variation in the outcome shifters $X$ to drive the probabilities to zero.

Finally, consider identification of the surplus parameters. Using that $S^{MTE}(x, z, u_S) = B^{MTE}(x, u_S) - C^{MTE}(z, u_S)$, we identify the marginal surplus parameter at $(x, z, u_S)$ such that $(x, u_S) \in \text{Supp}(X, P)$ and $(z, u_S) \in \text{Supp}(Z, P)$. By equation (2.2.7), we can integrate $S^{MTE}(x, z, u_S)$ using the appropriate weights to identify $S^{ATE}(x, z)$ and $S^{TT}(x, z)$ under the appropriate support conditions. For example, we identify $S^{ATE}(x, z)$ if $\text{Supp}(P|X = x) = [0, 1]$ and $\text{Supp}(P|Z = z) = [0, 1]$.

Thus, for identification of the benefit parameters we need sufficient variation in cost shifters conditional on the outcome shifters. For identification of the cost parameters, we need sufficient variation in the outcome shifters conditional on the cost shifters. For identification of the surplus parameters we need sufficient variation in both sets of regressors. We can thus identify the marginal cost, the average cost, and the cost of treatment on the treated parameters without direct information on the cost of treatment. Consequently, we can also identify the corresponding surplus parameters as well. Our ability to do so is directly related to the extent of variation in observed regressors that shift the benefit of the treatment.

We summarize our discussion in the form of a theorem:

**Theorem 2.** *Assume that equations (2.2.1)–(2.2.4) and our assumptions (A-1)–(A-4) hold.*

1. $B^{MTE}(x, u_S)$ *is identified for* $(x, u_S) \in$ *Supp*$(X, P)$; $C^{MTE}(z, u_S)$ *is identified for* $(z, u_S) \in$ *Supp*$(Z, P)$; *and* $S^{MTE}(x, z, u_S)$ *is identified for* $(x, z, u_S)$ *such that* $(x, u_S) \in$ *Supp*$(X, P)$ *and* $(z, u_S) \in$ *Supp*$(Z, P)$.

2. $B^{ATE}(x)$ *is identified if Supp*$(P|X = x) = [0, 1]$; $C^{ATE}(z)$ *is identified if Supp*$(P|Z = z) = [0, 1]$; $S^{ATE}(x, z)$ *is identified if Supp*$(P|X = x) = [0, 1]$ *and Supp*$(P|Z = z) = [0, 1]$.

3. $B^{TT}(x)$ *is identified if Supp*$(P|X = x) = [0, p_x^{max}]$; $C^{TT}(z)$ *is identified if if Supp*$(P|Z = z) = [0, p_z^{max}]$; $S^{TT}(x, z)$ *is identified if Supp*$(P|X = x) =$

$[0, p_x^{max}]$ *and* $Supp(P|Z = z) = [0, p_z^{max}]$.

**Remark 2.** As discussed in Remark 1, if there is no unobserved heterogeneity in costs of treatment, $U_C = 0$, then $C^{MTE}(z, u_S) = C^{TT}(z) = C^{ATE}(z)$. In the case of no unobserved heterogeneity in costs of treatment, we identify the cost of treatment on the treated and average cost parameters without the additional support conditions. Likewise, if there is no unobserved heterogeneity in the treatment effects, $U_1 - U_0 = 0$, we have $B^{MTE}(z, u_S) = B^{TT}(z) = B^{ATE}(z)$ and can thus identify all of the treatment effect parameters without additional support conditions.

We have thus far considered identification of $B^{ATE}(x) = \mu_1(X) - \mu_0(X)$, and of $C^{ATE}(z) = \mu_C(z)$. We can identify $\mu_1(X) - \mu_0(X)$ and $\mu_C(z)$ up to a location shift under slightly weaker conditions than those required to fully identify the functions. From the analysis of the previous section, we identify $B^{MTE}(x, p) = \mu_1(x) - \mu_0(x) + E(U_1 - U_0|U_S = p)$. By varying $x$ and holding $p$ constant, we trace out $\mu_1(x) - \mu_0(x)$ up to an additive constant. Likewise, consider $C^{MTE}(z, p) = \mu_C(z) + E(U_C|U_S = p)$. By varying $z$ and holding $p$ fixed for the marginal cost parameter, we identify $\mu_C(z)$ up to an additive constant. Given the preceding identification analysis, we can identify $B^{MTE}(x, p)$ over $x \in \text{Supp}(X|P = p)$ and $C^{MTE}(z, p)$ over $z \in \text{Supp}(Z|P = p)$, but not over the unconditional supports of $X$ and $Z$. Thus, we have identification of shifts in $\mu_1(x) - \mu_0(x)$ for $x \in \text{Supp}(X|P = p)$ and of $\mu_C(z)$ for $z \in \text{Supp}(Z|P = p)$ for some $p \in \text{Supp}(P)$. We do not identify $\mu_C(z_0) - \mu_C(z_1)$ if there does not exist a $p$ such that $z_0, z_1 \in \text{Supp}(Z|P = p)$. However, given a rank condition, we can combine information across different values of $p$ to identify $\mu_C(z)$ and $\mu_1(x) - \mu_0(x)$ up to an additive constant for all $z$ and $x$ in their unconditional supports. In particular, consider the following rank assumption:

(A-5) *$X$ and $P(X, Z)$ are measurably separated, i.e., any function of $X$ that almost surely equals a function of $P(X, Z)$ must be almost surely equal to a constant.*

**Theorem 3.** *Assume that equations (2.2.1)–(2.2.4) and our assumptions (A-1)–(A-5) hold. Then $\mu_C(\cdot)$ is identified over the support of $Z$ up to an additive constant, and $\mu_1(\cdot) - \mu_0(\cdot)$ is identified over the support of $X$ up to an additive constant.* □

For proof, see Appendix A

Measurable separability between $X$ and $P$ is a rank condition. As discussed by Florens et al. (2008), measurable separability is a relatively weak regularity condition in this context. See their paper for more discussion of this condition, including sufficient conditions for measurable separability.

Finally, we consider testable implications of $E(Y|X = x, P = p)$ as a function of $p$ that result from additional restrictions including those considered in Theorem 1.

**Theorem 4.** *Assume that equations (2.2.1)–(2.2.4) and (2.2.4) and our assumptions (A-1)–(A-4) hold.*

1. *Suppose that $U_1 - U_0$ is degenerate. Then $E(Y|X = x, P = p)$ is linear in $p$.*

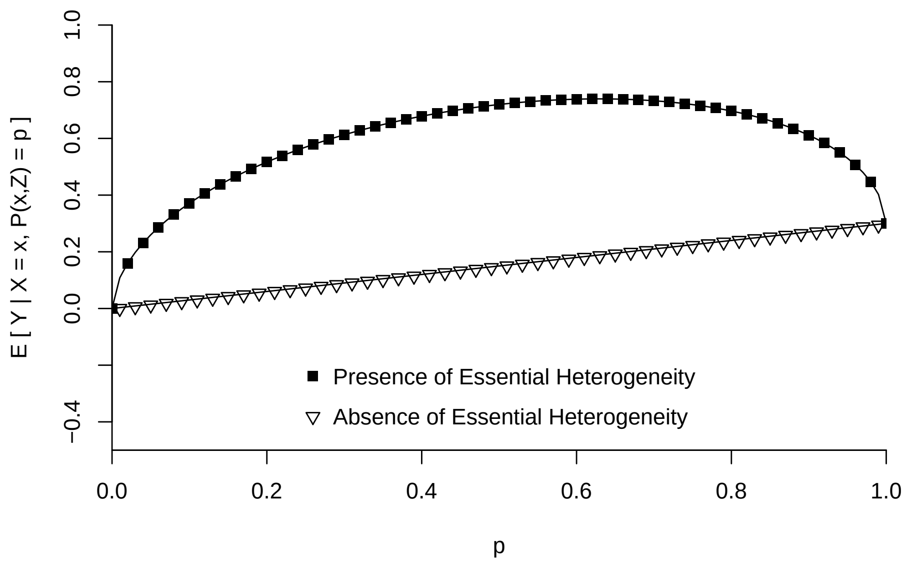
2. *Suppose $U_1 - U_0 \perp\!\!\!\perp U_C$. For a fixed $x$, consider a line $a(x) + b(x)p$, where $a(x) = E(Y|X = x, P(X, Z) = 0)$ and $b(x) = E(Y|X = x, P(X, Z) = 1) - E(Y|X = x, P(X, Z) = 0)$. Then $E(Y|X = x, P(X, Z) = p) \geq a(x) + b(x)p$ for all $p \in$Supp$(P|X = x)$.*

3. *Suppose $U_1 - U_0 \perp\!\!\!\perp U_C$, and suppose $U_1 - U_0$ and $U_C$ have log concave densities. Then $E(Y|X = x, P(X, Z) = p)$ is a concave function of $p$.* □

For proof, see Appendix A.

We now return to our numerical example. Recall that the unobservables in the numerical example are all normal but independently distributed, Figure 2.2 depicts the corresponding $E\left(Y \mid X = x, P(X, Z) = p\right)$. As predicted by Theorem 4, $E\left(Y \mid X = x, P(X, Z) = p\right)$ is a concave function of $p$ in this example. This is a direct consequence of the fact that those agents with a high propensity of treatment (low values of $u_S$) have the highest gains even after conditioning on observables. As $p$ increases, the share of individuals participating increases constantly, but at the same time the gain for agents at the margin decreases. Individuals with high values of $V$, who enter treatment only for high values of $p$, have the least to gain from treatment.

**Figure 2.2:** Testable Implication




## 2.3.2 Limited Information

Our analysis thus far has assumed choice equation (2.2.4), that $D = \mathbf{1}[S \geq 0]$ where $S = (Y_1 - Y_0) - C$. This implicitly assumes that agents have perfect foresight of their gross individual benefit of treatment $B = Y_1 - Y_0$ as well as cost $C$. In this section, we relax the choice model of equation (2.2.4) to allow limited information on the part of the agents, while maintaining the model for latent outcomes $(Y_0, Y_1)$ and cost $C$ of equations (2.2.2) and (2.2.3). We assume that agents form valid expectations of their outcomes and costs given the information that they have at the time of their treatment choice and that they select into treatment if the expected surplus is positive. We allow agents to know only some elements of $(X, Z)$, and to possibly have incomplete knowledge of $(U_0, U_1, U_C)$ and thus of their own idiosyncratic benefit and cost of treatment. We now show that the preceding analysis goes through with minor modifications, though it is now important to distinguish conditioning sets: what is known to the agent at the time of treatment choice (which might include some information not known to the econometrician), what is known to the econometrician (which might include some information not known to the agent at the time of treatment choice), and what is realized ex post. The essential change in our procedure in the case of incomplete information is that the marginal benefit of treatment identified by LIV must be projected onto the information set when selecting treatment to form the expected marginal benefit of treatment conditional on the information available to the agent. It is this coarsened version of $B^{MTE}$ that is used to identify the marginal cost parameter. In addition, only components of $X$ that are known to the agent at the time of treatment choice can aid in identification of the cost parameters. The exclusion restrictions for identification of the cost parameter are variables in $X$ that are not in $Z$ and that were known to the agent at the time of treatment selection.

Let $(X_I, Z)$ denote components of $(X, Z)$ that are observed by the agent when choosing whether to select into treatment.[9] Suppose that the agent's information set equals $(X_I, Z, U_I)$.[10] $U_I$ is the private information of the agent relevant to their own benefits and cost of treatment, and it is not observed by the econometrician.

We restate assumption (A-1) in the following way:

(A-6) *$(U_0, U_1, U_C, U_I)$ is independent of $(X, Z)$, and $X$ is independent of $Z$ conditional on $X_I$.*

Assumption (A-6) imposes the requirement that the private information of the agent is independent of the observed regressors. Note that, under this independence assumption

$$(U_0, U_1, U_C, U_I) \perp\!\!\!\perp (X_I, Z)$$

and

$$E(V|X, Z, U_I) = E(V|X_I, Z, U_I) = E(V|U_I),$$

using the definition $V = U_C - (U_1 - U_0)$. Assumption (A-6) implies that $(X, Z) \perp$

[9] We are assuming that agents know all components of $Z$, while allowing the agents to be ignorant of some components of $X$, as this restriction simplifies our notation and conforms to our empirical example of Section 2.4. The analysis directly extends (at the cost of somewhat more cumbersome notation) to allow agents to know only a subvector of $Z$ as well as only a subvector of $X$ at the time of selection into treatment.

[10] In other words, the information set of the agent equals $\sigma(X, Z, U_I)$, the sigma-algebra generated by $(X, Z, U_I)$.

$\perp U_I \mid (X_I, Z)$, so that $U_I$ does not help the agent predict elements of $(X, Z)$ that are not contained in $(X_I, Z)$. Thus, we allow the agents to have private information about their own idiosyncratic benefits $(U_1 - U_0)$ and costs $U_C$, though we invoke the restriction that the agents only know $(X_I, Z)$. Furthermore, Assumption (A-6) requires that, conditional on the components of $X$ known to the agent at the time of selecting into treatment, $Z$ does not help to predict those elements of $X$ not known at the time of treatment selection. This restriction is only imposed for notational convenience and can be easily relaxed.

We restate assumption (A-3) as:

(A-7) *The distribution of $\tilde{V} = E(V|U_I)$ is absolutely continuous with respect to Lebesgue measure, and the cumulative distribution function of $\tilde{V}$ is strictly increasing.*

An implication of (A-7) is that $E(V|U_I)$ is a nondegenerate random variable, and thus that agents have some nontrivial information on their own idiosyncratic cost or benefit from treatment when deciding whether to select into treatment. We maintain assumptions (A-2) and (A-4) as before.

Define $\mu_j^I(X_I) = E(Y_j|X_I)$ for $j = 0, 1$, and $\mu_C^I(Z) = E(C|Z)$, and note that given our independence assumptions and the law of iterated expectations, $\mu_j^I(X_I) = E(\mu_j(X)|X_I)$, $\mu_C^I(Z) = E(\mu_C(Z)|Z)$. Define $\mu_S^I(X_I, Z) = E(S|X_I, Z)$. Under our assumptions,

$$E(S|X_I, Z, U_I) = \mu_S^I(X_I, Z) - \tilde{V} = \mu_1^I(X_I) - \mu_0^I(X_I) - \mu_C^I(Z) - \tilde{V}.$$

The previous decision rule under perfect certainty, equation (2.2.4), is now replaced with

$$\text{(2.3.2)} \quad D = 1 \quad \text{if} \quad E(S|X_I, Z, U_I) \geq 0\,; \qquad D = 0 \quad \text{otherwise,}$$

where $E(S|X_I, Z, U_I)$ is the expected surplus from treatment, with the expectation conditional on the agents information set. We thus have

$$D = \mathbf{1}[\mu_S^I(X_I, Z) - \tilde{V} \geq 0],$$

where our independence assumptions imply $\tilde{V} \perp\!\!\!\perp (X_I, Z)$, and thus the selection model is of the same form as that used by Heckman and Vytlacil (1999), which allows us to use LIV to identify $B^{MTE}$. Redefining $U_S = F_{\tilde{V}}(\tilde{V})$ and $P(X_I, Z) = \Pr[D = 1|X_I, Z] = F_{\tilde{V}}(\mu_S^I(X_I, Z))$, we have

$$D = \mathbf{1}[P(X_I, Z) - U_S \geq 0],$$

with $U_S$ distributed unit uniform and independent of $(X, Z)$ and thus independent of $(X_I, Z)$.

Define $B_I^{MTE}(x_I, u_S) \equiv E(Y_1 - Y_0|X_I = x_I, U_S = u_S)$, $C_I^{MTE}(z, u_S) \equiv E(C|Z = z, U_S = u_S)$, and $S_I^{MTE}(x_I, z, u_S) \equiv B_I^{MTE}(x_I, u_S) - C_I^{MTE}(z, u_S)$, the marginal benefit, cost, and net surplus of treatment conditional on the agent's information set, where again by the law of iterated expectations and our independence assumptions

$$\begin{aligned}
B_I^{MTE}(x_I, u_S) &= E(B^{MTE}(X, u_S)|X_I = x_I, U_S = u_S) \\
&= E(B^{MTE}(X, u_S)|X_I = x_I) \\
C_I^{MTE}(z, u_S) &= E(C^{MTE}(Z, u_S)|Z = z, U_S = u_S) \\
&= E(C^{MTE}(Z, u_S)|Z = z).
\end{aligned}$$

Evaluating $S_I^{MTE}(x_I, z, u_S)$ at $u_S = P(x_I, z)$, we obtain

$$\begin{aligned}
S_I^{MTE}(x_I, z, P(x_I, z)) &= \mu_S^I(x_I, z) - E(V|U_S = P(x_I, z)) \\
&= \mu_S^I(x_I, z) - E(V|\tilde{V} = \mu_S^I(x_I, z)) \\
&= \mu_S^I(x_I, z) - E(V|E(V|U_I) = \mu_S^I(x_I, z)) \\
&= \mu_S^I(x_I, z) - E(E(V|U_I)|E(V|U_I) = \mu_S^I(x_I, z)) \\
&= \mu_S^I(x_I, z) - \mu_S^I(x_I, z) \\
&= 0,
\end{aligned}$$

where the second equality is obtained by plugging in the definition of $U_S$, the third equality is obtained by plugging in the definition of $\tilde{V}$, and the fourth equality is obtained using the law of iterated expectations and the fact that $E(V|U_I)$ is degenerate given $U_I$. Since $S_I^{MTE}(x_I, z, u_S) = B_I^{MTE}(x_I, u_S) - C_I^{MTE}(z, u_S)$, we have

$$B_I^{MTE}(x_I, u_S) = C_I^{MTE}(z, u_S) \text{ for } u_S \text{ such that } u_S = P(x_I, z).$$

Thus, identification of $B_I^{MTE}(x_I, P(x_I, z))$ provides identification of $C_I^{MTE}(z, P(x_I, z))$.

Since our model is a special case of Heckman and Vytlacil (1999), we can follow them in using LIV to identify $B^{MTE}(x, u_S)$ for $(x, u_S)$ in the support of $(X, P(X_I, Z))$. It is important to note that LIV does not identify the $B^{MTE}$ that is relevant to the agent's decision problem. LIV identifies $B^{MTE}(x, u_S) = E(Y_1 - Y_0|X = x, U_S = u_S)$, not $B_I^{MTE}(x_I, u_S) = E(Y_1 - Y_0|X_I = x_I, U_S = u_S)$. However, we can project the $B^{MTE}$ identified by LIV on the information known to the agent at the time of treatment selection and coarsen the set used to define and identify $B^{MTE}$, to identify the $B_I^{MTE}$ relevant to the agent's decision problem. It is the latter that is relevant for identifying cost. By the law of iterated expectations (and using that $X_I$ is degenerate given $X$ and that $U_I$ is independent of $X$), we obtain

(2.3.3)
$$B_I^{MTE}(x_I, u_S) = E(B^{MTE}(X, u_S)|X_I = x_I) = \int B^{MTE}(x, u_S) dF_X(x|X_I = x_I),$$

where $F_X(\cdot|X_I = x_I)$ is the cumulative distribution function of $X$ conditional on $X_I = x_I$. We directly identify $F_X(\cdot|X_I = x_I)$, and thus obtain identification of $B^{MTE}(x, u_S)$ for all $x \in \text{Supp}(X|X_I = x_I)$ which in turn implies identification of $B_I^{MTE}(x_I, u_S)$. Since, for a given $x$, we identify $B^{MTE}(x, u_S)$ if $u_S \in \text{Supp}(P(X_I, Z)|X = x)$, we thus identify $B_I^{MTE}(x_I, u_S)$ if

$$u_S \in \bigcap_{x \in \text{Supp}(X|X_I = x_I)} \text{Supp}(P(X_I, Z)|X = x).$$

In other words, to identify ex ante $B_I^{MTE}(x_I, u_S)$, we need to identify ex post $B^{MTE}(x, u_S)$ for every value $x$ that $X$ can take given $X_I = x_I$, and thus we need for $u_S$ to be an element of $\text{Supp}(P(X_I, Z)|X = x)$ for each value $x$ that $X$ can take given $X_I = x_I$. However, using the fact that $X_I$ is a subvector of $X$ and independence Assumption (A-6), it follows that $\text{Supp}(P(X_I, Z)|X) = \text{Supp}(P(X_I, Z)|X_I)$, and thus using equation 2.3.3 we identify $B_I^{MTE}(x_I, u_S)$ for $(x_I, u_S)$ in the support of $(X_I, P(X_I, Z))$. Using the fact that $B_I^{MTE}(x_I, P(x_I, z)) = C_I^{MTE}(z, P(x_I, z))$, we identify $C_I^{MTE}(z, u_S)$ for $(z, u_S)$ in the support of $(Z, P(X_I, Z))$. We have thus identified the marginal cost parameter, and can integrate it to obtain other cost parameters. We can also combine it with the benefit parameters to identify net surplus parameters as before. The only elements of $X$ that are useful for identifying the cost parameters are those elements which are known to the agent at the time of selection into treatment.

## 2.4 Empirical Illustration

### 2.4.1 Data and Estimation Strategy

Carneiro et al. (2011) estimate the marginal benefit of attending college for a sample of white males from the National Longitudinal Survey of Youth of 1979 (NLSY). We extend their analysis and apply our methodology to estimate the cost and surplus of attending college for the same sample and specification. In particular, following Carneiro et al. (2011), we separate individuals into two groups: persons with no college (D = 0) and persons with at least some college (D = 1). The outcome variable is the natural logarithm of the mean non-missing values of the hourly wage between 1989 and 1993. Schooling is measured in 1991 when individuals are between 28 and 34 years of age. To complete the setup of the generalized Roy model presented in Section 2.2, we need to specify the observed regressors that determine an agent's expected benefit and cost of treatment. The outcome equation contains all variable that determine the ex post benefits of treatment. For the choice equation, we specify the observed variables that are known to the agent at the time of the treatment decision and that affect either the (ex ante) benefits or costs of treatment.

Table 2.1 provides a full account of the observed variables used in our empirical analysis. We highlight two types of exclusion restrictions. To identify the $B^{MTE}(x, u_S)$, we require cost-shifters that are excluded from the outcome equations and thus do not affect the benefits of treatment. Short-run fluctuations in local labor market conditions at the time of the educational decision shift the opportunity cost of schooling, but do not directly affect future wages given permanent local labor market conditions. We also include tuition cost and distance to college as shifters that

**Table 2.1:** Specification

| | Outcome | *ex post* Benefits | Choice | *ex ante* Benefits | Cost |
|---|---|---|---|---|---|
| Years of Experience (in 1991) | X | X | | | |
| Current Local Wages (in 1991) | X | X | | | |
| Current Local Unemployment (in 1991) | X | X | | | |
| Permanent Local Unemployment (1973-2000) | X | X | X | X | |
| Permanent Local Wages (1973-2000) | X | X | X | X | |
| AFQT Scores | X | X | X | X | X |
| Mother's Education | X | X | X | X | X |
| Number of Siblings | X | X | X | X | X |
| Urban Residence | X | X | X | X | X |
| Cohort Dummies | X | X | X | X | X |
| Local Presence of Public College (age 14) | | | X | | X |
| Local Tuition at Public College (age 17) | | | X | | X |
| Local Wages (age 17) | | | X | | X |
| Local Unemployment (age 17) | | | X | | 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 1,747 white males, where 865 did receive some college education while 882 do not.

affect the direct cost of going to college. These variables enter the selection equation but are excluded from the outcome equations. To identify $C_I^{MTE}(z, u_S)$, we require access to variables that shift the benefit of treatment and are known to the agent at the time of selection but do not affect the cost of treatment. For this, we use information on the long-run labor market conditions in the location of residence in adolescence. Again, as we condition on current labor market conditions at the time of treatment, these regressors should affect the schooling decision only through their direct affect on future wages.

In addition, three variables are included in the outcome equations but not the selection equation because they are not in the individual's information set at the time of the college attendance decision. These are as follows: years of experience, earnings in the county of residence, and unemployment in the state of residence. All are measured in 1991, which is approximately 12 years after the agents' schooling decision is taken. We condition on earlier values of these variables in the selection equation. With this specification, we follow the analysis of Section 2.3.2 and allow agents to have imperfect foresight about future benefits. They only pursue a higher education, if their expected benefits exceed the cost.

As common elements (observables affecting benefits as well as cost of treatment) we include the Armed Forces Qualifying Test (AFQT) scores, mother's education, number of siblings, as well as dummy variables indicating urban residence at age 14, and cohort dummies. However, in what follows we keep this set of observables in the background to ease notation. Thus $X$ continues to denote the variables that shift the outcome equations (thus the benefit of treatment) but are excluded from the cost equation. The opposite is true for $Z$. $X_I$ denotes the subvector of $X$ that is known to the agent at the time of the decision of whether or not to attend college. Additional information about the data is provided in Appendix B.

Throughout we impose full independence between all observables and unobservables $(U_1, U_0, U_C)$ of the model, and specify a linear-in-parameters version of the generalized Roy model:

$$Y_1 = X\beta_1 + U_1$$
$$Y_0 = X\beta_0 + U_0$$
$$D = \mathrm{I}\left[X_I(\alpha_1 - \alpha_0) - Z\delta > U_C - (U_1 - U_0)\right].$$

Note that $(\beta_1 - \beta_0)$ is potentially different from $(\alpha_1 - \alpha_0)$ as $X_I$ not only affects the returns to education directly, but also helps to predict the ex post realization of $X \setminus X_I$. However, we impose $\mathrm{E}\left[X(\beta_1 - \beta_0) \mid X_I\right] = X_I(\alpha_1 - \alpha_0)$, thus agents form valid expectations about their outcomes. We present annualized returns to education, obtained by dividing our estimates by four which is the average difference in years of schooling between those with $D = 1$ and those with $D = 0$.

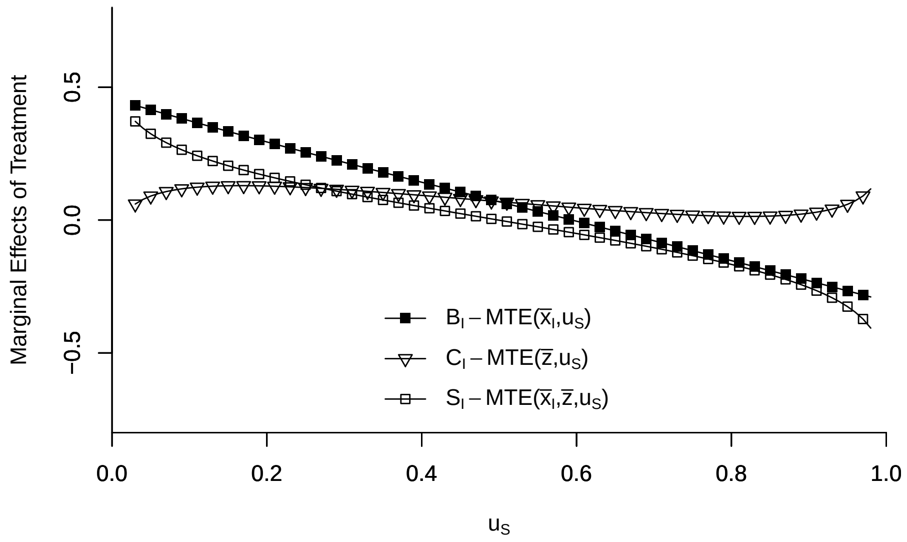
We implement the semiparametric LIV approach proposed in Heckman et al. (2006b). Given the remaining set of assumptions, the ex post benefit of treatment is given by $B^{MTE}(x, u_S) = x(\beta_1 - \beta_0) + \mathrm{E}\left[U_1 - U_0 \mid U_S = u_S\right]$. We estimate $\mathrm{E}\left[Y \mid X, P\right] = X\beta_0 + PX(\beta_1 - \beta_0) + K(P)$ and take its derivative with respect to $P$. We apply the partially linear regression method of Robinson (1988) to get an estimate for $(\beta_1, \beta_0)$. $K(\cdot)$ is estimated using locally quadratic regression. We then project the estimate for the $B^{MTE}(x, u_S)$ on the information set of the agent and use the resulting estimate of $(\alpha_1 - \alpha_0)$ as the scale normalization for the estimation of a Probit selection model to back out the corresponding marginal surplus and cost parameters.

Throughout, we assume that $(U_1, U_0, U_C)$ is independent of $(X, Z)$. One important consequence of imposing this assumption is that the marginal effects of treatment are identified over the marginal support of $P(X_I, Z)$. In our sample the range of support of $P(X_I, Z)$ is between 0.03 and 0.97 and does not cover the full unit interval. Thus, the conventional average effect parameters like the ATE are not identified in our data without additional parametric assumptions.

### 2.4.2 Results

Figure 2.3 presents our results for the ex ante marginal benefit, marginal cost, and marginal surplus parameters. We plot them as a function of $u_s$ and evaluate them at the sample mean of $(X_I, Z)$.

**Figure 2.3:** Marginal Effects of Treatment




Individuals with a high unobserved desire for treatment (low $u_S$) have the highest surplus and highest benefits from treatment participation. The estimated surplus is strongly positive for low values of $u_S$ and drops rapidly to be strongly negative for high values of $u_S$. The marginal benefits of treatment range from +43% to −28% and are strictly decreasing along the margins of $V$. For the marginal cost, the picture is more mixed. They are positive across the whole range of $V$ and range from 1% to 12%. Moving along $u_s$, they increase at first and then decline midrange before picking up again for high values of $u_s$. They are highest for those least likely to pursue a higher education.

Given our point of evaluation, the individuals at the margin are $u_S = P(x_I, z) = 0.49$. Individuals at this quantile of $V$ are just indifferent towards treatment. Their surplus is zero, as benefits are just offset by the cost incurred from participating.

## 2.5 Conclusion

Building on the pioneering analysis of Björklund and Moffitt (1987) this chapter extends the analysis of the marginal effects of treatment by Heckman and Vytlacil (1999, 2005, 2007b) using the marginal benefit of treatment $B^{MTE}$ to identify the subjective cost and surplus of treatment. We consider costs with perfect and imperfect foresight. An analysis of college going finds unobserved heterogeneity in both the benefits and costs of attending college, with agents selecting into college based on their idiosyncratic expected benefit and perceived cost of attending college. We find more heterogeneity in benefits than in subjective cost. Thus the variability in perceived benefits drives college attendance more than the variability in the expected costs.

# 3 Optimal Treatment Reallocation

## 3.1 Introduction

Econometric policy evaluation informs policy makers about the performance of alternative policy interventions. Its findings influence decisions on whether to continue or abolish a given program. However, exploiting possible effect heterogeneity allows to improve the current implementation of a program even without an augmentation of resources. Average outcomes are already affected by their mere reallocation.

Such considerations are particularly policy relevant in the context of active labor market programs (ALMP). This is true for several reasons. First, some unemployed seem to benefit more from treatment than others (Heckman et al., 1999a). Second, the role of resource constraints is particularly evident due to the large pool of the unemployed. Third, there are multiple competing approaches, and policy makers exercise considerable control over the assignment process of the unemployed to the numerous programs. And fourth, these programs consume a considerable amount of public funds, are subject to a high level of scrutiny, and feature prominently in the public debate.

In this chapter, we contribute to the empirical literature on optimal treatment assignment of active labor market programs. This literature predicts agents' outcomes

This chapter profited greatly from comments by James Heckman, Markus Frölich, Andrea Weber, Bernd Fitzenberger, Gesine Stephan, Christian Goldammer, Bryan Graham, Jens Mohrenweiser, Thomas Walter, Holger Stichnoth, Arne Uhlendorff, Stephan L. Thomsen, Steffen Reinhold, Stephan Dlugosz, and Hans-Martin von Gaudecker. I thank George Yates and Ben Skrainka for their advice regarding the computational implementation.

under alternative treatments and counterfactually assigns them according to some rule. The goal is to compare average benefits between the current and the counterfactual assignment to investigate possible improvements in the efficiency of such programs by the adoption of alternative assignment rules. Prominent examples include Lechner and Smith (2007), Staghoej et al. (2010), and the works by Eberts et al. (2002) and OECD (2002, 1998).

We extend this literature by explicitly linking the assignment mechanism to a criterion of optimality while addressing the presence of resource constraints. These constraints can take various forms. In the short run, the number of treatment slots is fixed due to local supply constraints or contractual agreements. In the long run, budget constraints matter. In this chapter, we focus on short run constraints and take the number of treatment slots as fixed throughout. When this is the case and there are multiple competing treatments, then the determination of an optimal assignment with respect to a specified criterion is a multidimensional task. Hence, these constraints fundamentally change the complexity of the assignment problem as not all agents can be assigned their first best.

We illustrate the relevance of our extensions with an application to the National Evaluation of Welfare-to-Work Strategies (NEWWS) dataset. NEWWS was an evaluation of welfare-to-work job training programs in the United States, where two competing strategies were implemented to determine their relative effectiveness. One strategy, called human capital development (HCD), focused on basic education and vocational skills training. The other strategy, labor force attachment (LFA), focused on job search and quickly finding employment, even at reduced wages. The program was evaluated by randomized assignment. These results provide us the benchmark for a comparison against alternative assignment rules.

To preview our main findings, we establish that the selected assignment mechanism has a significant effect on the scale of the program, its overall impact, and the relative effectiveness of the HCD and LFA approach. However, all these performance measures are affected by the choice of the objective function and its implementation. Resource constraints and limited information by policy makers restrict the feasible improvements by targeted treatment assignments. Finally, at least in small samples, uncertainty about the impact of an LFA or HCD assignment prohibits general statements regarding the relative performance of alternative assignment algorithms.

Our results confirm that the choice of the assignment mechanism is a powerful component of an optimal policy design. However, our considerations of resource constraints and limited information point to important caveats in the context of active labor market programs.

The plan of this chapter is as follows. Section 3.2 presents our conceptual framework to explore reallocation policies and links them to the econometrics of policy evaluation. We also discuss reallocation parameters and contrast them to the conventional treatment effect parameters. Section 3.3 provides a brief review of the related literature. In Section 3.4, we develop our empirical illustration. Section 3.5 summarizes our existing work, states its limitations, and previews our future research.

## 3.2 Conceptual Framework

We start by presenting our conceptual framework. We establish the required notation, discuss the role of assignment mechanisms for the efficiency of social programs, and define the objects of interest. We contrast the conventional treatment effects to reallocation effects by linking both parameters to the policy questions they address.

We follow the notation established in Heckman and Vytlacil (2007a) closely. Let $\Omega$ denote the set of agents $\omega \in \Omega$ and $\mathcal{S}$ the set of possible treatments with elements $s \in \mathcal{S}$. We define the outcome for agent $\omega$ receiving treatment $s$ as $Y(s, \omega)$. The $Y(s, \omega)$ are objective outcomes realized after treatment is selected. In principle, agent $\omega$ could participate in any of the programs in $\mathcal{S}$. Thus, we obtain a collection of potential outcomes $\{Y(s, \omega)\}_{s \in \mathcal{S}}$ for each agent $\omega$. $Y(s_0, \omega)$ may be the outcome for agent $\omega$ in the absence of any treatment participation, while $Y(s_1, \omega)$ is the outcome in treatment state $s_1$. We refer to the mechanisms that allocate agents $\omega$ to treatment $s$ as $\tau : \Omega \to \mathcal{S}$ and collect them in $\mathcal{T}$. Applying mechanism $\tau$ to the population $\Omega$ results in the allocation $\alpha \in \mathcal{A}$. Let $D_\tau(s, \omega)$ take value one if agent $\omega$ is assigned to treatment $s \in \mathcal{S}$ under mechanism $\tau$ and zero otherwise. Then, the observed outcome $Y(\omega)$ of agent $\omega$ is defined as

$$Y(\omega) = \sum_{s \in S} D_\tau(s,\omega) Y(s,\omega).$$

For each program $s$ there are $\bar{d}_s$ slots available. So, the following inequality must hold for each allocation $\alpha$ resulting from assignment mechanism $\tau$

$$\sum_{\omega \in \Omega} D_\tau(s,\omega) \leq \bar{d}_s \qquad \forall \quad s \in \mathcal{S}.$$

In our setting of active labor market programs, the set of agents corresponds to a population of unemployed who are potentially subject to alternative treatments such as job search assistance or public employment. The outcome refers to subsequent measures of labor market success, e.g. time unemployed or earnings in the next job. Possible assignment mechanisms that match agents to programs include random assignment, caseworker discretion, or self-selection.

Policy relevance of our analysis requires not only that there is treatment effect heterogeneity and that resources are scarce, but also that the allocation can be affected by policy makers. This is true in a variety of settings, but to varying degrees. Often, this ability is only indirect by affecting the incentives faced by agents when making their decisions, e.g. in education policy with tuition subsidies. However, in the case of active labor market programs, policy makers can directly assign the unemployed into different programs.

Figure 3.1 illustrates the role of alternative assignment mechanisms in matching agents to alternative treatments. We consider a pool of four unemployed $\{\omega_i\}_{i=1,..,4}$, four alternative treatments $\{s_j\}_{j=1,..,4}$, and three alternative assignment mechanisms $\{\tau', \tau'', \tau'''\}$. There are four treatment slots available: job application training ($s_1$), job training program ($s_2$), subsidized public employment ($s_3$), and job search assistance ($s_4$). The program slots are matched to agents either by random assignment ($\tau'$), by caseworker discretion ($\tau''$), or by self-selection ($\tau'''$).

Under random assignment ($\tau'$), $\omega_1$ receives a job application training, while $\omega_2$ takes up a job training program. A switch to caseworker assignment ($\tau''$) has no effect

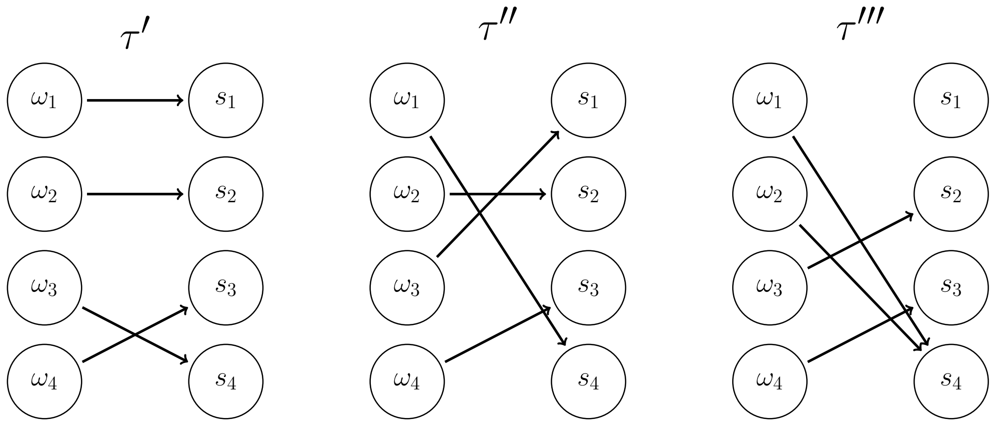



**Figure 3.1:** Assignment Examples

on $\omega_2$'s program, but $\omega_1$ is now sent to a job search assistance program. However, self-selection by the unemployed ($\tau'''$) leads to an unfeasible allocation as the resource constraints are violated. In this case, $\omega_1$ and $\omega_2$ both opt for the job search assistance program. Agent $\omega_4$ ends up in public employment under any mechanism. So, different mechanisms lead to agents receiving different treatments. Some mechanisms may yield allocations that violate the resource constraints.

So far, we have focused on the assignment mechanisms and the resulting matches between agents and programs. However, each match affects subsequent outcomes. Next, we turn to potential parameters of interest that capture this effect.

**Individual Effect of Treatment** The individual effect of treatment $\Delta_{s_j,s_k}$ for agent $\omega$ is the difference in objective outcomes across treatments $s_j$ and $s_k$

$$\Delta_{s_j,s_k} = Y(s_j,\omega) - Y(s_k,\omega)\,, \quad s_j \neq s_k.$$

This is the individual level causal effect. If treatment affects agents in different ways, then $\Delta_{s_j,s_k}$ varies in the population and there is treatment effect heterogeneity.

Unfortunately, we rarely observe the same agent in different states and are thus not able to identify the causal effect of treatment on an individual level. That is why econometricians reformulate the parameter of interest and focus on the average effect of treatment at the population level instead.

We now present the conventional treatment effect parameters and contrast them to reallocation effects. For both, we focus on the policy questions they address and the role of resource constraints.

**Conventional Treatment Effects** Economists usually focus their attention on the average treatment effect (ATE), the average treatment effect on the treated (TT), and the average treatment effect on the untreated (TUT). Comparing treatment $s_j$ to $s_k$, these are

$$\begin{aligned} ATE(s_j, s_k) &= \mathrm{E}\left[Y(s_j, \omega) - Y(s_k, \omega)\right] \\ TT(s_j, s_k) &= \mathrm{E}\left[Y(s_j, \omega) - Y(s_k, \omega) \mid D(s_j, \omega) = 1\right] \\ TUT(s_j, s_k) &= \mathrm{E}\left[Y(s_j, \omega) - Y(s_k, \omega) \mid D(s_j, \omega) = 0\right]. \end{aligned}$$

The $ATE(s_j, s_k)$ captures the average effect of moving a random individual of the population from base state $s_k$ to treatment $s_j$. Moreover, it is also the effect of moving a whole population from a universal policy $s_k$ to a universal regime of $s_j$. The $TT(s_j, s_k)$ only considers the subpopulation of those matched with treatment $s_j$. It measures the average effect of the program among that subpopulation compared to alternative $s_k$. The $TUT(s_j, s_k)$ is defined analogously. In the presence of treatment effect heterogeneity these parameters are potentially different depending on the assignment process. If agents select their treatment status based on their own gains, then those who choose treatment $s_j$ and not $s_k$ will profit more from $s_j$, thus $TT(s_j, s_k) > TUT(s_j, s_k)$. If agents are assigned at random, then all parameters are equal. Identification, inference, and estimation of these parameters are well studied in the econometrics of policy evaluation.[1]

However, the conventional treatment effect parameters are of limited policy relevance, and the role of resource constraints remains vague. 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 treatments at random. The $TT$ is informative if the complete elimination of a program already in place is considered. Conversely, if the

[1] See Heckman and Vytlacil (2007b,a) and Blundell and Costas Dias (2009) for excellent surveys.

same program is examined for compulsory participation, then the $TUT$ is the policy relevant parameter. All these scenarios alter the amount of the required resources.

As the role of resource constraints remains unclear, we now present the reallocation effects. They address this issue directly.

**Reallocation Effects** Reallocation policies differ from typical policies as the available resources remain unchanged. It is just the pairing between individuals and treatments that is varied. If the benefits of treatment are heterogeneous in the relevant population, then differences in the allocation affect the efficiency of treatments.

Two useful parameters have been introduced in the econometrics literature by Graham et al. (2007, 2009). They define the *Average Reallocation Effect* and the *Maximum Reallocation Effect*. We adapt both to the context of optimal treatment reallocations. The two parameters consider the efficiency of treatment $s_j$ compared to $s_k$ among the subpopulation that receives treatment $s_j$ under two alternative assignment mechanisms.

First, we define the $ARE$ as

$$ARE(s_j, s_k, \tau', \tau'') = \mathrm{E}\left[\, \Delta_{s_j,s_k} \mid D_{\tau''}(s_j, \omega) = 1\right] - \mathrm{E}\left[\, \Delta_{s_j,s_k} \mid D_{\tau'}(s_j, \omega) = 1\right]$$
$$\text{s.t.} \qquad \sum_{\omega \in \Omega} D_\tau(s_j, \omega) \leq \bar{d}_{s_j} \qquad \forall \quad s \in \mathcal{S} \quad \text{and} \quad \tau \in \{\tau', \tau''\}$$

, where $(\tau', \tau'') \in \mathcal{T}$ denote any two alternative assignment mechanisms. Parts of the subpopulations over which the average is taken, might be identical as some agents are matched with treatment $s_j$ under both assignment rules. Returning to the example of Figure 3.1, the $ARE(s_1, s_4, \tau', \tau'')$ captures the difference in benefits of assigning the job application training $s_1$ (instead of job search assistance $s_4$) to agent $\omega_3$ and not $\omega_1$. The change in the allocation is induced by replacing random assignment ($\tau'$) with assignment at a caseworker's discretion ($\tau''$).

Second, the $MRE$ compares the mechanism $\tau^{\max}$, which maximizes the average benefits among those who end up in treatment $s_j$, to an alternative $\tau'$.

$$\begin{aligned} MRE(s_j, s_k, \tau^*, \tau') = \mathrm{E}\left[\, \Delta_{s_j,s_k} \mid D_{\tau^*}(s_j,\omega) = 1\right] - \mathrm{E}\left[\, \Delta_{s_j,s_k} \mid D_{\tau'}(s_j,\omega) = 1\right] \\ \text{s.t.} \qquad \sum_{\omega\in\Omega} D_\tau(s_j,\omega) \leq \bar{d}_{s_j} \qquad \forall \quad s \in \mathcal{S} \quad \text{and} \quad \tau \in \{\tau^*, \tau'\} \end{aligned}$$

, where $\tau^* = \text{argmax}_{\tau\in\mathcal{T}} \quad \mathrm{E}\left[\, \Delta_{s_j,s_k} \mid D_\tau(s_j,\omega) = 1\right]$.

So far, the reallocation effects deal with each treatment $s$ separately. Yet, a social planner that aims to maximize aggregate overall efficiency has not only to take the efficiency of each treatment into account separately, but also possible side effects of assignments. An allocation rule that maximizes the gains among those participating in treatment $s_j$ potentially results in very low gains among the agents who are assigned to $s_k$. The allocation problem faced by a planner is

(3.2.1)

$$\begin{aligned} \max_{\tau\in\mathcal{T}} \quad & \sum_{s\in\mathcal{S}} \pi_s \, \mathrm{E}\,[\, \Delta_{s,b} \mid D_\tau(s,\omega) = 1\,] \\ \text{s.t.} & \sum_{\omega\in\Omega} D_\tau(s,\omega) \leq \bar{d}_s \qquad \forall \quad s \in \mathcal{S} \quad \text{and} \quad \tau \in \mathcal{T} \end{aligned}$$

, where $\pi_s$ denotes the share of slots available for program $s$ and the subscript $b$ denotes the baseline outcome, i.e. the outcome in the absence of any treatment.

The planner's assignment problem is more complex in the presence of resource constraints and multiple treatments as it is now a multidimensional instead of a one-dimensional task. First, resource constraints lead to some agents not receiving any treatment even though they would benefit from participation. Second, even if an agent is assigned to a treatment, this is not necessarily the one where the individual gains are largest. It is not the absolute advantage but the comparative advantage that determines the aggregate optimal assignment schedule. In the existing literature on optimal treatment assignment these two issues are ignored.

In light of these complications, we draw on operations research for solution methods to assign agents to multiple treatments in the presence of resource constraints. Let

$X = (x_{i,j})$ be a binary assignment matrix such that

$$x_{i,j} = \begin{cases} 1 & \text{if agent } \omega_i \text{ is assigned to treatment } s_j \\ 0 & \text{otherwise.} \end{cases}$$

Now, the planner's maximization problem can be written as a standard linear sum assignment problem (LSAP).

$$\begin{aligned} \max \quad & \sum_{i=1}^{I} \sum_{j=1}^{J} \mathrm{E}\left[\Delta_{s_j, s_b}\right] x_{i,j} \\ \text{s.t.} \quad & \\ & \sum_{j=1}^{J} x_{i,j} = 1 \qquad i = 1, ..., I \\ & \sum_{i=1}^{I} x_{i,j} = \bar{d}_{s_j} \qquad j = 1, ..., J \\ & x_{i,j} \in \{0, 1\} \qquad i = 1, ..., I, \ j = 1, ..., J \end{aligned}$$

The maximization constraints ensure that each agent is assigned to only one treatment and that resource constraints are not violated.

Figure 3.2 demonstrates the difference between the existing work on optimal treatment assignment of ALMP and our work on optimal treatment reallocations.

We return to our previous example but now include the individual effect of each assignment on the agent. We consider the case of four agents $\{\omega_i\}_{i=1,..,4}$ and four alternative treatment slots $\{s_j\}_{j=1,..,4}$. The matrix entry $(i, j)$ contains the reemployment probability of agent $\omega_i$ after participation in treatment $s_j$. Consider Figure 3.2a, in which the light gray indicates each agents' initial assignment. It might be the outcome of a random allocation or a caseworker's assignment. Initially, agent $\omega_2$ ends up with treatment slot $s_4$ leading to a reemployment probability of 18%, while agent $\omega_4$ ends up in $s_2$ which yields a reemployment probability of 25%. Now, Figure 3.2b depicts the allocation resulting from an optimal assignment, while Figure 3.2c shows the allocation from an optimal reallocation. The assignment from each approach is indicated by the entries shaded in dark gray, while the light gray refers

| | $s_1$ | $s_2$ | $s_3$ | $s_4$ |
|---|---|---|---|---|
| $\omega_1$ | 0.25 | 0.20 | 0.12 | 0.27 |
| $\omega_2$ | 0.10 | 0.45 | 0.25 | 0.18 |
| $\omega_3$ | 0.05 | 0.15 | 0.33 | 0.26 |
| $\omega_4$ | 0.23 | 0.25 | 0.45 | 0.14 |

**(a)** Initial Assignment

| | $s_1$ | $s_2$ | $s_3$ | $s_4$ |
|---|---|---|---|---|
| $\omega_1$ | 0.25 | 0.20 | 0.12 | 0.27 |
| $\omega_2$ | 0.10 | 0.45 | 0.25 | 0.18 |
| $\omega_3$ | 0.05 | 0.15 | 0.33 | 0.26 |
| $\omega_4$ | 0.23 | 0.25 | 0.45 | 0.14 |

**(b)** Optimal Assignment

| | $s_1$ | $s_2$ | $s_3$ | $s_4$ |
|---|---|---|---|---|
| $\omega_1$ | 0.25 | 0.20 | 0.12 | 0.27 |
| $\omega_2$ | 0.10 | 0.45 | 0.25 | 0.18 |
| $\omega_3$ | 0.05 | 0.15 | 0.33 | 0.26 |
| $\omega_4$ | 0.23 | 0.25 | 0.45 | 0.14 |

**(c)** Optimal Reallocation

**Figure 3.2:** Contrasting Assignment Mechanisms

to the initial assignment.

In the literature on optimal treatment assignment, each agent is assigned to their first best program. However, as agents $\omega_3$ and $\omega_4$ are both assigned to $s_3$, this violates the resource constraints. In our work on optimal treatment reallocations, we honor the resource constraints and maximize equation (3.2.1) given the available resources. This has a considerable effect on the resulting allocation. Only agents $\omega_2$ and $\omega_4$ are assigned their first best. For the other two agents this is not possible as their favored slots are already filled. Instead of $s_3$, agent $\omega_1$ is assigned to $s_1$, and $\omega_3$ ends up in $s_4$ instead of $s_3$. Turning to a social planner's perspective, the initial assignment produced an average reemployment probability of 15%, while a first best rule results in an increase to 37%. Imposing resource constraints moderates the effectiveness and yields an average reemployment probability of 35%.

## 3.3 Related Literature

Our focus is on reallocation policies in the context of active labor market programs. That is why we restrict our literature review to the related work on optimal treatment assignment, the implementation attempts of targeting and profiling systems, and the limited work on optimal reallocations. We briefly reference work on linear sum assignment problems as well.

Manski (2004), Dehejia (2005), Hirano and Porter (2009), and Frölich (2008) comprise the group of papers that discuss optimal treatment assignment with a methodological focus. Manski (2004) considers a utilitarian social planner who must choose among a set of feasible statistical treatment rules using the minimax-regret criterion. Dehejia (2005) uses data from a randomized evaluation of the Greater Avenues for Independence (GAIN) program to study the effect of exploiting treatment effect heterogeneity on program performance while considering statistical uncertainty about potential outcomes. Focusing on asymptotic optimality theory for statistical treatment rules, Hirano and Porter (2009) derive treatment assignment rules that are asymptotically optimal under different loss functions. Frölich (2008) develops a method that allows the combination of two datasets to build a statistical treatment model when the information available about previous clients is partly unavailable for current assignment decisions due to time delays in data availability. Moreover, he provides results on statistical inference about a recommended treatment choice.

The empirical literature on optimal assignment predicts the potential outcomes for the studied subpopulation and hypothetically reassigns people to the program where they have the best chances of success. All focus is on the efficacy of the allocation, and there is no role for resource constraints. When large improvements in the average treatment effects are documented, this is interpreted as an argument in support of the implementation of a statistical treatment model. For Sweden, Frölich (2008) finds large gains relative to caseworker assignment in the context of a rehabilitation program. Turning to Germany, the findings from Caliendo et al. (2008) are more modest - they only find minor efficiency gains from using an impact-based statistical treatment rule in the context of job creation schemes. Still for Germany, Biewen et al. (2007) focus on the effect of public sector sponsored training programs. They conclude that better targeting along the lines of gender, region, and unemployment duration would increase the aggregate impacts.

We go beyond the empirical strand of the literature by considering optimal treatment reallocations, i.e. we account for the presence of resource constraints. In doing so, we move closer to the assignment problem faced by a job market agency.

Lechner and Smith (2007) and Staghoej et al. (2010) offer the first cautious attempts in this direction. For Switzerland, Lechner and Smith (2007) account for resource constraints on a local level by introducing "Need Based" and "Effect Based" assignment mechanisms, each motivated by equity and efficiency considerations respectively. They report that the ordering of agents does not matter much for the performance of a program. They compare the assignment by a caseworker to a random assignment and do not find any significant differences in employment rates one year after the start of program participation. For Denmark, Staghoej et al. (2010) argue that the implementation of a statistical treatment rule (compared to the current method of caseworker assignment) would decrease the average duration in unemployment by up to 30%. They impose supply constraints on the national level.

The limitation of the work by Lechner and Smith (2007) and Staghoej et al. (2010) is the lack of a clearly articulated link of their proposed assignment rules to an objective function. In their simulation exercises, they allow for multiple treatment options. Yet, they construct a one-dimensional ordering of agents and assign each their first best treatment choice conditional on availability. They determine the sequencing of agents based on an auxiliary index (Lechner and Smith, 2007) or by random ordering of the sample (Staghoej et al., 2010). In doing so, both avoid the combinatorial problem of tackling a multidimensional assignment problem. We take up this challenge and extend their work in an important way. Nevertheless, we also implement their proposed algorithms for our empirical illustration to maintain a close link to their analysis.

The existing work on optimal treatment assignment has already influenced policy making. A variety of countries have implemented profiling and targeting systems (at least for testing purposes).

Profiling systems compute a score for the risk of becoming long-term unemployed. People with a score above a certain threshold are eligible for intensive assistance programs. Profiling systems have been implemented on a nationwide basis and have been evaluated in the United States and Australia. In the United States, a prominent implementation is the Worker Profiling and Reemployment Services (WPRS) system (Wandner, 1997). Its goal is to identify unemployment insurance (UI) claimants likely to exhaust their regular UI entitlement and to offer them job search assistance during the early weeks of unemployment. The evaluation found that the statistical model successfully identifies the claimants most in need and that the attached services reduced most UI outcomes (Dickinson et al., 1997). Building on this work, O'Leary et al. (2005) claim that a targeted reemployment bonus would yield net benefits to the unemployment insurance trust fund if targeted at those with the longest predicted unemployment spells. In Germany and Canada systems are under development but have not been implemented yet.[2]

Targeting systems rely on estimates of potential outcomes for all agents and all programs. The unemployed are assigned to the program in which they are most likely to succeed. For the case of Switzerland, Behncke et al. (2009) report on a randomized experiment of a targeting system, the Statistically Assisted Program Selection (SAPS) system, aimed at assisting caseworkers in choosing an active labor market program for the unemployed. However, they found that caseworkers mostly ignored the offered advice and no further development or implementation efforts are undertaken at this point.

We are interested in maximizing the impact of a given amount of resources. Thus, our work is related to the developing literature that tries to add statistical content to output maximizing reallocation policies.

Graham et al. (2007, 2009) introduce reallocation problems in the econometrics lit-

[2]Eberts et al. (2002) and OECD (1998, 2002) provide an overview on the different implementations in a variety of countries. See Stephan et al. (2006) for the Treatment Effect and Prediction (TrEffeR) project in Germany.

erature and define estimands that capture their effect. Graham et al. (2009) consider a class of reallocations that includes the status quo allocation, a random allocation, and perfect positive (or negative) assortative matching. They do not search for an optimal assignment rule but compare specific rules. Nevertheless, they provide asymptotic results for statistical inference about the effect of the considered reallocations. Bhattachary and Dupas (2012) deal with a social planner's maximization of mean welfare when a binary treatment can be allocated but is in limited supply. For this case, they provide consistency and distributional results. In a change of perspective, they also consider the dual value, i.e. the minimum resources needed to attain a specific level of welfare via an efficient treatment assignment protocol. Linking their work to the literature on model selection, they address the issue of covariate choice for a targeted treatment assignment and show its dependence on the available (finite) sample. Bhattachary (2009) studies the effect of optimally assigning individuals to peer groups to maximize social gains from heterogeneous peer effects. Graham (2011), with a focus on the presence of situations with social spillovers, provides a recent overview on econometric methods for the analysis of such reallocation policies.

As illustrated in Section 3.2, the determination of an optimal assignment in the presence of resource constraints and multiple treatments is a nontrivial task due to the multidimensionality of the assignment problem. Given the formulation of the reallocation problem as a linear sum assignment problem (LSAP), we draw on operations research for suitable solution methods.[3] Burkard and Cela (2011) provide a survey on the state of the art of the solution methods. Numerous applications, ranging from the personnel assignment (Ewashko and Dudding, 1971; Meggido and Tamir, 1971), production planning (Veinott and Wagner, 1962; Zangwill, 1969), or tracking moving objects in space (Brogan, 1989), were previously formulated and solved as a LSAP.[4]

Despite the existing literature our contribution is clear. We determine an optimal assignment of agents to multiple ALMP measures in the presence of resource con-

[3]See Burkard et al. (2009) for an introduction to this line of research and comprehensive overview of the available solution methods.
[4]Ahuja et al. (1995) provide an overview on numerous applications.

straints. In contrast to earlier work, our proposed assignment mechanism is tightly linked to the adopted objective function that maximizes the efficiency of the allocated resources.

## 3.4 Empirical Illustration

We now present our empirical illustration. We apply the developed conceptual framework to the case of an active labor market program. We estimate the conventional average treatment effects, but also document considerable treatment effect heterogeneity. We exploit this variation and investigate the effect of optimal reallocations on program performance under alternative information sets. We link our work to the existing literature by contrasting the previously proposed assignment mechanisms to an optimal reallocation. We ensure full transparency of our implementation by publishing and documenting our computational routines online.[5]

We draw our data from the National Evaluation of Welfare-to-Work Strategies (NEWWS). NEWWS was a U.S. evaluation, conducted by the MDRC, of different strategies for welfare-to-work job training programs.[6] It was designed to evaluate the performance of competing strategies. We briefly summarize these below:

- **Standard Services (SS):** standard local services outside of the NEWWS programs.
- **Human Capital Development (HCD):** focus on basic education and vocational skills training.
- **Labor Force Attachment (LFA):** focus on job search activities to get participants employed as quickly as possible, even at reduced wages.

The evaluation was based on a randomized controlled trial (RCT) conducted at three sites which had specifically implemented both the HCD and the LFA approach for the participants. Each site conducted a three-way randomization between the two competing strategies and a control group, which received the standard local services. The benefits of the program generated by randomized assignment will serve

[5] `http://www.policy-lab.org/software.html`

[6] Additional information is provided in Freedman et al. (2000) and Hamilton et al. (2001).

as a benchmark for the results that can be achieved using alternative assignment rules.

We use data on a total of 5,768 agents from the Atlanta site. Earnings data is available for each agent for two years (by quarter) prior to treatment, and for five years (by quarter) post treatment.

We postulate the policy goal of maximizing total earnings over a period of five years. Table 3.1 shows the raw difference in average outcomes for each treatment alternative.

**Table 3.1:** Total Earnings Within Five Years

| Policy | Total Earnings | vs. SS |
|---|---|---|
| SS | 19,722 | — |
| HCD | 20,979 | 1,257 |
| LFA | 21,778 | 2,056 |

Agents assigned to the SS earned a total of \$19,722 on average over a period of 5 years, which is less than in either of the alternative states. Under random assignment, the performance of the LFA approach is superior to the HCD alternative. Compared with SS, the improvement in total earnings is largest for the LFA approach with \$2,056. This is roughly 65% higher than the benefits generated by the HCD program, where total earnings improved by \$1,257 on average.

A focus on average differences only, neglects possible treatment effect heterogeneity based on observable and unobservable agent characteristics. We model this heterogeneity by fitting a linear-in-parameters equation for the earnings dynamics,

$$Y_t = X\beta_t + U_t$$

, where $Y_t$ denotes realized earnings in period $t$. From the perspective of the econometrician, $X$ are the observed and $U_t$ the unobserved determinants of earnings.

Following Cunha et al. (2005), we decompose $U_t$ into a permanent $\theta$ and a transitory $\epsilon_t$ component using a factor structure assumption. More formally,

$$U_t = \alpha_t \theta + \epsilon_t$$

, where $\theta$ denotes the factor common to the unobservable earnings components over time. The factor loading $\alpha_t$ may vary, and thus $\theta$ may affect earnings differently in each period. $(\theta, \epsilon_t)$ are both unobservable components of the variation in $Y_t$. We refer to $\theta$ as the "Permanent Earnings" factor in what follows.

Nonparametric identification of the joint dependencies among the unobservables of the model $U_t$ can be achieved under the conditions outlined in Cunha et al. (2010). Each observation on $Y_t$ provides a signal on $\theta$ but also contains noise due to the disturbances $\epsilon_t$. Nonetheless, orthogonality conditions allow to separate the two and to identify the distribution of $\theta$.

We break the agents' labor market experiences into four periods. Prior to treatment, we pool earnings into three periods with two quarters each. Post treatment, we collapse all five years into one period.

Prior treatment, zero earnings are observed for roughly 70% of the sample within each period. To account for this fact, we specify a normal Tobit model for each period $t = 1, .., 3$.

$$Y_t = \begin{cases} Y_t & \text{if} \quad X\beta_t + \alpha_t \theta + \epsilon_t > 0 \quad \text{where} \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2_{\epsilon_t}) \\ 0 & \text{otherwise} \end{cases}$$

The unobservables $(\theta, \{\epsilon_t\})$ are independent of the observables $X$. The idiosyncratic components $\{\epsilon_t\}$ are independent across time and independent of the factor $\theta$.

Post treatment, we specify a normal linear regression model by treatment status $d = s, h, l$. The linear regression specification fits the data better than the Tobit alternative as only 15% of the sample experience zero earnings over the five years post treatment.

(3.4.1)
$$Y_d = X\beta_d + \alpha_d\theta + \epsilon_d \qquad \text{with} \quad \epsilon_d \sim \mathrm{N}(0, \sigma^2_{\epsilon_d})$$

The unobservables $(\theta, \{\epsilon_d\})$ are independent of the observables $X$. The idiosyncratic components $\{\epsilon_d\}$ are independent across treatment states and independent of $\theta$.

Prior and post treatment earnings are determined by dummy variables indicating martial status, the presence of more than two children in the household, the presence of a small child, race, gender, and educational attainment. In addition, earnings are determined by linear terms in age at random assignment and the "Permanent Earnings" factor. Additional descriptive statistics and more details about the dataset are provided in Appendix C.

We approximate the distribution of $\theta$ 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 $\theta$ is distributed as a univariate mixture of $K$ normals with share parameter $\pi_k$, mean $\mu_k$, and variance $\sigma^2_k$,

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

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

We collect all parameters of the model in $\Psi$. Conditional on $\theta$ and the relevant observables, the observed outcomes are all independent. Thus, the individual likelihood can be written as

$$\mathcal{L}(\Psi) = \int_{\Theta} \prod_{t=1}^{3} f(Y_t \mid X, \theta; \Psi) \prod_{d=1}^{3} f(Y_d \mid X, \theta; \Psi)^{\mathrm{I}[D=d]} dF_\theta(\theta)$$

, where $f(\cdot)$ denotes a density function, and $F_\theta(\cdot)$ is the cumulative distribution function of $\theta$ over its support $\Theta$.

$\theta$ needs to be integrated out of the individual likelihood, which leads to a complex nature of the likelihood function. For this reason, we implement a full Bayesian approach for the estimation of the model and rely on Markov Chain Monte Carlo (MCMC) techniques.[7] 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). 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 $M = 10,000$ iterations are used for postestimation inference.

We start our analysis with the conventional average treatment effects and the underlying treatment effect heterogeneity.

We sample from the posterior distribution of the conventional average treatment effects as follows. First, we simulate a sample of $N = 3,000$ agents for each of the $m = 1, \ldots, M$ remaining iterations. We draw a set of observable characteristics $X$ with replacement from the original dataset and simulate the unobservables $(\theta, \{\epsilon_d\})$ from their distributions which are parametrized by $(\{\pi_k^{(m)}\}, \{\mu_k^{(m)}\}, \{\sigma_k^{2(m)}\})$ and $\{\sigma_{\epsilon_d}^{(m)}\}$ respectively. Second, we construct potential outcomes $\{Y_d^{(m)}\}$ based on $(\{\beta_d^{(m)}\}, \{\alpha_d^{(m)}\})$ according to equation (3.4.1). Finally, we calculate the $ATE_d^{(m)}$ for each iteration and each treatment alternative,

$$ATE_d^{(m)} = \frac{1}{N} \sum_{n=1}^{N} (Y_{d,n}^{(m)} - Y_{s,n}^{(m)}) \qquad \text{for} \quad d = h, l.$$

Table 3.2 reports our results.

[7] 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.

**Table 3.2:** Average Effect of Treatment

| Overall | | |
|---|---|---|
| Treatments | Estimate | 95% Confi. |
| HCD | 1,341** | -155 / 2,847 |
| LFA | 1,972*** | 440 / 3,491 |

**Notes**: based on 10,000 simulations of the model; Estimate = mean of posterior distribution; Confi. = confidence interval of the posterior distribution; Level of Significance = ***/**/* if zero lies outside the 99%/95%/90% confidence interval of the posterior distribution.

By virtue of the experimental research design, the estimates are nearly identical to the reported average differences in Table 3.1. On average and under random assignment, the benefits generated by the LFA program are higher than by the HCD alternative.

The average benefits of treatment mask considerable treatment effect heterogeneity due to observable and unobservable agent characteristics. This heterogeneity can be exploited by a targeted treatment assignment. However, for our contribution to the existing literature to be meaningful, this is not enough. A truly multidimensional assignment problem requires variation in the benefits from the alternative treatments within agents as well.

In the NEWWS dataset, this is in fact the case - the impact of the HCD and LFA program varies between and within agents. Table 3.3 provides a detailed account on the sources of the underlying differential variation. There, we test whether the determinants of an agent's post treatment earnings have a differential marginal effect (ME) on the benefits generated by the HCD or LFA approach.[8]

[8]We rescaled the earnings variable to $10,000 units for the purpose of our estimation.

**Table 3.3:** Treatment Effect Heterogeneity

| Overall | | |
|---|---|---|
| | HCD vs. LFA | |
| Covariate | Estimate | 95% Confi. |
| Intercept | -0.030 | -0.168 / 0.110 |
| Single Parent | -0.015 | -0.047 / 0.018 |
| Two Children | 0.008 | -0.024 / 0.039 |
| Small Child | -0.010 | -0.043 / 0.023 |
| Black | -0.003 | -0.069 / 0.064 |
| Age | 0.001 | -0.001 / 0.004 |
| Female | 0.001 | -0.079 / 0.084 |
| Education | 0.019 | -0.012 / 0.050 |
| Permanent Earnings | -0.047** | -0.094 / 0.000 |
| Variance | 0.008*** | 0.003 / 0.013 |

**Notes**: based on 10,000 simulations of the model; Estimate = mean of posterior distribution; Confi. = confidence interval of the posterior distribution; Level of Significance = ***/**/* if zero lies outside the 99%/95%/90% confidence interval of the posterior distribution.

For each of the $M$ iterations and each covariate $c$, we calculate the following test statistic $\delta_c^{(m)}$,

$$\delta_c^{(m)} = \underbrace{(\beta_{h,c}^{(m)} - \beta_{s,c}^{(m)})}_{\text{ME on } \Delta_{h,s}} - \underbrace{(\beta_{l,c}^{(m)} - \beta_{s,c}^{(m)})}_{\text{ME on } \Delta_{l,s}}.$$

Positive values of $\delta_c^{(m)}$ are associated with a higher marginal effect of covariate $c$ on the benefits in the HCD program than in the case of an LFA assignment. The opposite is true for negative values. Only "Permanent Earnings" has a significant negative differential effect. Thus, all else equal, as "Permanent Earnings" increases, so does the relative benefit of an LFA assignment compared to the HCD approach. A potential reason is the locking-in effect of the HCD program, which lasts considerably longer than the LFA program. Participants in either program might reduce their search intensity while participating (e.g. van Ours (2004)). However, agents with high "Permanent Earnings" tend to have a higher attachment to the labor market. So, at least over a period of five years, the relative benefit of an agent from participating in the HCD program declines with "Permanent Earnings" due to the prolonged absence from the labor market.

Now, we start to investigate the effects of alternative assignment mechanisms on program performance. Throughout, we take an ex post perspective. We study what benefits could have been realized for alternative assignment rules.

We simulate a random sample of 5,768 agents. First, we fix all estimated parameters to their posterior means. Second, we draw a set of observable characteristics $X$ with replacement from the original dataset. Third, we simulate the unobservables of the model $(\theta, \{\epsilon_t\}, \{\epsilon_d\})$. Together, this allows to construct observed and potential outcomes $(\{Y_t\}, \{Y_d\})$ and calculate individual benefits $\{\Delta_{d,s}\}$. Fourth, we allocate the simulated agents to the treatment alternatives by applying a variety of assignment mechanisms.

Most assignment mechanisms exploit information about predicted potential outcomes $\{\hat{Y}_d\}$. We calculate them by fixing the parameters $(\{\beta_d\}, \{\alpha_d\})$ at their posterior means and plugging in the agents' characteristics $(X, \hat{\theta})$,

(3.4.2)
$$\hat{Y}_d = X\beta_d + \alpha_d\hat{\theta} \qquad \text{for} \quad d = s, h, l.$$

As the "Permanent Earnings" factor is not directly observable, we predict its conditional mean $\hat{\theta}$ for each individual using only the earnings observations prior to treatment $\{Y_t\}$. In doing so, we make sure that the information used to predict potential outcomes resembles that of a social planner at the time of the allocation decision. In principle, there are two sources for differences between predicted and realized outcomes. They arise due to the random earnings components $\{\epsilon_d\}$ and prediction error in $\theta$. In the first part of this section, we abstract from uncertainty due to $\{\epsilon_d\}$ and set them to zero. This allows to study the properties of the assignment algorithms free of randomness due to sample size. Later, we study the impact of uncertainty in small samples for the comparative performance of alternative assignments rules in detail.

When we impose resource constraints, we designate 1,935 slots for the HCD program and 1,887 slots for the LFA program. This agrees with the original dataset.

Table 3.4 presents our main results. For each alternative assignment algorithm, we report the average observed outcomes, the average benefits among those assigned to each treatment, the slots associated with each alternative, as well as the average overall benefits.

For the "Optimal" assignment, we draw on the Hungarian algorithm (Kuhn, 1955) to solve the maximization problem of equation 3.2.1, i.e. to maximize predicted overall benefits for a limited amount of resources.

In addition, we present the results from a "Random" and a "First Best" assignment. These are commonly applied in the existing literature on optimal treatment assignment. For the "Random" assignment, we impose resource constraints. We assign each of the 5,768 generic agents a treatment state at random (among those available) with equal probability. For the "First Best" assignment, each agent is assigned to the treatment with the highest predicted potential outcome disregarding resource constraints. We break potential ties at random.

A "First Best" assignment results in average outcomes of $22,116, which is the highest among the three alternatives. However, the scale of the program increases from a total of 3,822 to 5,376 slots. Thus, only 382 agents have negative predicted benefits from either program and remain in the SS. Imposing resource constraints moderates average outcomes to $20,897 for the "Random" assignment and to $21,870 for the "Optimal" assignment. Acknowledging resource constraints but exploiting treatment effect heterogeneity, i.e. moving from a "Random" to an "Optimal" assignment, allows to roughly double average overall benefits from $1,595 to $3,062. Interestingly, the HCD program outperforms the LFA approach under both algorithms that attempt to exploit treatment effect heterogeneity. Only under "Random" assignment average benefits among the LFA participants are larger than for the HCD program. Note also that the average benefits among those assigned to the LFA are much smaller under the "First Best" than under the "Optimal" assignment. The marginal agents added to the LFA program experience only small benefits as

**Table 3.4:** Main Algorithms: Observables and Predicted Factor

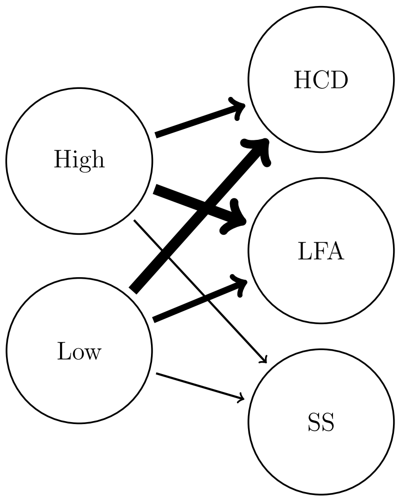
| | Overall | HCD | | LFA | | Overall | |
|---|---|---|---|---|---|---|---|
| Algorithms | Outcomes | Benefits | Slots | Benefits | Slots | Benefits | Slots |
| Optimal | 21,870 | 3,228 | 1,935 | 2,892 | 1,887 | 3,062 | 3,822 |
| Random | 20,897 | 1,229 | 1,935 | 1,970 | 1,887 | 1,595 | 3,822 |
| First Best | 22,116 | 2,948 | 1,887 | 2,167 | 3,489 | 2,441 | 5,376 |

**Notes**: based on 5,768 simulated individuals from the estimates of the model; outcomes and benefits in $1.

the scale of the program increases from 1,887 to 3,489 slots under the “First Best” regime.

In Figure 3.3, we split the simulated sample by the level of the predicted “Permanent Earnings” component $\hat{\theta}$. We show the “Optimal” assignment to either of the three alternative states. An arrow represents the assignment, with its thickness indicating the relative proportions of each group that end up in either of the possible states. As suggested by Table 3.3, those with a high level of “Permanent Earnings” largely end up in the LFA program as they are likely to profit the most from it. However, due to resource constraints and other factors determining an agent’s relative benefit of treatment, some do end up in the HCD and SS programs as well. Those with a low level of “Permanent Earnings” mainly end up in the HCD approach.

**Figure 3.3:** “Permanent Earnings” Assignment




The results based on “Random” assignment make clear the limited policy relevance of the conventional treatment effects. They differ, by construction, from the $ATE$ reported in Table 3.2 only due to finite sample uncertainty. Thus, the $ATE$ is only of interest to policy makers if the program is applied in the future with random assignment as the allocation mechanism as well.

We briefly summarize our main results. The assignment mechanism has a profound impact on the scale of the program, as well as its overall performance and the relative efficiency of treatment alternatives. Thus, the choice of the assignment mechanism is an important component of an optimal policy design.

Our main contribution is the determination of the optimal assignment, as clearly defined by equation (3.2.1), of agents to alternative treatments in the presence of resource constraints. However, as noted in the literature review, Lechner and Smith (2007) and Staghoej et al. (2010) tentatively address this issue. Yet, they do not provide a clear link to any well-specified criterion function. This allows them to avoid the multidimensional assignment problem. We now discuss their approaches in detail and compare them to our "Optimal" algorithm.

Lechner and Smith (2007) propose a "Need Based" and an "Effect Based" assignment rule. For the "Need Based" assignment, agents are ordered based on an indicator of need. Those in highest need, i.e. those with the lowest predicted outcome in the absence of any treatment, are assigned their preferred treatment among those available. They introduce this assignment rule to reflect possible equity concerns. The "Effect Based" assignment is motivated by efficiency considerations instead. Agents are ordered by the difference between the most positive and second most positive benefit from treatment. Those with the highest difference are allocated first. Once a treatment alternative is filled, the ranking is updated as if that treatment was not available. The intuition is the following: those who benefit the most from their first best treatment choice are assigned first. Finally, for the "First Come" procedure implemented in Staghoej et al. (2010), agents are ordered at random and assigned their first best among the programs still available.

We apply their proposed assignment mechanisms to our simulated sample. Table 3.5 presents the results.

Looking at the "Effect Based" assignment, overall average benefits are reduced to \$2,796 compared to \$3,062 based on the "Optimal" assignment. In addition, the average benefits within each treatment state differ. Following an "Effect Based" assignment, the benefits among LFA participants are smaller compared to the case of an "Optimal" assignment. The opposite is true for the HCD program. A "First Come" assignment performs worst among the rules that try to exploit treatment ef-

**Table 3.5:** Constrained Algorithms: Observables and Predicted Factor

| | Overall | HCD | | LFA | | Overall | |
|---|---|---|---|---|---|---|---|
| Algorithms | Outcomes | Benefits | Slots | Benefits | Slots | Benefits | Slots |
| Optimal | 21,870 | 3,228 | 1,935 | 2,892 | 1,887 | 3,062 | 3,822 |
| Need Based | 21,620 | 2,996 | 1,935 | 2,367 | 1,887 | 2,686 | 3,822 |
| Effect Based | 21,693 | 3,472 | 1,935 | 2,102 | 1,887 | 2,796 | 3,822 |
| First Come | 21,442 | 2,625 | 1,935 | 2,202 | 1,887 | 2,416 | 3,822 |

**Notes**: based on 5,768 simulated individuals from the estimates of the model; outcomes and benefits in $1.

fect heterogeneity. Overall average benefits are roughly 20% lower when set against the "Optimal" assignment. In short, an attempt to avoid a multidimensional assignment problem leads to worse performances among those algorithms, which have efficiency as their declared goal.

A conflict between efficiency and equity arises if those in greatest need do not have the most to gain from treatment participation. The overall average benefits are smaller when following the "Need Based" assignment ($2,686) compared to an "Optimal" assignment ($3,062). In contrast to Lechner and Smith (2007), there is a trade-off in our sample.

Next, we turn to the crucial role of information that is available to the policy maker at the time of the treatment assignment. We start with the additional value of information provided by the factor structure assumption. Afterwards, we study the relative performance of alternative assignment rules in small samples.

Again, we apply the main assignment algorithms to our simulated sample. This time, however, we only use the observable characteristics $X$ for the prediction of potential outcomes $\{\hat{Y}_d\}$ and neglect the information about $\theta$. Thus,

(3.4.3)
$$\hat{Y}_d = X\beta_d \qquad \text{for} \quad d = s, h, l.$$

Table 3.6 presents the results. For comparison, we repeat our previous results in the bottom half of the table.

The "Optimal" assignment algorithm exploits the additional information most effectively. Adding "Permanent Earnings" to the prediction step allows for an increase in overall average benefits by $375 from $2,687 to $3,062. For the "First Best" assignment the increase is more moderate with $194. The performance of both algorithms improves when using more information. In this case, the predicted benefits provide a more accurate picture of the realized benefits. Turning to the case of "Random" assignment, the results are very similar by construction as no information is exploited anyway. Interestingly, in the case of an "Optimal" assignment and when only the observables are used in the prediction step, then the LFA program outperforms the

**Table 3.6:** Main Algorithms: Value of Information

| | Observables Only | | | | | | |
|---|---|---|---|---|---|---|---|
| | Overall | HCD | | LFA | | Overall | |
| Algorithms | Outcomes | Benefits | Slots | Benefits | Slots | Benefits | Slots |
| Optimal | 21,621 | 2,364 | 1,935 | 3,018 | 1,887 | 2,687 | 3,822 |
| Random | 20,943 | 1,380 | 1,935 | 1,953 | 1,887 | 1,663 | 3,822 |
| First Best | 21,921 | 2,295 | 1,315 | 2,231 | 4,023 | 2,247 | 5,338 |

| | Observables and Predicted Factor | | | | | | |
|---|---|---|---|---|---|---|---|
| | Overall | HCD | | LFA | | Overall | |
| Algorithms | Outcomes | Benefits | Slots | Benefits | Slots | Benefits | Slots |
| Optimal | 21,870 | 3,228 | 1,935 | 2,892 | 1,887 | 3,062 | 3,822 |
| Random | 20,897 | 1,229 | 1,935 | 1,970 | 1,887 | 1,595 | 3,822 |
| First Best | 22,116 | 2,948 | 1,887 | 2,167 | 3,489 | 2,441 | 5,376 |

**Notes**: based on 5,768 simulated individuals from the estimates of the model; outcomes and benefits in $1.

HCD approach. Just as under "Random" assignment. However, adding the information about "Permanent Earnings" turns this ranking around. So, the available information about treatment effect heterogeneities not only determines the magnitude of the improvements that can be achieved by targeted assignment rules. It also affects the relative performance of alternative programs for a given assignment rule.

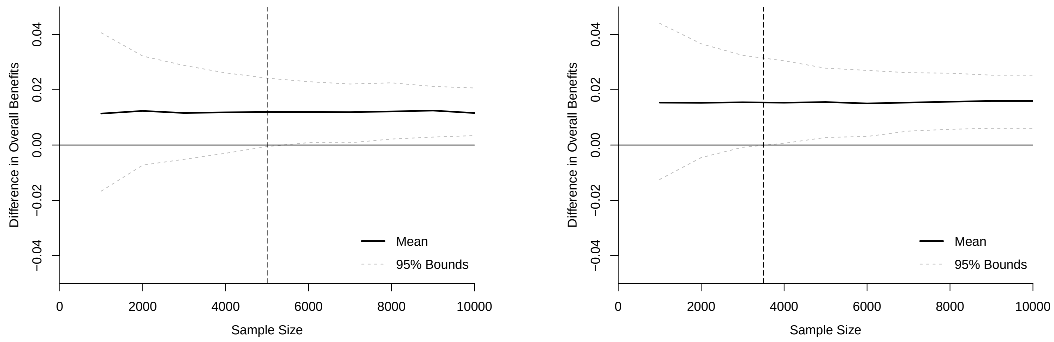
At last, we investigate the behavior of the assignment mechanisms in small samples. More precisely, we study what sample size is required for the "Optimal" assignment to reliably outperform the "Random" allocation. After the prediction of potential outcomes, there is still considerable uncertainty about the actual benefits due to the unknown realization of $(\theta, \{\epsilon_d\})$. So, at least in small samples, the realized overall average benefits resulting from an "Optimal" assignment might turn out to be lower than in the case of a "Random" allocation.

We simulate samples of varying size with replacement from the original dataset. We impose resource constraints by keeping the relative share of the HCD and LFA slots in the same proportion as in the original dataset. For each of the samples, we predict potential outcomes using equation (3.4.2) and (3.4.3) and determine the assignment schedule for a "Random" and "Optimal" rule. In the next step, we draw $\{\epsilon_d\}$ for all agents in the random samples and construct the difference in average overall benefits between the two algorithms. We repeat the last step 1,000 times.

Figure 3.4 reports the results for the two alternative information sets. The thick line marks the average difference between the two assignments, whereas the dashed line indicates the 0.025 and 0.975 quantiles. The vertical line emphasizes the sample size which is required so that the "Optimal" assignment outperforms the "Random" alternative at least 95% of the time.

On average, the "Optimal" assignment outperforms the "Random" allocation for each sample size. This is true regardless of the information set used. The average performance difference is unaffected by increases in sample size, but the variation in observed differences declines. However, all this is true by construction of the model. The prediction error washes out due to the independence and mean zero assumptions. What is interesting, however, is that it takes a sample of roughly

**Figure 3.4:** Algorithm Performance and Uncertainty




**(a)** Observables Only **(b)** Observables and Predicted Factor

**Notes:** benefits in $100,000.

3,500 agents for the “Optimal” assignment to reliably outperform the “Random” assignment when predicted “Permanent Earnings” are included. If only observable characteristics are used, then this number increases to 5,000 agents. The increase in the required sample size results from the additional source of uncertainty due to the omission of the “Permanent Earnings” factor.

The last set of results both emphasize the key role of the information available to the policy maker at the time of the assignment decision. The available information determines not only whether there is the possibility to exploit any treatment effect heterogeneity at all, but also the required sample size that allows for any meaningful improvements in program performance by an optimal assignment rule.

## 3.5 Conclusion

We presented a framework that links reallocation policies to the econometrics of policy evaluation and argued that this connection is particularly policy relevant for active labor market programs. We extended the existing literature on optimal treatment assignment by acknowledging the presence of resource constraints and linking the assignment mechanism to a criterion of optimality.

We illustrated the relevance of our contribution by an application to the National Evaluation of Welfare-to-Work Strategies (NEWWS) dataset. The assignment mechanism matters for the scope and the performance of the program. Overall average benefits could have been doubled by an optimal reallocation - all this without an augmentation of resources. However, the available information mattered for the magnitude of possible improvements.

There are several limitations to our current approach. First, our analysis is limited by the ex post focus on fixed program capacities. Yet, the design of an optimal policy demands an ex ante perspective where budgetary constraints are more relevant. Second, policy relevance of our research requires that the statistical treatment model used to predict the potential outcomes remains valid over time and when applied to a new population.

We intend to address both these issues using administrative data from the German Federal Employment Agency. Germany has a long tradition of extensive active labor market programs covering a wide range of approaches (job creation in the public sector, public training programs, etc.). Thus, it offers a promising opportunity to investigate the economic relevance of the issues raised in this chapter on a large scale.

We interpret our contribution as a starting point that motivates a more elaborate investigation into the assignment problem faced by a job market agency. Additional information about our research agenda is provided online.[9]

[9] `http://www.policy-lab.org/researchprojects.html`

# Bibliography


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.

Ahuja, R. K., Magnanti, T. L., Orlin, J. B., and Reddy, M. R. (1995). Applications of Network Optimization. In Ball, M. O., Magnanti, C. L., Monma, C. L., and Neuhauser, G. L., editors, *Network Models*, volume 7, pages 1–83. North-Holland Publishing Company, 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.

Bagnoli, M. and Bergstrom, T. (2005). Log-Concave Probability and Its Applications. *Econometric Theory*, 26(2):445–469.

Behncke, S., Frölich, M., and Lechner, M. (2009). Targeting Labour Market Programmes - Results from a Randomized Experiment. *Swiss Journal of Economics and Statistics*, 145(3):221–268.

Bhattachary, D. (2009). Inferring Optimal Peer Assignment From Experimental Data. *Journal of the American Statistical Association*, 104(486):486–500.

Bhattachary, D. and Dupas, P. (2012). Inferring Welfare-Maximizing Treatment Assignment under Budget Constraints. *Journal of Econometrics*, 167(1):168–196.

Biewen, M., Fitzenberger, B., Osikominu, A., and Waller, M. (2007). Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany. *IZA Discussion Paper*, No. 2885.

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.

Blundell, R. and Costas Dias, M. (2009). Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources*, 44(3):565–640.

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.

Brogan, W. L. (1989). Algorithm for Ranked Assignments with Applications to Multiobject Tracking. *Journal of Guidance, Control, and Dynamics*, 12(3):357–364.

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.

Burkard, R. and Cela, E. (2011). Linear Assignment Problems and Extensions. *Unpublished Manuscript*.

Burkard, R., Dell'Amico, M., and Martello, S. (2009). *Assignment Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Caliendo, M., Hujer, R., and Thomsen, S. L. (2008). Identifying Effect Heterogeneity to Improve the Efficiency of Job Creation Schemes in Germany. *Applied Economics*, 40(9):1101–1122.

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.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics*, 1:451–488.

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.

Dehejia, R. H. (2005). Program Evaluation as a Decision Problem. *Journal of Econometrics*, 125(1-2):141–173.

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.

Dickinson, K. P., Suzanne, K. D., and Decker, P. T. (1997). *Evaluation of Worker Profiling and Reemployment Services Systems.* U.S. Department of Labor, 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.

Eberts, R. W., O'Leary, C. J., and Wandner, S. A., editors (2002). *Targeting Employment Services.* W. E. Upjohn Institute for Employment Research, Kalamazoo, MI.

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

Ewashko, T. A. and Dudding, R. C. (1971). Application of Kuhn's Hungarian Assignment Algorithm to Posting Servicemen. *Operations Research*, 19(4):991.

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.

Flinn, C. and Heckman, J. J. (1982). New Methods for Analyzing Structural Models of Labor Force Dynamics. *Journal of Econometrics*, 18(1):115–168.

Florens, J.-P., Meghir, C., Heckman, J. J., and Vytacil, E. J. (2008). Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects. *Econometrica*, 76(5):1191–1206.

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.

Freedman, S., Friedlander, D., Hamilton, G., Rock, J., Mitchell, M., Nudelman, J., Schweder, A., and Storto, L. (2000). *Evaluating Alternative Welfare-to-Work Approaches: Two-Year Impacts for Eleven Programs.* Manpower Demonstration Research Corporation, New York, NY.

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.

Frölich, M. (2008). Statistical Treatment Choice: An Application to Active Labor Market Programs. *Journal of the American Statistical Association*, 103(482):547–558.

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

Graham, B. S. (2011). Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers. In Benhabib, J., Bisin, A., and Jackson, M., editors, *Handbook of Social Economics*, volume 1B, pages 965–1052. North-Holland Publishing Company, Amsterdam, Netherlands.

Graham, B. S., Imbens, G. W., and Ridder, G. (2007). Redistributive Effects for Discretely-Valued Inputs. *Unpublished Manuscript.*

Graham, B. S., Imbens, G. W., and Ridder, G. (2009). Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis. *Unpublished Manuscript.*

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.

Hamilton, G., Freedman, S., Gennetian, L., Michalopoulos, C., Walter, J., Adams-Ciardullo, D., and Gassman-Pines, A. (2001). *How Effective Are Different*

*Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs*. U.S. Department of Health and Human Services, Washington, DC.

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.

Harberger, A. C. and Jenkins, G. P. (2002). *Cost-Benefit Analysis*. Edward Elgar Publishers, Northampton, MA.

Heckman, J. J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, 42(4):679–694.

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

Heckman, J. J. (1992). Randomization and Social Policy Evaluation. In Manski, C. and Garfinkel, I., editors, *Evaluating Welfare and Training Programs*, pages 201–230. Harvard University Press, Cambridge, MA.

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. (2010). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policies. *Journal of Economic Literature*, 48(2):356–398.

Heckman, J. J. and Honore, B. E. (1990). The Empirical Content of the Roy Model. *Econometrica*, 58(5):1121–1149.

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., LaLonde, R., and Smith, J. (1999a). The Economics and Econometrics of Active Labor Market Programs. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3A, pages 1865–2073. North-Holland Publishing Company, Amsterdam, Netherlands.

Heckman, J. J., Lochner, L., and Taber, C. (1999b). 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. and Sedlacek, G. (1985). Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market. *Journal of Political Economy*, 93(6):1077–1125.

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. and Smith, J. A. (1995). Assessing the Case for Social Experiments. *The Journal of Economic Perspectives*, 9(2):85–110.

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. (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8):4730–4734.

Heckman, J. J. and Vytlacil, E. J. (2001a). Insturmental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect. In Lechner, M. and Pfeiffer, F., editors, *Econometric Evaluations of Active Labor Market Policies in Europe*, pages 1–15. Physica Verlag, Heidelberg, Germany.

Heckman, J. J. and Vytlacil, E. J. (2001b). 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. (2001c). 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.

Hirano, K. and Porter, J. R. (2009). Asymptotics for Statistical Treatment Rules. *Econometrica*, 77(5):1683–1701.

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 Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

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

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, UK.

Karlin, S. (1968). *Total Positivity*. Stanford University Press, Stanford, CA.

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

Kuhn, H. W. (1955). The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.

Lechner, M. and Smith, J. (2007). What is the Value Added by Caseworkers? *Labour Economics*, 14(2):135–151.

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

Manski, C. F. (2004). Statistical Treatment Rules for Heterogeneous Populations. *Econometrica*, 72(4):1221–1246.

Meggido, N. and Tamir, A. (1971). An O(n log n) Algorithm for a Class of Matching Problems. *SIAM Journal of Computing*, 7(2):154–157.

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.

OECD (1998). *Early Identification of Jobseekers at Risk of Long-Term Unemployment: The Role of Profiling*. OECD Publication Service, Paris, France.

OECD (2002). *OECD Employment Outlook*. OECD Publication Service, Paris, France.

O'Leary, C. J., Decker, P. T., and Wandner, S. A. (2005). Cost-Effectiveness of Targeted Reemployment Bonuses. *Journal of Human Resources*, 40(4):270–279.

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.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931–954.

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.

Shaikh, A. and Vytlacil, E. J. (2011). Partial Identification in Triangular Systems of Equations with Binary Dependent Variables. *Econometrica*, 79(3):949–955.

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

Staghoej, J., Svarer, M., and Rosholm, M. (2010). Choosing the Best Training Programme: Is There a Case for Statistical Treatment Rules. *Oxford Bulletin of Economics and Statistics*, 72(2):172–201.

Stephan, G., Rässler, S., and Schewe, T. (2006). Das TrEffeR-Projekt der Bundesagentur für Arbeit: Die Wirkung von Maßnahmen Aktiver Arbeitsmarktpolitik. *Zeitschrift für Arbeitsmarkt- und Berufsforschung*, 39(3-4):447–465.

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.

Tinbergen, J. (1956). *Economic Policy: Principles and Design*. North-Holland Publishing Company, Amsterdam, Netherlands.

van Ours, J. C. (2004). The Locking-In Effect of Subsidized Jobs. *Journal of Comparative Economics*, 32(1):37–52.

Veinott, A. F. and Wagner, H. M. (1962). Optimal Capacity Scheduling - Part I and II. *Operations Research*, 10(4):518–547.

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

Vytlacil, E. J. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, 70(1):331–341.

Wandner, S. A. (1997). Early Reemployment for Dislocated Workers: A New United States Initiative. *International Social Security Review*, 50(4):95–112.

Yildiz, N. and Vytlacil, E. J. (2007). Dummy Endogenous Variables in Weakly Separable Models. *Econometrica*, 75(3):759–779.

Zangwill, W. I. (1969). A Backlogging Model and a Multi-Echelon Model of a Dynamic Economic Lot Size Production System - A Network Approach. *Management Science*, 15(9):506–527.

# A Technical Appendix

## A.1 Proofs of Theorems

### *Theorem 1*

*Proof.* Assertion (1) was proven in the text. For Assertion (2), first consider the cost parameters. $C^{ATE}(z) - C^{TT}(z) = E(U_C) - E(U_C|Z = z, D = 1)$, and $E(U_C|Z = z, D = 1) = \int E(U_C|Z = z, X = x, U_S \leq P(x, z))dF_{X|Z,D}(x|z, 1) = \int E(U_C|U_S \leq P(x, z))dF_{X|Z,D}(x|z, 1)$ using $(X, Z) \perp\!\!\!\perp (U_C, U_S)$. Thus, using that $U_S = F_V(V)$, it will be sufficient to show that $E(U_C|V \leq t) \leq E(U_C)$ for all $t$, and thus sufficient to show that $\Pr[U_C \leq s|U_C - (U_1 - U_0) \leq t] \geq \Pr[U_C \leq s]$ for all $s$. Using Bayes' rule, this is equivalent to $\Pr[U_C - (U_1 - U_0) \leq t|U_C \leq s] \geq \Pr[U_C - (U_1 - U_0) \leq t]$, and this last assertion can now easily be shown using $U_C \perp\!\!\!\perp (U_1 - U_0)$. We can thus conclude that $C^{ATE}(z) - C^{TT}(z) \geq 0$. The same argument *mutatis mutandis* shows that $B^{ATE}(x) - B^{TT}(x) \leq 0$. Now consider assertion (3). The densities of $U_C$ and $U_1 - U_0$ being log concave is equivalent to their densities being Polya frequency functions of order two (Karlin, 1968). Using that $U_1 - U_0 \perp\!\!\!\perp U_C$, one can now easily verify that $(U_C, U_C - (U_1 - U_0))$ and $(-(U_1 - U_0), U_C - (U_1 - U_0))$ have joint densities that are totally positive of order 2 (TP2). By Joe (1997) (Theorems 2.2, 2.3), $(U_C, U_C - (U_1 - U_0))$ and $(-(U_1 - U_0), U_C - (U_1 - U_0))$ having TP2 densities implies that $U_C$ and $-(U_1 - U_0)$ are stochastically increasing in $U_C - (U_1 - U_0)$ and thus stochastically increasing in $U_S$ using that $U_S$ is a strictly monotonic function of $U_C - (U_1 - U_0)$. Thus $E(U_C|U_S = u_S)$ is increasing in $u_S$ while $E(U_1 - U_0|U_S = u_S)$ is decreasing in $u_S$, establishing the assertion. □

### *Theorem 3*

*Proof.* Consider part (i) of the theorem. Let $\mu_{10}(\cdot) = \mu_1(\cdot) - \mu_0(\cdot)$, and let $\Upsilon(p) = E(U_1 - U_0 \mid U_S = p)$. From our previous analysis, we have

$$\text{(A.1.1)} \quad \frac{\partial}{\partial p} E(Y|X = x, P = p) = \mu_{10}(x) + \Upsilon(p) \qquad \text{a.e. } (x, p).$$

Let $(\mu^*_{10}, \Upsilon^*)$ denote candidate functions that also satisfy equation (A.1.1). We then have $\mu^*_{10}(x) - \mu_{10}(x) = \Upsilon(p) - \Upsilon^*(p)$ for a.e. $(x, p)$. By the rank condition (A-5), we have that, for some constant $C$: $\mu^*_{10}(x) - \mu_{10}(x) = C$ for a.e. $x$, and $\Upsilon^*(p) - \Upsilon(p) = -C$ for a.e. $p$. We thus have that $\mu^*_{10}(x) + \Upsilon^*(p) = \mu_{10}(x) + \Upsilon(p) = B^{MTE}(x, p)$ for a.e. $x$ and a.e. $p$. We have thus established identification of $B^{MTE}(x, p)$ for $(x, p) \in$

$\text{Supp}(X) \times \text{Supp}(P)$. The same argument *mutatis mutandis* shows identification of $C^{MTE}(z, u_S)$ for $(z, u_S) \in \text{Supp}(Z) \times \text{Supp}(P)$, and we thus have identification of $S^{MTE}(x, z, u_S)$ for $(x, z, u_S) \in \text{Supp}(X) \times \text{Supp}(Z) \times \text{Supp}(P)$. Parts (ii) and (iii) of the theorem now follow using part (i) of the theorem and the representation of the ATE and TT parameters as integrals of the MTE parameters. □

## Theorem 4

*Proof.* Assertion (1) follows from equation (2.3.1) and $B^{MTE}(x, u_S) = \mu_1(x) - \mu_0(x)$ if $U_1 - U_0$ is degenerate. Assertion (2) follows from $E(Y|X = x, P(X,Z) = 1) - E(Y|X = x, P(X,Z) = 0) = B^{ATE}(x)$, $[E(Y|X = x, P(X,Z) = p) - E(Y|X = x, P(X,Z) = 0)]/p = E(B|X = x, P(X,Z) = p, D = 1)$, and that $B^{ATE}(x) \leq E(B|X = x, P(X,Z) = p, D = 1)$ by the arguments used to prove Assertion (2) of Theorem 1. Assertion (3) follows from equation (2.3.1) and Assertion (3) of Theorem 1. □

# B NLSY Data

The dataset is based on the National Longitudinal Survey of Youth of 1979 (NLSY79).[1] 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.

**Educational Choice** Individuals are separated into two groups. The first comprises high school dropouts and high school graduates, while the second is made up

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

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 B.1:** 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 B.2:** 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 B.3:** 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.

# C NEWWS Data

We present a brief description of the Nation Evaluation of Welfare-to-Work Strategies (NEWWS) dataset. The following paragraphs are directly from various documentation sources including the webpage which is available at `http://aspe.hhs.gov/hsp/NEWWS`. Additional information is provided in Freedman et al. (2000) and Hamilton et al. (2001).

The Department of Health and Human Services (HHS) undertook a study of the effectiveness of welfare-to-work programs. The NEWWS evaluation is a study of the effectiveness of eleven mandatory welfare-to-work programs in seven locales: Atlanta, Georgia; Columbus, Ohio; Detroit and Grand Rapids, Michigan; Oklahoma City, Oklahoma; Portland, Oregon; and Riverside, California. Program impacts were evaluated by comparing outcomes for a randomly assigned experimental group subject to program requirements with outcomes for control groups. As part of the National Evaluation of Welfare-to-Work Strategies (NEWWS), the effects of two approaches to preparing welfare recipients for employment were compared in three sites (Atlanta, Grand Rapids, and Riverside). In one approach, the human capital development approach (HCD), individuals were directed to avail themselves of education services and, to a lesser extent, occupational training before they sought work, under the theory that they would then be able to get better jobs and keep them longer. In the other approach, the labor force attachment approach (LFA), individuals were encouraged to gain quick entry into the labor market, even at low wages, under the theory that their work habits and skills would improve on the job and they would thereby be able to advance themselves.

The evaluation used a random assignment design to get reliable results. Sample members were followed for five years from the time they entered the study. Comprehensive data on economic outcomes, including information on quarterly Unemployment Insurance-reported earnings and monthly Temporary Assistance for Needy Families (TANF) and Food Stamp payments was collected. A broad range of data was collected through surveys including data on educational attainment, family composition, housing status, wage progression, employment, child care, depression, and total family income. In addition, effects on the well-being of the children of the mothers in the study was evaluated. Four types of child outcomes were measured: cognitive development and academic achievement; safety and health; problem behavior and emotional well-being; and social development. Assessments in each of these areas will be compared across research groups two and five years after the mothers entered the survey sample. An overview on the research using this dataset is provided online.

Some sample members have missing values for these measures. These must be imputed prior to calculating impacts. The public use file contains imputed values for these measures based on mean substitution by site and the sample member's educational attainment at random assignment.

Next, we present the descriptive statistics for the sample used in the analysis.

### Individual Characteristics

**Table C.1:** Covariates

| Atlanta | | | |
|---|---|---|---|
| Individual Characteristics | SS | HCD | LFA |
| Single Parent, Ever Married | 0.401 | 0.414 | 0.391 |
| Two Children | 0.337 | 0.334 | 0.318 |
| Three or More Children | 0.299 | 0.288 | 0.298 |
| Any Child 0-5 Years Old | 0.427 | 0.417 | 0.430 |
| Black | 0.939 | 0.958 | 0.951 |
| Age | 33.062 | 33.144 | 33.018 |
| Female | 0.965 | 0.962 | 0.965 |
| High School Diploma, GED | 0.609 | 0.606 | 0.609 |
| Observations | 1,946 | 1,935 | 1,887 |

### Outcome

**Table C.2:** Outcomes

| Atlanta | | | |
|---|---|---|---|
| Earnings | SS | HCD | LFA |
| prior treatment, 6 - 5 quarters | 1,041 | 1,013 | 976 |
| prior treatment, 4 - 3 quarters | 755 | 710 | 738 |
| prior treatment, 2 - 1 quarters | 561 | 481 | 573 |
| post treatment, 2 - 21 quarters | 19,722 | 20,979 | 21,778 |
| Observations | 1,946 | 1,935 | 1,887 |

**Notes**: in units of $1.

# D Supplementary Material

**Table D.1:** 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).

# E Estimation Material

**Table E.1:** 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.

# F Additional Material

# Affidavit

I hereby confirm that my thesis entitled “Essays in the Econometrics of Policy Evaluation” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

# Curriculum Vitae

## Philipp Eisenhauer

---

### Education

| | |
|---|---|
| since 09/2007 | Ph.D. student at the Graduate School of Economic and Social Sciences (GESS) in Mannheim, Germany<br>Specialization: Econometrics and Computational Statistics<br>Dissertation: *Essays in the Econometrics of Policy Evaluation*, Supervisor: Prof. Dr. Dr. h.c. mult. Wolfgang Franz |
| 09/2005–02/2006 | Ph.D. studies on an Erasmus Grant at the European Center for Advanced Research in Economics and Statistics (ECARES) in Brussels, Belgium |
| 10/2002–08/2007 | Diploma studies in economics at the University of Mannheim, Germany<br>Specialization: Public Policy<br>Thesis: *Assessing Intergenerational Income Mobility in Germany*, Supervisor: PD Dr. Friedhelm Pfeiffer |

December 12, 2012