IHS Economics Series
Working Paper 297
June 2013

Doubly Robust Estimation of Causal Effects with Multivalued Treatments

S. Derya Uysal

IRIHS Institutional Repository at IHS Vienna

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Impressum

Author(s):
S. Derya Uysal

Title:
Doubly Robust Estimation of Causal Effects with Multivalued Treatments

ISSN: Unspecified

2013 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

All IHS Working Papers are available online:
http://irihs.ihs.ac.at/view/ihs_series/
This paper is available for download without charge at:
https://irihs.ihs.ac.at/id/eprint/2207/

297

Reihe Ökonomie

Economics Series

Doubly Robust Estimation of Causal Effects with Multivalued Treatments

S. Derya Uysal

297

Reihe Ökonomie

Economics Series

Doubly Robust Estimation of Causal Effects with Multivalued Treatments

S. Derya Uysal

June 2013

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

S. Derya Uysal
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria.
☎: +43/1/599 91-156
email: uysal@ihs.ac.at

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share "work in progress" in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschafts-wissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

This paper provides doubly robust estimators for treatment effect parameters which are defined in multivalued treatment effect framework. We apply this method on a unique data set of British Cohort Study (BCS) to estimate returns to different levels of schooling. Average returns are estimated for entire population, as well as conditional on having a specific educational achievement. The analysis is carried out for female and male samples separately to capture possible gender differences. The results indicate that, on average, the percentage wage gain due to higher education versus any other lower educational attainment is higher for highly educated females than highly educated males.

Keywords

Multivalued treatment, returns to schooling, doubly robust estimation

JEL Classification

C21, J24, I2

Comments
Supported by funds of the Oesterreichische Nationalbank (Anniversary Fund, project number: 14986)

Contents

1 Introduction

Estimation of the causal effects of a binary treatment under the conditional independence assumption has been studied extensively in the program evaluation literature (see for example Wooldridge 2007, Heckman et al. 2007, Imbens 2004, Rosenbaum & Rubin 1983, among others). While the literature dealing with the binary treatment variables is comprehensive, the discussion on multivalued treatment variables is more recent and sparse. Given that in many empirical applications the programs which are being evaluated offer more alternatives than just one possible treatment the methods dealing more general treatment effect regimes are particularly valuable. For example, from policy makers point of view it is usually more preferable to get information on the causal effects of different labor market programs rather than just looking at the effect of participating in any one of the programs versus not participating. Similarly, the effects of different doses of a drug or, as in this paper, the effects of different levels of educational attainment might be more interesting than just looking at the binary cases.

This study provides a simple method to estimate the causal effects of a multivalued treatment variable which possesses a property known as *double robustness*. In general, doubly robust methods combine two estimations methods each of which estimates the same parameter(s) of interest but uses different model specifications. Thus, doubly robust methods require both models specifications. The advantage of using both model specifications is that the parameter(s) of interest can be consistently estimated even if one of the model specifications is wrong which is not the case when the methods are used alone.[1] In other words, using doubly robust methods provides the practitioners more chances to get consistent estimators. Since in most of the applied works it is not possible to determine whether the model is correctly specified or not, having a doubly robust estimator for the parameter of interest might be quite useful. The method proposed here is closely related with the method used by Hirano & Imbens (2001) for the binary treatment case. We generalize the doubly robust estimator they are using for the potential outcome model with a multivalued treatment variable. The asymptotic distribution of the generalized doubly robust estimator is derived based on the results by Wooldridge (2002) and Wooldridge (2007). Additionally, the small sample properties of the proposed method and the underlying single methods are evaluated by a small Monte Carlo study. The interest of the simulation experiment lies on the demonstration of the double robustness property of the combination method under misspecified models as well as comparison of the small sample properties under correct model specifications. Furthermore, the doubly robust method is applied on the unique data set of British Cohort Study to estimate the returns to different levels of schooling.

Although this study is related to the several papers in different branches of the literature, it also differs from the existing literature in many ways. The interest on

[1] For further discussion on double robustness see Robins & Rotnitzky (1995), Robins et al. (1995), Robins & Ritov (1997), Hirano & Imbens (2001), Wooldridge (2007), Bang & Robins (2005), Tan (2006*a*), Tan (2006*b*), Tan (2010).

multivalued treatment effect in the program evaluation literature has been increasing mostly after Imbens (2000) and Lechner (2001). Imbens (2000) and Lechner (2001), almost simultaneously, define the assumptions, treatment effect parameters and the potential outcome framework for multivalued treatment parameter.[2] Following these papers, several papers contribute to the literature by extending the existing methods such as matching, weighting and regression, for different treatment parameters when a whole range of treatments are available (see Lechner 2002, Frölich 2004, Blundell et al. 2005). Tan (2010) considers the combination of the regression and weighting methods for multivalued treatment parameter. In fact, Tan (2010) investigates theoretical properties of another type of doubly robust estimator for unconditional means. Not only the form of the doubly robust estimator we are considering here is different, but also with our proposed model under the below described setup we are able to get doubly robust estimators for the conditional treatment effects. Another related paper is by Cattaneo (2010). He provides very general results on the efficient semi parametric estimation of multivalued treatment effects. Different from Cattaneo (2010) we only consider the parametric estimation of the probabilities. Despite the similarities, this study provides a contribution by explicitly investigating a specific type of doubly robust estimator for the conditional and unconditional mean effects in multivalued treatment effect framework. On the other hand, not only the econometric method we consider is different from the existing literature, but also the empirical study here differs in several ways from existing literature. First, the econometric approach proposed has not been applied on this question previously. Furthermore, due to the doubly robustness property of the proposed method, the results can be interpreted with more confidence. Another difference is that the returns to schooling are estimated using education as a multivalued treatment variable instead of a binary treatment variable or years of education. Taking into account the multivalued nature of education can provide further insights regarding the returns to education. Moreover, using the highest degree achieved as a treatment variable makes it possible to account for the fact that different levels of educational qualifications do not differ only in years but also in qualitative input they provide. Last but not least, the usage of the unique data set 1970 British Cohort Study (BCS70) with extensive control measures on cognitive and noncognitive ability as well as child's behavior justifies the identifying assumption at a reliable degree. Given that many recent papers like Heckman et al. (2006), Carneiro et al. (2007), Heineck et al. (2010), Uysal & Pohlmeier (2011), Blanden et al. (2007), Feinstein (2000) and Murasko (2007) provide empirical evidence on the importance of noncognitive and cognitive skills in determining different outcomes such as school performance, earnings, labor force participation, and job finding success, it is advantageous that the BCS70 gives the possibility to measure certain dimensions of noncognitive skills and cognitive skills besides the usual control variables.

The organization of the paper is as follows: Section 2 introduces the parameters of interest for multivalued treatment and proposes a weighted regression method to get doubly robust estimators of the treatment parameters of interest. In Section 3,

[2]See also Hirano & Imbens (2004) and Imai & van Dyk (2004) for the extension of this idea to the continuous treatment variable.

theoretical results on double robustness are illustrated by means of a small Monte Carlo Study. Section 4 motivates the empirical study and describes the data set used for the application. Moreover, in Section 4 the proposed estimator is applied to estimate causal effects of different educational levels on earnings and the estimation results are discussed in detail. Finally, Section 5 summarizes the main results and concludes the paper.

2 Econometric Method

The basic setup for the proposed doubly robust estimation method is based on Imbens (2000) and Lechner (2001). The interest lies in the causal effects of the treatment on some outcome variable, where the treatment of interest, T_i, takes the integer values between 0 and K. Consider N units which are drawn from a large population. For each individual i, $i = 1, ..., N$, in the sample the triple (Y_i, T_i, X_i) is observed. $D_{it}(T_i)$ is the indicator of receiving the treatment t for individual i:

$$D_{it}(T_i) = \begin{cases} 1, & \text{if} \quad T_i = t \\ 0, & \text{otherwise} \end{cases}$$

The vector of characteristics (covariates) for the i^{th} individual is denoted by X_i. For each individual there is a set of potential outcomes $(Y_{i0}, \ldots, Y_{iK})$. Y_{it} denotes the outcome for each individual i, for which $T_i = t$ where $t \in \mathfrak{T} = \{0, \ldots, K\}$. Only one of the potential outcomes is observed depending on the treatment status. Adopting the potential outcomes framework pioneered by Rubin (1974), the observed outcome, Y_i, can be written in terms of treatment indicator, $D_{it}(T_i)$, and the potential outcomes, Y_{it},:

$$Y_i = \sum_{t=0}^{K} D_{it}(T_i) Y_{it}. \tag{2.1}$$

Lechner (2001) defines several pairwise treatment effects. The first is the average effect of the treatment m relative to treatment l. It measures the mean effect of treatment over the entire population:

$$\tau^{ml} = \mathrm{E}\left[Y_{im} - Y_{il}\right] = \mu_m - \mu_l. \tag{2.2}$$

The second treatment effect is the expected effect for an individual randomly drawn from the population of participants who receive the treatment m:

$$\gamma^{ml|m} = \mathrm{E}\left[Y_{im} - Y_{il} \middle| T_i = m\right] = \mu_{m|m} - \mu_{l|m} \tag{2.3}$$

The average treatment effects τ^{ml} and τ^{lm} are symmetric, i.e. $\tau^{ml} = -\tau^{lm}$, but $\gamma^{ml|m} \neq -\gamma^{lm|l}$. $\gamma^{ml|m}$ measures the effect of the treatment m with respect to the treatment l for the subpopulation of individuals who receive the treatment m. On the other hand, $-\gamma^{lm|l}$ measures the treatment effect of m with respect to the l for the subpopulation of individuals who receive the treatment l.

Since only one of the potential outcomes is observed, the above defined average treat-

ment effects cannot be identified from observed data without further assumptions. For the rest of the paper the Conditional Independence Assumption as defined by Imbens (2000) assumed to be satisfied:

Definition 1. *Conditional Independence Assumption (CIA)*
$Y_{it} \perp D_{it}(T_i) | X_i$, $\forall\ t \in \mathfrak{T}$, *where* $\perp$ *stands for independence.*

This implies that the assignment to the treatment is weakly unconfounded given pre-treatment variables X. As noted by Imbens (2000), this assumption is similar to the missing at random assumption of Rubin (1976) and Little & Rubin (1987) in the missing data literature. Under this assumption one can identify $\mathrm{E}[Y_{it}]$ by adjusting for X:

$$\begin{aligned} \mathrm{E}[Y_{it} | X_i] &= \mathrm{E}[Y_{it} | D_{it}(T_i) = 1, X_i] = \mathrm{E}[Y_i | D_{it}(T_i) = 1, X_i] \\ &= \mathrm{E}[Y_i | T_i = t, X_i] \qquad \forall t \in \mathfrak{T} \end{aligned}$$

Thus, the unconditional means can be estimated by averaging these conditional means, i.e.

$$\mu_t \equiv \mathrm{E}[Y_{it}] = \mathrm{E}[\mathrm{E}[Y_{it} | X_i]]. \tag{2.4}$$

Based on this identification result one can use regression adjustment to estimate $K+1$ conditional mean functions by a parametric regression as in the binary treatment case (see for example Hirano & Imbens 2001, Rubin 1977, for the regression adjustment of a binary treatment variable). The conditional mean functions of the potential outcomes are specified as follows:

$$\mathrm{E}[Y_{it} | X_i] = \mathrm{E}[Y_i | T_i = t, X_i] = \beta_{0t} + X_i' \beta_{1t}, \tag{2.5}$$

where $\beta_t = [\beta_{0t} \quad \beta_{1t}']'$ is the vector of unknown parameters and β_{1t} has the same dimension as X_i. After estimating the parameter vector β_t the treatment effect parameters, τ^{ml} and $\gamma^{ml|m}$, can be estimated by the following:

$$\hat{\tau}^{ml} = (\hat{\beta}_{0m} - \hat{\beta}_{0l}) + \frac{1}{N} \sum_{i=1}^{N} X_i' (\hat{\beta}_{1m} - \hat{\beta}_{1l}) \tag{2.6}$$

$$\hat{\gamma}^{ml|m} = (\hat{\beta}_{0m} - \hat{\beta}_{0l}) + \frac{1}{N_m} \sum_{i: D_{im}(T_i)=1} X_i' (\hat{\beta}_{1m} - \hat{\beta}_{1l}), \tag{2.7}$$

where N_t is the number of observations who take part in the treatment $T_i = t$. Instead of specifying the $(K+1)$ regression models, one can define one regression equation depending on the treatment parameter of interest to get estimates of μ_l or $\mu_{l|m}$ directly (see Appendix A for the derivations). Using the definition of the observed outcome in Equation (2.1), the regression model can be rewritten as in Equation (2.8) to estimate the unconditional means, μ_t, as parameters of the regression model.

$$Y_i = \sum_{t=0}^{K} \mu_t D_{it}(T_i) + \sum_{t=0}^{K} D_{it}(T_i)(X_i - \bar{X})' \alpha_t + \varepsilon_i \tag{2.8}$$

where $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$. The parameters μ_t and α_t are estimated by minimizing the

objective function which is the sum of squared residuals:

$$\min_{\mu_t,\alpha_t} \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \sum_{t=0}^{K}\mu_t D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)(X_i - \bar{X})'\alpha_t\right)^2 \equiv \min_{\mu_t,\alpha_t}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2. \quad (2.9)$$

If the conditional mean function in Equation (2.5) is correctly specified $\hat{\mu}_t \xrightarrow{p} \mu_t = \mathrm{E}[Y_{it}]$. Thus, using the estimators for $\hat{\mu}_m$ and $\hat{\mu}_l$, τ^{ml} can be estimated as:

$$\hat{\tau}^{ml} = \hat{\mu}_m - \hat{\mu}_l. \quad (2.10)$$

If the interest lies in the treatment effect parameter $\gamma^{ml|m}$, one can reformulate the regression model in Equation (2.8) as follows:

$$Y_i = \sum_{t=0}^{K}\mu_{t|m}D_{it}(T_i) + \sum_{t=0}^{K} D_{it}(T_i)(X_i - \bar{X}_m)'\alpha_{t|m} + \varepsilon_i \quad (2.11)$$

where $\bar{X}_m = \frac{1}{N_m}\sum_{i:D_{im}(T_i)=1} X_i$. The minimization problem for this regression model is given by:

$$\min_{\mu_{t|m},\alpha_{1t|m}} \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \sum_{t=0}^{K}\mu_{t|m}D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)(X_i - \bar{X}_m)'\alpha_{t|m}\right)^2 \equiv \min_{\mu_{t|m},\alpha_{1t|m}}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2. \quad (2.12)$$

The coefficients of the treatment indicator variables, $\mu^m_{t|m}$, estimate $\mathrm{E}[Y_{it}|\, T_i = m] \equiv \mu_{t|m}$ consistently if the conditional mean of Y_{it} is correctly specified. Thus,

$$\hat{\gamma}^{ml|m} = \hat{\mu}_{m|m} - \hat{\mu}_{l|m}. \quad (2.13)$$

Another estimation approach is to construct propensity score weighting type estimators for the relevant treatment effect parameters. For weighting type estimators, one needs to generalize the concept of propensity score for the case of multivalued treatment effect.[3] Imbens (2000) defines the Generalized Propensity Score as follows:

Definition 2. *The Generalized propensity score (GPS) is the conditional probability of receiving a particular level of the treatment given the pre-treatment variables:*

$$r(t,x) \equiv \Pr[T_i = t\,|X_i = x] = \mathrm{E}[D_{it}(T_i)|\, X_i = x]. \quad (2.14)$$

Using the GPS Imbens (2000) shows that, similar to the binary treatment case, one can identify the unconditional means of the potential outcomes by weighting:

$$\mathrm{E}\left[\frac{Y_i D_{it}(T_i)}{r(t,X_i)}\right] = \mathrm{E}[Y_{it}] \quad (2.15)$$

[3] In binary treatment analysis, the conditional probability of receiving the treatment is called the Propensity score.

Based on this identification result, the treatment effect estimators are given by:

$$\hat{\tau}^{ml} = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i D_{im}(T_i)}{\hat{r}(m,X_i)} - \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i D_{il}(T_i)}{\hat{r}(l,X_i)} \tag{2.16}$$

$$\hat{\gamma}^{ml|m} = \frac{1}{N_m}\sum_{i=1}^{N} Y_i D_{im}(T_i) - \frac{1}{N_m}\sum_{i=1}^{N} D_{il}(T_i) Y_i \frac{\hat{r}(m,X_i)}{\hat{r}(l,X_i)} \tag{2.17}$$

where $\hat{r}(t,X_i)$ is the estimated GPS. One can estimate $r(t,X_i)$ by discrete response models if the multivalued treatment does not have a logical ordering, or by ordered response models if the treatment corresponds to ordered levels (Imbens 2000).

To get doubly robust estimators for the treatment effect parameters we propose to combine the GPS weighting approach with the regression approach. Basically, we are using a weighted regression method with the weights related to the weighting identification. Hirano & Imbens (2001) use the same approach to estimate binary treatment effects. By generalizing their approach for multivalued treatment we increase the applicability of doubly robust methods on more general treatment regimes. The double robustness for the proposed estimation method implies that if the weights are estimated based on a correct GPS specification or if the potential outcomes are correctly specified, the resulting estimator will be consistent. The doubly robust estimator of τ^{ml} can be derived by estimating the regression model in Equation (2.8) by a weighted least squares regression with the following estimated weights:

$$\sum_{t=0}^{K}\frac{D_{it}(T_i)}{\hat{r}(t,X_i)}. \tag{2.18}$$

Thus, the minimization problem for doubly robust estimation is given by

$$\min_{\mu_t,\alpha_t}\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=0}^{K}\frac{D_{it}(T_i)}{\hat{r}(t,X_i)}\right)\left(Y_i - \sum_{t=0}^{K}\mu_t D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)(X_i-\bar{X})'\alpha_t\right)^2. \tag{2.19}$$

The resulting estimators, $\hat{\mu}_t^w$, are consistent for μ_t if (i) the conditional mean of Y_{it} is correctly specified, (ii) the conditional mean of $D_{it}(T_i)$ is correctly specified or (iii) both. By using $\hat{\mu}_m^w$ and $\hat{\mu}_l^w$ instead of the unweighted regression estimators $\hat{\mu}_m$ and $\hat{\mu}_l$ in Equation (2.10), the treatment effect τ^{ml} is estimated doubly robustly (see A for the demonstration of the double robustness), i.e.:

$$\tau_{dr}^{ml} = \hat{\mu}_m^w - \hat{\mu}_l^w. \tag{2.20}$$

For doubly robust estimation of $\gamma^{ml|m}$, one can use the regression model given in Equation (2.11) with the following weights:

$$\sum_{t=0}^{K} D_{it}(T_i)\frac{\hat{r}(m,X_i)}{\hat{r}(t,X_i)}. \tag{2.21}$$

Accordingly, the weighted regression estimators of $\mu_{t|m}$ and $\alpha_{1t|m}$ solve the following minimization problem

$$\min_{\mu_{t|m},\alpha_{1t|m}} \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=0}^{K} D_{it}(T_i)\frac{\hat{r}(m,X_i)}{\hat{r}(t,X_i)}\right)\left(Y_i - \sum_{t=0}^{K}\mu_{t|m}D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)(X_i - \bar{X}_m)'\alpha_{t|m}\right)^2. \tag{2.22}$$

$\hat{\mu}^w_{t|m}$ for $t = 0, \ldots, K$, which is derived as the solution to the above given minimization problem, is doubly robust estimator of $\mu_{t|m}$. Hence, $\hat{\mu}^w_{m|m}$ and $\hat{\mu}^w_{l|m}$ are used to estimate $\gamma^{ml|m}$ doubly robustly:

$$\gamma^{ml|m}_{dr} = \hat{\mu}^w_{m|m} - \hat{\mu}^w_{l|m}. \tag{2.23}$$

To estimate the standard errors, one might use bootstrap methods as it has been done in most of the applications in programm evaluation or one could estimate the standard errors based on the asymptotic variance. In the following, we derive the asymptotic distribution for the estimators of the treatment parameters where the GPS is estimated by multinomial response models, however one easily follow the results for ordered response models. First, the asymptotic distribution of the estimators which are solutions to the minimization problems given by Equations (2.19) and (2.22) has to be derived. It is important to consider that the weights are estimated. The approach of Wooldridge (2007) and Wooldridge (2002) for two step estimation with generated regressors is used to derive the asymptotic distribution. Wooldridge (2007) derives the asymptotic distribution for the estimates of a weighted regression with binary treatment variable. This can be easily adjusted for the case of multivalued treatment effect. The advantage of using the models in Equation (2.8) and (2.11) is that by deriving the asymptotic distribution of the parameter estimates one also obtains the asymptotic distribution of $\hat{\mu}^w_t$ and $\hat{\mu}^w_{t|m}$. Since the treatment parameters of interest are simple functions of $\hat{\mu}^w_t$ and $\hat{\mu}^w_{t|m}$, simple application of the Delta Method will be sufficient to derive the asymptotic distribution of the treatment parameters.

Let $r(t, X_i; \psi_t)$ be the parametric model for $r(t, x)$, i.e. $\Pr[T_i = t \,|X_i] = r(t, X_i; \psi_t)$, where $\boldsymbol{\psi} \in \boldsymbol{\Psi} \subset \mathbb{R}^{M\times(K+1)}$ with $\boldsymbol{\psi} = [\psi_0' \quad \psi_1' \ldots \psi_K']'$. The estimator $\hat{\boldsymbol{\psi}}$ solves a conditional likelihood problem of the form

$$\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \sum_{i=1}^{N} \ln L(\boldsymbol{\psi}; D_{it}(T_i), X_i) = \sum_{i=1}^{N}\sum_{t=0}^{K} D_{it}(T_i)\ln r(t, X_i; \psi_t).$$

Since the probabilities sum up to one, parameter identification requires a normalization such as $\psi_0 = \mathbf{0}$. Thus the individual score functions of dimension $M \times 1$ are given by:

$$c_{ti}(\boldsymbol{\psi}; D_{it}(T_i), X_i) \equiv \frac{\partial \ln L(\boldsymbol{\psi}; D_{it}(T_i), X_i)}{\partial \psi_t}, \quad t = 1, \ldots, K.$$

Let θ be $P \times 1$ parameter vector contained in a parameter space $\Theta \subset \mathbb{R}^P$. θ denotes

either (μ_t, α_t) or $(\mu_{t|m}, \alpha_{t|m})$. Thus, $\hat{\theta}$ solves the following minimization problem:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \hat{\omega}_i \varepsilon_i^2,$$

where ε_i is the sum of squared residuals for the corresponding regression model and $\hat{\omega}_i = \sum_{t=0}^{K} \frac{D_{it}(T_i)}{r(t,X_i;\hat{\psi}_t)}$ or $\hat{\omega}_i = \sum_{t=0}^{K} D_{it}(T_i) \frac{r(m,X_i;\hat{\psi}_m)}{r(t,X_i;\hat{\psi}_t)}$ depending on the treatment parameter of interest. Since the estimation problem in the multivalued treatment case is same as the binary treatment case, Theorem 3.1 in Wooldridge (2007) applies immediately.[4] Define $s_i = s(Y_i, X_i, T_i; \theta, \psi) \equiv \omega_i \frac{\partial \varepsilon_i^2}{\partial \theta}$ as the $P \times 1$ weighted score of the (unweighted) objective function $q(\cdot)$, $H(Y_i, X_i; \theta) = \frac{\partial^2 \varepsilon_i^2}{\partial \theta \partial \theta'}$ as the $P \times P$ Hessian of the objective function $q(\cdot)$. Under standard regularity conditions,

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, A^{-1} D A^{-1}\right), \tag{2.24}$$

where $A \equiv \mathrm{E}\left[H(Y_i, X_i; \theta)\right]$, $D \equiv \mathrm{E}\left[e_i e_i'\right]$, $e_i \equiv s_i - \mathrm{E}\left[s_i c_i'\right]\left[\mathrm{E}\left[c_i c_i'\right]\right]^{-1} c_i$, $c_i \equiv c_i(\boldsymbol{\psi}) = [c_{1i}' \dots c_{Ki}']'$ is the $MK \times 1$ score for the MLE of $\boldsymbol{\psi}$. Since the term D in the asymptotic distribution includes the score of the first step estimation, the resulting asymptotic distribution for second step takes into account that the weights are estimated. Wooldridge (2007) proposes consistent estimators of A and D in the binary treatment framework, which can be generalized to the following for multivalued treatment case:

$$\hat{A} \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{\omega}_i H(Y_i, X_i; \hat{\theta}) \tag{2.25}$$

and

$$\hat{D} \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{e}_i \hat{e}_i' \tag{2.26}$$

are consistent estimators of A and D where the
$\hat{e}_i \equiv \hat{s}_i - (N^{-1} \sum_{i=1}^{N} \hat{s}_i \hat{c}_i')(N^{-1} \sum_{i=1}^{N} \hat{c}_i \hat{c}_i')^{-1} \hat{c}_i$ are the $P \times 1$ residuals from the multivariate regression of $\hat{s}_i$ on $\hat{c}_i$ and hatted quantities are evaluated at $\hat{\theta}$ or $\hat{\boldsymbol{\psi}}$. Since the treatment effects τ^{ml} and $\gamma^{ml|m}$ are estimated as differences of regression parameters (Equations (2.20) and (2.23)), a straightforward application of Delta-method is sufficient to derive the variances of τ^{ml} and $\gamma^{ml|m}$ after getting a variance-covariance estimate of $\hat{\theta}$.

[4] Wooldridge (2007) derives in Theorem 3.1 the asymptotic distribution of the weighted regression parameter with estimated weights under CIA, where the weights are the estimated probabilities of receiving a binary treatment. Since his results follow the maximum likelihood theory (generalized conditional information matrix equality) and standard results on M-Estimation, the application of the theorem in multivalued treatment problem under CIA requires a straightforward adjustment of the score function. See for example Wooldridge (2002) Section 13.7 and Newey (1985).

3 Monte Carlo Evidence

This section presents a small Monte Carlo study to demonstrate the double robustness of the proposed method. Simulations are based on 2000 Monte Carlo samples with sample sizes $n = 500$, 2000 and 8000.[5] The data generating processes of $D_i^*(t)$ and Y_{it} for $t \in \mathfrak{T} = \{0, 1, 2\}$ are given below in Table 3.1.

Table 3.1: DGPs for $D_i^*(t)$ and Y_{it}

DGP1	$D_i^*(t) = \psi_{0t} + \psi_{1t}X_{i1} + \psi_{2t}X_{i2} + \psi_{3t}X_{i3} + \nu_{it}$ $Y_{it} = \beta_{0t} + \beta_{1t}X_{i1} + \beta_{2t}X_{i2} + \beta_{3t}X_{i3} + \varepsilon_{it}$
DGP2	$D_i^*(t) = \psi_{0t} + \psi_{1t}X_{i1} + \psi_{2t}X_{i2} + \psi_{3t}X_{i3} + \nu_{it}$ $Y_{it} = \beta_{0t} + \beta_{1t}X_{i1} + \beta_{2t}X_{i2} + \beta_{3t}X_{i3} + \beta_{4t}X_{i3}^2 + \varepsilon_{it}$
DGP3	$D_i^*(t) = \psi_{0t} + \psi_{1t}X_{i1} + \psi_{2t}X_{i2} + \psi_{3t}X_{i3} + \psi_{4t}X_{i3}^2 + \nu_{it}$ $Y_{it} = \beta_{0t} + \beta_{1t}X_{i1} + \beta_{2t}X_{i2} + \beta_{3t}X_{i3} + \varepsilon_{it}$

The value of the treatment variable, T_i, and the observed outcome variable, Y_i, are generated by the following observation rules:

$$T_i = \arg\max_{t \in \mathfrak{T}} \{D_i^*(t)\} \tag{3.27}$$

$$D_{it}(T_i) = \mathbb{1}\{T_i = t\} \tag{3.28}$$

$$Y_i = \sum_{t=0}^{K=2} D_{it}(T_i) Y_{it}. \tag{3.29}$$

X_{1i}, X_{2i} and X_{3i} are correlated uniform random variables distributed over $[-0.5, 0.5]$ with the correlation matrix V_X which is given by

$$V_X = \begin{bmatrix} 1.0 & 0.7 & 0.6 \\ 0.7 & 1.0 & 0.6 \\ 0.6 & 0.6 & 1.0 \end{bmatrix}.$$

Error terms ν_{i0}, ν_{i1} and ν_{i2} are drawn from independent *Gumbel* (0,1) distribution. This implies a multinomial logistic model for the GPS. ε_{i0}, ε_{i1} and ε_{i2} are independent standard normal variables. Table 3.2 summarizes the parameter values.

Table 3.2: Parameter Values for the Simulation Study

	Treatment Model					**Outcome Model**				
t	ψ_{0t}	ψ_{1t}	ψ_{2t}	ψ_{3t}	ψ_{4t}^*	β_{0t}	β_{1t}	β_{2t}	β_{3t}	β_{4t}^*
0	0	0	0	0	0	0	0.5	0.5	0.5	0.5
1	1	1	1	1	1	1	0.5	0.5	0.5	0.5
2	2	2	2	2	2	2	0.5	0.5	0.5	0.5

Note: ψ_{4t}^* is only used for DGP3 and β_{4t}^* is only used for DGP2.

[5] The sample sizes are unconventionally large, because otherwise with three treatment groups the number of observations in each group would have been too small. The data generation process used here creates subsamples with the treatment $T_i = 0$, $T_i = 1$ and $T_i = 2$ approximately 10%, 25%, 65% of the total observations, respectively.

For all three DGPs, the unconditional means of the potential outcomes, $\mathrm{E}[Y_{it}] = \mu_t$ $\forall t \in \mathfrak{T}$, the treatment parameters τ^{ml} as well as $\gamma^{ml|m}$ for all possible combinations of m and l are estimated by three methods: weighting, regression and the doubly robust method. Weighting model requires specification of GPS model, whereas regression method requires specification of outcome model. The doubly robust method requires both specifications. The GPS is estimated by multinomial logit based on the following model specification:

$$r(t, x_i) \equiv \Pr[T_i = t \,|X_i] = \frac{\exp(\psi_{0t} + \psi_{1t}X_{i1} + \psi_{2t}X_{i2} + \psi_{3t}X_{i3})}{\sum_{j=0}^{2}\exp(\psi_{0j} + \psi_{1j}X_{i1} + \psi_{2j}X_{i2} + \psi_{3j}X_{i3})}, \tag{3.30}$$

and the outcome model for Y_{it} is specified as follows:

$$\mathrm{E}[Y_{it}|\, X_i] = \beta_{0t} + \beta_{1t}X_{i1} + \beta_{2t}X_{i2} + \beta_{3t}X_{i3}. \tag{3.31}$$

The model specification given in Equation 3.30 is correct for DGP1 and DGP2, but it is wrong for DGP3. Thus, weighting estimators which relies on the estimated GPS based on this model specification will not be consistent for DGP3, but will be consistent for the other DGPs. The outcome model in Equation (3.31) is only correct for DGP1 and DGP3. Hence, the regression estimators will be inconsistent for DGP2. However, the doubly robust estimators which use both model specifications will be consistent for all three DGPs, since for each DGP at least one of the model specifications is correct.

Table 3.3: Summary of Monte Carlo Results

		DGP1			**DGP2**			**DGP3**		
		Both Correct			**Outcome Wrong**			**GPS Wrong**		
	N	500	2000	8000	500	2000	8000	500	2000	8000
WE	ABIAS	0.01	0.00	0.00	0.01	0.00	0.00	**0.02**	**0.02**	**0.02**
	ASE	0.19	0.09	0.04	0.19	0.09	0.05	0.18	0.09	0.04
	ARMSE	0.19	0.09	0.04	0.19	0.09	0.05	0.18	0.09	0.05
REG	ABIAS	0.00	0.00	0.00	**0.01**	**0.01**	**0.01**	0.00	0.00	0.00
	ASE	0.16	0.08	0.04	0.16	0.08	0.04	0.16	0.08	0.04
	ARMSE	0.16	0.08	0.04	0.16	0.08	0.04	0.16	0.08	0.04
DR	ABIAS	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	ASE	0.18	0.08	0.04	0.18	0.08	0.04	0.18	0.09	0.04
	ARMSE	0.18	0.08	0.04	0.18	0.08	0.04	0.18	0.09	0.04

AABIAS: average of absolute bias, ASE: average standard error, ARMSE: average root of the mean squared error over twelve parameter estimates.

The results of the Monte Carlo experiment are summarized in Table 3.3. Here, only the averages of absolute biases (AABIAS), standard errors (ASE) and root of the mean squared errors (ARMSE) over twelve parameters for each DGP are reported. Tables B.1-B.3 display detailed simulation results for each μ_t, τ^{ml} and $\gamma^{ml|m}$. The results clearly demonstrate the double robustness of the proposed estimation method. Under correct specification of the relevant models, all three methods estimate the parameters consistently. The most efficient method is the regression method, followed by the doubly robust method. The efficiency difference, however, is negligible.

Interestingly, the efficiency of weighting method is only slightly less than the doubly robust method. This might be due to the treatment homogeneity, i.e. treatment effects do not change with the covariates. Under both types of misspecifications, doubly robust estimators stay consistent, whereas the misspecification of the outcome model leads to inconsistent regression estimators and misspecification of the GPS leads to inconsistent weighting estimators, i.e. the biases do not decrease as the sample size increases. Obviously this Monte Carlo study does not consider more general cases like heterogeneous treatment or overlap problems, but it demonstrates the double robustness of the proposed method under misspecification of one of the models. A more comprehensive Monte Carlo study with a more general design is necessary to evaluate the properties of these methods more in detail. This, however, is beyond the scope of this paper.

4 Empirical Study

In the empirical part of this study, the returns to education at different levels are estimated by the doubly robust estimation method explained in Section 2. Estimation of causal effects of education on earnings is not trivial. Card (1999) provides a comprehensive review on problems associated with the estimation of the returns to education. The identification of the casual effects requires some strong assumptions on the selection to the participation mechanism either in terms of unobservables or observables. The usual method under assumption of selection on unobservables is the Instrumental Variable (IV) method, where the biggest challenge is to find a valid instrument. Card (1999), Card (2001) review the empirical results based on IV methods. On the other hand, if the assumption of conditional independence, i.e. selection on observables, is satisfied, there is no need for an instrumental variable and the causal effects can be identified by controlling for observable characteristics. This assumption however puts strong requirement on the data set. The variables available in a data set should be rich enough such that none of the important confounders of the treatment and outcome variable is left out. Due to data limitations, there are few studies where the causal effects are estimated by methods based on CIA (for example Blundell et al. 2005, Pohlmeier & Pfeiffer 2004, Flossmann & Pohlmeier 2006). The strong data requirement is not a restriction for this study because the data set provides standard control variables like gender, family background etc., as well as variables which are less common in surveys, such as several IQ measures, noncognitive skill measures and behavioral measures. The richness of the control variables makes unobserved ability problem less severe. Moreover, since all the variables are measured during the childhood before the measurement of the wages or any schooling choice is made, the problem of reverse causality is also avoided.

4.1 Data

1970 British Cohort Study (BCS70) is a longitudinal study which includes all the children born in the UK in the first week of April 1970. Since BCS70 began, there have been seven full data collection exercises in order to monitor the cohort members' health, education, social and economic circumstances. These took place when

respondents were aged 5, in 1975, aged 10, in 1980, aged 16, in 1986, aged 26, in 1996, aged 30, in 1999-2000, aged 34, in 2004-2005, and aged 38, in 2008-2009. For the empirical study, we use the birth survey, the surveys at the age of 10 and 30. The birth survey provides background information on the newborn and the parents. The second sweep, additional to the classical variables, includes very comprehensive measures on noncognitive and cognitive abilities, as well as child's behavioral problems. The last sweep is used to construct wages and the highest qualification level attained. After dropping all observations with missing information on any of the variables used and removing the children with congenital abnormalities, the sample used consists of 2424 males and 2261 females.

The richness of the measures available in the data set is important for the justification of the econometric method used in the current study. The crucial assumption is the conditional independence assumption which is not testable; though, it requires having all the important variables which affect both the treatment variable and the outcome variable in the data. Blundell et al. (2005) use another longitudinal study from UK and estimate the returns to schooling using various methods which rely on CIA. They use a rich data set and show some evidence that the assignment to the treatment is unconfounded given pre-treatment variables X. The data set used here contains equivalent information to their data set and some other measures on noncognitive, cognitive ability, as well as child's behavior. Since the data used here contain comprehensive measures, the CIA is not too unrealistic to hold. Another important issue for the CIA is that the covariates have to be unaffected by the treatment. This requirement in our study is fulfilled since the covariates are measured before the minimum school leaving age in the UK.

The outcome variable is the log hourly wages at the time of the fifth sweep. Therefore our sample consists of individuals who were employed at the time of the fifth sweep. We use the information on the last net payment they received, the period the period that corresponds to the payment and the weekly working hours.[6] The educational attainment, the treatment variable, is measured in detail in BCS70. For the empirical analysis, the educational attainment is categorized in four groups, which have a sequential nature. Table 4.1 summarizes which qualifications the categories include.

[6]In order to construct the outcome variable, as well as the treatment variable, the Appendix by Bynner et al. (2000) is followed.

Table 4.1: Educational qualifications and mapping to level of qualification

T	Level	General (Academic)	Vocationally-related (Applied)	Occupational (Vocational)
0	**No qualification**	No qualification GCSE grade D-G CSEs grades 2-5 Scottish standard grades 4-5 Other Scottish school qualification	Foundation GNVQ Other GNVQ	NVQ level 1 Other NVQ Units towards NVQ RSA Cert/Other Pitmans level 1 Other vocational qualifications HGV
1	**O-level**	GCSE grade A*-C O levels grade A-C O levels grade D-E CSE grade 1 Scottish standard grades 1-3 Scottish lower or ordinary grades	Intermediate GNVQ BTEC First Certificate BTEC First Diploma	NVQ level 2 Apprenticeships City Guilds Part 2/Craft/Intermediate City Guilds Part 1 Other RSA First Diploma Pitmans level 2
2	**A-Level**	A level AS levels Scottish Highers Scottish Cert of 6th Year Studies	Advanced GNVQ BTEC National Diploma ONC/OND	NVQ level 3 City Guilds Part 3 Final Advanced Craft RSA Advanced Diploma Pitmans level 3
3	**Higher Education**	Degree HE Diploma Higher Degree	BTEC Higher Certificate/Diploma HNC/HND	NVQ level 4-5 Professional degree qualifications Nursing/paramedic Other teacher training qualification City Guilds Part 4 Career Ext/Full Tech RSA Higher Diploma/PGCE

Figure 1 illustrates the distribution of log hourly wages by gender and Figure 2 illustrates the distribution of the wages by education level for males and females separately. Figure 1 does not indicate a big difference in the unconditional wage distribution of females and males. On the other hand, if we look at the wage distributions by educational attainment, we see that the distributions differ for both males and females. As expected, the most observable difference is between higher education and no qualification.

In the second sweep of BCS70, there are several measures related to the child's cognitive ability. Three different tests are used to construct indices to measure the cognitive ability. The first test is called "*Friendly Math Test (FMT)*". This test was

developed especially for the use of BCS70. It consists of a total 72 multiple choice questions and covers the rules of arithmetic, numbers skills, fractions, algebra, geometry and statistics. The variable is constructed as the number of correctly answered questions.

The second test is the "*Shortened Edinburgh Reading Test (SERT)*". It is the shortened version of the Edinburgh Reading Test developed by Godfrey Thomson Unit. The shortened test contains 67 items which examine vocabulary, syntax, sequencing, comprehension and retention. The variable to control for the reading ability is constructed as the sum of the correctly answered questions. The last test is the "*British Ability Scale (BAS)*". This test of cognitive attainment aims at measuring something akin to IQ (Elliot et al. 1978). They are two verbal and two non-verbal subscales. Verbal subscales comprise word definitions (37 items) and word similarities (42 items). Non-verbal subscales comprise recall of digits (34 items) and matrices (28 items). For each scale, the variables are constructed as the number of correct answers.

Furthermore, there are two tests related to the noncognitive abilities of the child: "(Lawseq) Self-Esteem Scale" and "(Caraloc) Locus of Control Scale" available in the second sweep. The Self-Esteem Scale was developed by Lawrence (1973). Lawrence (1973) defines self-esteem as a person's evaluation of his self-image in relation to his ideal self. The questions used in the survey are listed in the upper part of Table C.1. There are 16 questions, four of which are distractor questions. The distractor questions are marked with a star in the table. Children answer the questions with "Yes", "No" or "I do not know". The index to measure self-esteem is constructed following Lawrence (1996). All "No" answers get two points except for question 1. For question 1, answering with "Yes" is worth to two points. "I do not know" is worth for one point for all questions. The distractor questions do not contribute to the measure. High scores indicate higher self-esteem. The second noncognitive skill measure is constructed based on the Locus of Control questions. The concept of Locus of Control introduced by Rotter (1966) refers to an individual's perception about the underlying main causes of the events in his/her life. According to this concept, individuals range between externaliser and internaliser. Externalisers believe that the events in his/her life are caused by external factors like fate or luck. On the other hand, internalisers believe that the events in his/her life are caused by his/her personal decisions and efforts. The questionnaire was constructed from various tests to measure the locus of control (Gammage 1975). The children are asked 20 questions, to which they answer with "Yes", "No" or "I do not know". There are five distractor questions. From the answers a one dimensional scale is constructed as a measure of the degree of internalization. Each "No" response counts as one point, except for the question ten where "Yes" equals one point. The distractor questions do not count for the locus of control index. High scores indicate greater locus of control, i.e. higher degree of internalizing. The questions are listed in Table C.1.

In addition to the above mentioned measures, in the second sweep of BCS70 mothers have completed a set of questions which are related to the behavioral difficulties

of the child. Two different scales are used to construct indices to measure the behavior disorder. The first one is "Rutter Parental 'A' Scale of Behavior Disorder" (Rutter 1967, Rutter et al. 1970) and the second one is "Conner's Hyperactivity Scale" (Conners 1969). The list of related questions is in Table C.2. For both scales, mothers had to make a vertical mark through the line alongside each statement to indicate to what extent the child shows the behavior described. The line corresponds to a scale from 0 to 100. 0 refers to "does not apply" and 100 refers to "certainly applies". The overall Rutter score and Connor score for a cohort member at the age of 10 is the sum across the individual variables. Categorical ratings were calculated for each scale by dividing scores into three levels of severity: "normal" scores less than the 80th percentile, "moderate" problem scores between the 80th and 95th percentile and "severe" problem scores above the 95th percentile (this is a simplified version of the technique adopted in a paper by Thompson et al. 2003).[7]

In addition to the cognitive and noncognitive ability measures as well as the behavior disorder measures, some other information on the child's family background like mother's age and mother's education at child's birth, as well as the ethnicity and the gender of the child from the birth survey are included as controls. The information on household income, total number of children in the household and father's social class are taken from the second sweep. Furthermore, an indicator variable for whether the child lived with both parents since birth till age of 10 is included. Description of the variables and summary statistics are given in Tables C.3 and C.4, respectively.

4.2 Empirical Results

There are a number of studies dealing with estimation of the returns to schooling. In empirical studies education is usually taken as a binary treatment variable.[8] Considering multivalued treatment provides the opportunity for a better characterization of the returns to education at different levels. The proposed method described in Section 2, which is doubly robust against misspecification, is used in the econometric analysis. As noted earlier, the advantage of using a doubly robust estimator is that the treatment parameter of interest can be consistently estimated even if one of the underlying methods relies on a misspecified model.

The GPS is estimated by ordered logit, where the dependent variable is different levels of education: no qualification, O-Level, A-Level and higher Education.[9] The regression results of ordered logit estimation are represented in Table C.5. In Table C.6, the average partial effects are presented for interested readers. Although

[7] The percentiles are calculated using the raw data.

[8] Conti et al. 2011 use British Cohort Study in order to estimate the returns to education on non-market outcomes as well as earnings. The study here differs from their study in three important ways. First of all, they use Bayesian econometric methods using factor models. Thus, identification relies on different assumptions. Second, although they provide estimates of returns to education on earnings, the emphasis of their study is non-market outcomes. Furthermore, the education is measured as a dummy variable in their study not as a multivalued treatment variable.

[9] For robustness check, the probabilities are also estimated by sequential logit. Treatment effect estimates do not change qualitatively or quantitatively.

the GPS estimation results are not a direct interest of this study, it is worth mentioning that the results provide supporting evidence on the importance of cognitive and noncognitive abilities. The Locus of Control Scale, the Shortened Edinburgh Reading Test, the Friendly Math Test as well as some scales of British Ability Test are significant. For both males and females, being internaliser in terms of Locus of Control Scale decreases the probability of having no qualification, whereas increases the probability of having all other degrees. The magnitude of the positive effect is the highest for the O-Level. Higher scores in the Shortened Edinburgh Reading Test and in the Friendly Math Test decrease the probability of having no qualification and increase the probability of having any other degree. For females, similar effects are observable for word similarities scale and word definitions scale of BAS. For males, the word definitions scales and matrices scale of the BAS have significant effects on the probabilities. These results provide further evidence on the effects of cognitive and noncognitive abilities.

In the evaluation literature it is common to inspect the histogram estimates visually to determine lack of overlap. The histogram estimates of the GPS for individuals with $T_i = t$ and $T_i \neq t$ for each $t = \{0, 1, 2, 3\}$ for male and female samples are plotted in Figures 3 and 4. For females, the boundaries of the histograms have some gaps; however the probabilities over two different groups are distributed over the same interval. For males it seems even less problematic. Thus, there is no need to apply any common support adjustment.

After estimating the weights, the proposed doubly robust estimator is used as explained in Section 2 with corresponding weights to estimate the treatment effect parameters ($\gamma^{ml|m}$, τ^{ml}, $-\gamma^{lm|m}$) as well as the expected earnings for each level of education (μ_t). Mean Estimates of the log of earning by education levels are presented in Figure 5. There are significant differences in estimated earnings of females and males for different educational levels. Males earn on average more than females even after controlling for the covariates. For each education level, the expected earnings for males are larger than for females.

The estimated treatment effect parameters are summarized in Table 4.2 below. The results for females and males are reported in the upper and lower part of the table, respectively. All possible pairwise comparisons for four levels of education are considered. The reported numbers are % wage gains due to the treatment m relative to l. Average effect of m relative to l is estimated for three groups: (i) for the subpopulation $T_i = m$ ($\gamma^{ml|m}$); (ii) for the entire population (τ^{ml}), and (iii) for the subpopulation $T_i = l$ ($-\gamma^{lm|m}$). τ^{ml} is estimated as in Equation (2.20) and $\gamma^{ml|m}$ is estimated as in Equation (2.23) for all values of m and l. If $\gamma^{ml|m}$ is higher than τ^{ml} and τ^{ml} is higher than $-\gamma^{lm|m}$, the treatment is "efficient" in terms of the allocation of individuals to the particular treatment level m, i.e. the individuals who would benefit at most from the treatment level m are allocated into this treatment. The difference between $\gamma^{ml|m}$ and τ^{ml}, as well as the difference between τ^{ml} and $-\gamma^{lm|m}$ is called the "sorting gain" (Heckman & Li 2004). For example, if we consider the return of higher education (m) over no qualification (l), positive sorting gains would

imply that the individuals with higher ability are allocated to the appropriate educational institutions. However, negative sorting gains would indicate that there may be individuals with lower qualifications who should have received a higher educational degree according to their abilities.[10]

Table 4.2: Estimated Treatment Parameters: Average effect of m relative to l

	m	l	$\hat{\gamma}^{ml\|m}$	$\hat{\tau}^{ml}$	$-\hat{\gamma}^{lm\|m}$
Females	Higher Education	No Qualification	**19.0*****	**18.8*****	**16.8*****
	Higher Education	O-Level	**20.7*****	**19.2*****	**18.4*****
	Higher Education	A-Level	**14.0*****	**15.1*****	**17.5*****
	A-Level	No Qualification	2.9	3.7	2.4
	A-Level	O-Level	2.9	4.2	2.6
	O-Level	No Qualification	0.4	-0.5	1.4
	m	l	$\hat{\gamma}^{ml\|m}$	$\hat{\tau}^{ml}$	$-\hat{\gamma}^{lm\|m}$
Males	Higher Education	No Qualification	**18.5*****	**22.1*****	**23.3*****
	Higher Education	O-Level	**13.3*****	**16.8*****	**18.3*****
	Higher Education	A-Level	**11.9*****	**14.1*****	**15.3*****
	A-Level	No Qualification	**9.4*****	**8.0*****	**7.2*****
	A-Level	O-Level	3.5	2.7	2.8
	O-Level	No Qualification	**5.7****	**5.3***	3.4

Note: % wage gains due to the treatment are reported. *** 1% significance level, ** 5% significance level, * 10% significance level. Standard errors are calculated based on the asymptotic variance.

Females who have received higher education earn on average 19 % more by getting higher education instead of no qualification. The wage gain due to higher education compared to no qualification for the entire female sample is 18.8%. On the other hand, the percentage wage gain due to higher education for females who do not have any qualifications would be 16.8. Although, the differences between treatment effect estimates are very small, positive sorting gains are observed for females when returns to higher education is compared to no qualification. If we compare the corresponding results for males (18.5%, 22.1%, 23.3%), the ascending order ($\gamma^{ml|m} < \tau_{ml} < -\gamma^{lm|m}$)of the percentage wage gains indicate negative sorting gains. Males without any qualification would earn 23.3% more if they had received higher education, whereas those with a higher education degree earn only 18.5% more due to the higher education. This implies that the selection of males into higher education is "inefficient". Similarly, when higher education is compared with O-level, positive sorting gains are observed for females (20.7%, 19.2%, 18.4%) but negative for males (13.3%, 16.8%, 18.3%). This situation changes for females if the returns of higher education versus A-Level is compared: here the sorting gains are negative for both females (14%, 15.1%, 17.5%) and males (11.9%, 14.1%, 15.3%). Other pairwise comparisons for females do not yield any significant results. For males, significant gains due to A-Level over no qualification with positive sorting (9.4%, 8.0%, 7.2%) are observed. The gain over O-level over no qualification is also significant for males whose highest qualification is O-Level. The overall percentage wage gain due to O-level over no qualification is 5.3% but it is only significant at

[10] Flossmann & Pohlmeier (2006) use the same argument to when comparing the average causal returns for the German school tracks.

10% significance level.

Another striking result is that the percentage wage gain due to higher education versus any other lower educational attainment is higher for highly educated females than highly educated males. The difference is highest for higher education versus O-level. The percentage gain due to the higher education over O-level is 20.7% for highly educated females, whereas the same pairwise comparison for the males who received higher education suggests only 13.3% wage gain. On the other hand, males who have lower educational attainments like A-Level and O-Level benefit more from these attainments than females when compared to no qualification. If we consider the average returns of different levels of education for the entire population by gender, we see that the returns to higher education over no qualification are higher for males (22.1%) than for females (18.8%). However, the expected effect of higher education compared to O-level and A-level is higher for females (19.2%, 15.1%) than for males (16.6%, 14.1%).

5 Conclusion

In this paper, the returns to different levels of education are estimated using a novel estimation method. The proposed method is an extension of the doubly robust ATE and TT estimator in binary treatment case to the multivalued treatment evaluation problem under CIA. It combines the regression adjustment approach with the weighting approach. Regression adjustment requires model specifications for the conditional mean functions and the weighting approach requires model specification for the generalized propensity score. The advantage of the combination of the two methods is that the doubly robust method estimates the treatment parameter of interest consistently even if one of the models is misspecified. The estimation procedure is defined explicitly and the asymptotic distribution of the parameters is derived. The results generalize the studies by Hirano & Imbens (2001) and Wooldridge (2007) for multivalued treatment variables. Furthermore, a small Monte Carlo experiment is used to demonstrate the double robustness property of the proposed estimation method. The results clearly indicate that even if only one of the underlying models is correctly specified, the proposed method gives consistent estimates of the treatment parameters. Under correct specification of the outcome model, the regression adjustment is the most efficient one, but the efficiency difference between regression and doubly robust method is slight. These results indicate that the use of the doubly robust methods to estimate treatment parameters in multivalued treatment evaluation provide protection against misspecification at almost no efficiency costs. Given that in an empirical study it is difficult to be sure about the correctness of the model, doubly robust methods provides more chances to hit the correct model with no significant costs. Furthermore, the generalization introduced here is applicable to a wide range of program evaluation questions, since it deals with multivalued treatment effect. Given that the method is just a weighted regression; any standard econometric method can be used in practice.

In the empirical part of this paper, the unique data set of the British Cohort Study is used. Availability of various measures for cognitive and noncognitive ability as well as for behavioral disorder beyond the standard covariates makes CIA more likely to be valid. The effects are analyzed separately for males and females. The estimated earnings of females and males for different educational levels are found to be significantly different. The expected earnings for males are larger than that of females for any education level. The estimated treatment effects of different educational levels are also shown to be different for males and females. The percentage wage gains due to the higher education, A-Level and O-level over no qualification are higher for males than females if the entire population is considered. However, females' percentage wage gains due to the higher education over A-Level and O-Level are higher than males' gains. The results also suggest that highly educated women gain more from higher education than highly educated men on average.

References

Bang, H. & Robins, J. M. (2005), 'Doubly robust estimation in missing data and causal inference models', *Biometrics* **61**(4), 962–973.

Blanden, J., Gregg, P. & Macmillan, L. (2007), 'Accounting for intergenerational income persistence: Noncognitive skills, ability and education*', *The Economic Journal* **117**(519), C43–C60.

Blundell, R., Dearden, L. & Sianesi, B. (2005), 'Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey', *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **168**(3), 473–512.

Bynner, J., Butler, N., Ferri, E., Shepherd, P. & Smith, K. (2000), The design and conduct of the 1999-2000 surveys of the National Child Development Study and the 1970 British Cohort Study, CLS Cohort Studies Working Paper 1, Centre for Longitudinal Studies, Institute of Education University of London.

Card, D. (1999), Chapter 30 the causal effect of education on earnings, *in* O. C. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Part A of *Handbook of Labor Economics*, Elsevier, pp. 1801 – 1863.

Card, D. (2001), 'Estimating the return to schooling: Progress on some persistent econometric problems', *Econometrica* **69**(5), 1127–1160.

Carneiro, P., Crawford, C. & Goodman, A. (2007), Which skills matter?, *in* D. Kehoe, ed., 'Practice Makes Perfect: The Importance of Practical Learning', Social Markets Foundation, London, pp. 22–38.

Cattaneo, M. D. (2010), 'Efficient semiparametric estimation of multi-valued treatment effects under ignorability', *Journal of Econometrics* **155**(2), 138 – 154.

Conners, C. K. (1969), 'A teacher rating scale for use in drug studies with children', *American Journal of Psychiatry* **126**, 884–888.

Conti, G., Heckman, J. J., Lopes, H. F. & Piatek, R. (2011), Constructing economically justified aggregates: An application of the early origins of health, Working paper, Department of Economics, University of Chicago.

Elliot, C., Murray, D. J. & Pearson, J. (1978), British ability scales; manual 3: Directions for administration and scoring and manual 4: Tables of abilities and norms,

Technical report, National Foundation for Educational Research in England and Wales, Windsor, United Kingdom.

Feinstein, L. (2000), The relative economic importance of academic, psychological and behavioural attributes developed on childhood, Discussion Paper CEPDP 443, Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Flossmann, A. & Pohlmeier, W. (2006), 'Causal returns to education: A survey in empirical evidence for Germany', *Journal of Economics and Statistics* **226**(1), 6–23.

Frölich, M. (2004), 'Programme evaluation with multiple treatments', *Journal of Economic Surveys* **18**(2), 181–224.

Gammage, P. (1975), Socialization, schooling and locus of control, Technical report, Ph.D. thesis, University of Bristol.

Heckman, J. J. & Li, X. (2004), 'Selection bias, comparative advantage and heterogeneous returns to education: Evidence from China in 2000', *Pacific Economic Review* **9**(3), 155–171.

Heckman, J. J., Stixrud, J. & Urzua, S. (2006), '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., Vytlacil, E. J., Heckman, J. J. & Vytlacil, E. J. (2007), Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation, *in* J. J. Heckman & E. E. Leamer, eds, 'Handbook of Econometrics', Vol. 6, Part B, Elsevier, pp. 4779 – 4874.

Heineck, G., Anger, S., Heineck, G. & Anger, S. (2010), 'The returns to cognitive abilities and personality traits in Germany', *Labour Economics* **17**(3), 535 – 546.

Hirano, K. & Imbens, G. W. (2001), 'Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization', *Health Services and Outcomes Research Methodology* **2**, 259–278.

Hirano, K. & Imbens, G. W. (2004), The propensity score with continuous treatments, *in* A. Gelman & X.-L. Meng, eds, 'Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: : An Essential Journey with Donald Rubin's Statistical Family', John Wiley & Sons, Ltd, pp. 73–84.

Imai, K. & van Dyk, D. A. (2004), 'Causal inference with general treatment regimes', *Journal of the American Statistical Association* **99**(467), 854–866.

Imbens, G. W. (2000), 'The role of the propensity score in estimating dose-response functions', *Biometrika* **87**(3), 706–710.

Imbens, G. W. (2004), 'Nonparametric estimation of average treatment effects under exogeneity', *Review of Economics and Statistics* **86**(1), 4–29.

Lawrence, D. (1973), *Improved Reading through Counselling*, London: Ward Lock Educational.

Lawrence, D. (1996), *Enhancing self-esteem in the classroom*, Paul Chapman Publishing.

Lechner, M. (2001), Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, *in* M. Lechner & F. Pfeiffer, eds, 'Econometric Evaluation of Labour Market Policies', Heidelberg: Physica, pp. 43–58.

Lechner, M. (2002), 'Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies', *Review of Economics and Statistics* **84**(2), 205–220.

Little, R. J. A. & Rubin, D. B. (1987), *Statistical Analysis with Missing Data*, New York: John Wiley.

Murasko, J. E. (2007), 'A lifecourse study on education and health: The relationship between childhood psychosocial resources and outcomes in adolescence and young adulthood', *Social Science Research* **36**(4), 1348–1370.

Newey, W. K. (1985), 'Maximum likelihood specification testing and conditional moment tests', *Econometrica* **53**(5), 1047–1070.

Pohlmeier, W. & Pfeiffer, F. (2004), Returns to education and individual heterogeneity, ZEW Discussion Papers 04-34, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Robins, J. M. & Ritov, Y. (1997), 'Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models', *Statistics in Medicine* **16**(3), 285–319.

Robins, J. M. & Rotnitzky, A. (1995), 'Semiparametric efficiency in multivariate regression models with missing data', *Journal of the American Statistical Association* **90**(429), 122–129.

Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1995), 'Analysis of semiparametric regression models for repeated outcomes in the presence of missing data', *Journal of the American Statistical Association* **90**(429), 106–121.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

Rotter, J. (1966), 'Generalized expectancies for internal versus external control of reinforcement', *Psychological Monographs* **80**, 1–28.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and non-randomized studies', *Journal of Educational Psychology* **66**, 688 – 701.

Rubin, D. B. (1976), 'Inference and missing data', *Biometrika* **63**(3), 581–592.

Rubin, D. B. (1977), 'Assignment to treatment group on the basis of a covariate', *Journal of Educational Statistics* **2**(1), 1–26.

Rutter, M. (1967), 'A children's behaviour questionnaire for completion by children: preliminary findings', *Journal of Child Psychology and Psychiatry* **8**, 1–12.

Rutter, M., Tizard, J. & Whitmore, K. (1970), *Education, health and behaviour*, London: Longmans.

Tan, Z. (2006*a*), 'A distributional approach for causal inference using propensity scores', *Journal of the American Statistical Association* **101**(476), 1619–1637.

Tan, Z. (2006*b*), 'Regression and weighting methods for causal inference using instrumental variables', *Journal of the American Statistical Association* **101**(476), 1607–1618.

Tan, Z. (2010), 'Bounded, efficient and doubly robust estimation with inverse weighting', *Biometrika* **97**(3), 661–682.

Thompson, A., Hollis, C. & Richards, D. (2003), 'Authoritarian parenting attitudes as a risk for conduct problems – results from a British National Cohort Study', *European Child and Adolescent Psychiatry* **12**, 84–91.

Uysal, S. D. & Pohlmeier, W. (2011), 'Unemployment duration and personality', *Journal of Economic Psychology* **32**(6), 980 – 992.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

Wooldridge, J. M. (2007), 'Inverse probability weighted estimation for general missing data problems', *Journal of Econometrics* **141**(2), 1281 – 1301.

A Proofs

Derivation of Equation (2.8)

Due to Equation 2.5, the following regression model can be written for the potential outcomes:

$$Y_{it} = \beta_{0t} + X_i'\beta_{1t} + u_{it} \tag{A.1}$$

Equation 2.8 can be derived by combining 2.1 and A.1:

$$\begin{aligned}
Y_i &= \sum_{t=0}^{K} D_{it}(T_i) Y_{it} \\
Y_i &= \sum_{t=0}^{K} D_{it}(T_i)[\beta_{0t} + X_i'\beta_{1t} + u_{it}] \qquad + / - \sum_{t=0}^{K} D_{it}(T_i)\,\mathrm{E}\left[X_i\right]'\beta_{1t} \\
Y_i &= \sum_{t=0}^{K} D_{it}(T_i)[\beta_{0t} + \mathrm{E}\left[X_i\right]'\beta_{1t}] + \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]'\beta_{1t} + \sum_{t=0}^{K} D_{it}(T_i) u_{it} \\
Y_i &= \sum_{t=0}^{K} D_{it}(T_i)\,\mathrm{E}\left[\mathrm{E}\left[Y_{it}|\, X_i\right]\right] + \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]'\beta_{1t} + \sum_{t=0}^{K} D_{it}(T_i) u_{it} \\
Y_i &= \sum_{t=0}^{K} D_{it}(T_i)\,\mathrm{E}\left[Y_{it}\right] + \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]'\beta_{1t} + \sum_{t=0}^{K} D_{it}(T_i) u_{it} \\
Y_i &= \sum_{t=0}^{K} \mu_t D_{it}(T_i) + \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]'\alpha_t + \varepsilon_i
\end{aligned}$$

From the above derivation, we see that the coefficient of $D_{it}(T_i)$ stands for $\beta_{0t} + \mathrm{E}\left[X_i\right]'\beta_{1t}$, therefore it identifies the unconditional mean of Y_{it} if the conditional mean is correctly specified. Last two equalities show that $\beta_{1t} = \alpha_t$. In the regression model given by Equation (2.8), the unknown population mean $\mathrm{E}\left[X_i\right]$ is replaced by the sample mean, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$. Thus, if the conditional mean functions in Equation (2.5) is the correct specification, $\hat{\mu}_t \xrightarrow{p} \mu_t = \mathrm{E}\left[Y_{it}\right]$. By adding and subtracting $\sum_{t=0}^{K} D_{it}(T_i)\,\mathrm{E}\left[X_i|\, D_{im}(T_i) = 1\right]'\beta_{1t}$ in the second equality above, one gets Equation (2.11).

Double Robustness

First consider the unweighted regression adjustment to demonstrate that the consistency of treatment effect parameter depends on correct specification of the conditional mean function. Let θ be $P \times 1$ parameter vector contained in a parameter

space $\Theta \subset \mathbb{R}^P$. θ stands for (μ_t, α_t) in Equation (2.8).

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \sum_{t=0}^{K} \mu_t D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]' \alpha_t \right)^2$$

Consider the first $K+1$ first order conditions related to this minimization problem is given by:

$$\frac{1}{N} \sum_{i=1}^{N} D_{is}(T_i) \left(Y_i - \sum_{t=0}^{K} \mu_t D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]' \alpha_t \right) \quad \text{for} \quad s = 0, \ldots, K$$

If the following population counterparts of the above given moment functions have zero expectations, than the resulting parameter estimators will be consistent.

$$\begin{aligned}
&\mathrm{E}\left[D_{is}(T_i) \left(Y_i - \sum_{t=0}^{K} \mu_t D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]' \alpha_t \right) \right] \quad \text{for} \quad s = 0, \ldots, K \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) \left(Y_i - \sum_{t=0}^{K} \mu_t D_{it}(T_i) - \sum_{t=0}^{K} D_{it}(T_i)[X_i - \mathrm{E}\left[X_i\right]]' \alpha_t \right) \middle| X_i \right] \right] \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) Y_{is} - \mu_s D_{is}(T_i) - D_{is}(T_i)[X_i - \mathrm{E}\left[X_i\right]]' \alpha_s) \middle| X_i \right] \right] \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) \middle| X_i \right] \mathrm{E}\left[(Y_{is} - \mu_s - [X_i - \mathrm{E}\left[X_i\right]]' \alpha_s) \middle| X_i \right] \right] \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) \middle| X_i \right] \mathrm{E}\left[(Y_{is} - \beta_{0s} - \mathrm{E}\left[X_i\right]' \alpha_s - [X_i - \mathrm{E}\left[X_i\right]]' \alpha_s) \middle| X_i \right] \right] \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) \middle| X_i \right] \mathrm{E}\left[(Y_{is} - (\beta_{0s} + X_i' \alpha_s) \middle| X_i \right] \right] \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) \middle| X_i \right] \left[\mathrm{E}\left[Y_{is} \middle| X_i \right] - \mathrm{E}\left[\beta_{0s} + \mathrm{E}\left[X_i\right]' \beta_{1s} \middle| X_i \right] \right] \right] \\
&= \mathrm{E}\left[\mathrm{E}\left[D_{is}(T_i) \middle| X_i \right] \left[\mathrm{E}\left[Y_{is} \middle| X_i \right] - (\beta_{0s} + \mathrm{E}\left[X_i\right]' \beta_{1s}) \right] \right]
\end{aligned}$$

From the first to second equality we use law of iterated expectations. The third equality uses the fact that $D_{it}(T_i)$ is only once equal to one and K times it takes the value zero. By multiplying Equation (2.1) by $D_{is}(T_i)$ one can easily show that $D_{is}(T_i)Y_i = D_{is}(T_i)Y_{is}$. From third to fourth equality we apply CIA. For the next equality we use the definition of μ_s and the equality of $\beta_{1s} = \alpha_s$. The last equality shows that the expectation is equal to zero only if the true conditional mean of Y_{is} is equal to $\beta_{0s} + \mathrm{E}\left[X_i\right]' \beta_{1s}$, i.e. second term in the expectation is equal to zero. Otherwise the expectation would not be zero and the estimators would not be consistent.

We can now apply the similar arguments to show the double robustness of the weighted regression estimators. Consider weighted regression with the weighted objective function. In that case the first $K+1$ first order conditions yield the

following moment functions:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{D_{is}(T_i)}{r(s,X_i;\hat{\psi}_s)}\left(Y_i-\sum_{t=0}^{K}\mu_t D_{it}(T_i)-\sum_{t=0}^{K}D_{it}(T_i)[X_i-\mathrm{E}\left[X_i\right]]'\alpha_t\right)\quad\text{for}\quad s=0,\ldots,K.$$

The population counterpart of the above given moment function is given by

$$\begin{aligned}
&\mathrm{E}\left[\frac{D_{is}(T_i)}{r(s,X_i;\hat{\psi}_s)}\left(Y_i-\sum_{t=0}^{K}\mu_t D_{it}(T_i)-\sum_{t=0}^{K}D_{it}(T_i)[X_i-\mathrm{E}\left[X_i\right]]'\alpha_t\right)\right]\quad\text{for}\quad s=0,\ldots,K\\
&=\mathrm{E}\left[\mathrm{E}\left[\frac{D_{is}(T_i)}{r(s,X_i;\hat{\psi}_s)}\left(Y_i-\sum_{t=0}^{K}\mu_t D_{it}(T_i)-\sum_{t=0}^{K}D_{it}(T_i)[X_i-\mathrm{E}\left[X_i\right]]'\alpha_t\right)\middle|\, X_i\right]\right]\\
&=\mathrm{E}\left[\mathrm{E}\left[\frac{D_{is}(T_i)}{r(s,X_i;\hat{\psi}_s)}Y_{is}-\mu_s\frac{D_{is}(T_i)}{r(s,X_i;\hat{\psi}_s)}-\frac{D_{is}(T_i)}{r(s,X_i;\hat{\psi}_s)}[X_i-\mathrm{E}\left[X_i\right]]'\alpha_s)\middle|\, X_i\right]\right]\\
&=\mathrm{E}\left[\frac{\mathrm{E}\left[D_{is}(T_i)|\, X_i\right]}{r(s,X_i;\hat{\psi}_s)}\mathrm{E}\left[(Y_{is}-\mu_s-[X_i-\mathrm{E}\left[X_i\right]]'\beta_{1s})|\, X_i\right]\right]\\
&=\mathrm{E}\left[\frac{\mathrm{E}\left[D_{is}(T_i)|\, X_i\right]}{r(s,X_i;\hat{\psi}_s)}\left[\mathrm{E}\left[Y_{is}|\, X_i\right]-(\beta_{0s}+\mathrm{E}\left[X_i\right]'\beta_{1s})\right]\right]
\end{aligned}$$

The derivation follows similar steps as in unweighted regression estimation. The double robustness property can be seen from the last equality. If the conditional mean for Y_{is} is correctly specified, the second term in the expectation will be equal to zero, thus the whole expression will be equal to zero even with a wrong specified GPS model. Moreover, if the GPS model $r(s,X_i;\psi_s)$ is a correct specification for the conditional mean of $D_{is}(T_i)$ the first term in the expectation will be equal to one. In that case, due to properties of the linear model the whole expectation will be equal to zero even if the conditional mean of Y_{is} is not correctly specified.

B Tables: Monte Carlo Evidence

Table B.1: Monte Carlo Results: Correct specifications

N=500	WEIGHTING			REGRESSION			DOUBLY ROBUST		
	BIAS	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**
μ_0	-0.002	0.197	0.197	0.001	0.179	0.179	0.000	0.190	0.190
μ_1	0.001	0.097	0.097	0.001	0.095	0.095	0.001	0.096	0.096
μ_2	0.000	0.059	0.059	0.000	0.058	0.058	0.000	0.058	0.058
τ^{01}	-0.004	0.216	0.216	0.000	0.200	0.200	-0.001	0.210	0.210
τ^{02}	-0.003	0.207	0.207	0.000	0.189	0.189	0.000	0.199	0.199
τ^{12}	0.001	0.113	0.113	0.000	0.110	0.110	0.001	0.111	0.111
$\gamma^{01\|0}$	0.002	0.171	0.171	0.003	0.169	0.169	0.002	0.171	0.171
$\gamma^{02\|0}$	0.000	0.165	0.165	0.000	0.157	0.157	0.000	0.161	0.161
$\gamma^{10\|1}$	0.000	0.183	0.183	0.000	0.181	0.181	0.000	0.184	0.184
$\gamma^{12\|1}$	-0.001	0.111	0.111	-0.001	0.110	0.110	-0.001	0.110	0.110
$\gamma^{20\|2}$	0.004	0.241	0.241	-0.001	0.210	0.210	0.000	0.230	0.230
$\gamma^{21\|2}$	-0.002	0.120	0.120	-0.002	0.115	0.115	-0.002	0.117	0.117

N=2000	WEIGHTING			REGRESSION			DOUBLY ROBUST		
	BIAS	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**
μ_0	0.000	0.094	0.094	0.000	0.085	0.085	0.000	0.091	0.091
μ_1	0.001	0.049	0.049	0.001	0.048	0.048	0.001	0.048	0.048
μ_2	0.000	0.029	0.029	0.000	0.029	0.029	0.000	0.029	0.029
τ^{01}	-0.001	0.106	0.106	-0.001	0.097	0.097	0.000	0.103	0.103
τ^{02}	0.000	0.098	0.098	0.000	0.090	0.090	0.000	0.095	0.095
τ^{12}	0.001	0.057	0.057	0.001	0.056	0.056	0.001	0.056	0.056
$\gamma^{01\|0}$	0.000	0.084	0.084	0.000	0.084	0.084	0.000	0.084	0.084
$\gamma^{02\|0}$	0.000	0.080	0.080	0.000	0.076	0.076	0.000	0.078	0.078
$\gamma^{10\|1}$	0.000	0.089	0.089	0.000	0.088	0.088	0.000	0.089	0.089
$\gamma^{12\|1}$	0.000	0.057	0.057	0.000	0.056	0.056	0.000	0.056	0.056
$\gamma^{20\|2}$	0.000	0.114	0.114	0.000	0.100	0.100	-0.001	0.109	0.109
$\gamma^{21\|2}$	-0.002	0.060	0.060	-0.002	0.058	0.058	-0.002	0.058	0.058

Table B.2: Monte Carlo Results: Wrong Outcome Model

N=500	WEIGHTING			REGRESSION			DOUBLY ROBUST		
	BIAS	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**
μ_0	-0.002	0.197	0.197	-0.059	0.166	0.176	-0.006	0.187	0.187
μ_1	0.001	0.097	0.097	-0.021	0.095	0.098	0.001	0.096	0.096
μ_2	0.000	0.059	0.059	0.019	0.058	0.061	0.000	0.058	0.058
τ^{01}	-0.004	0.216	0.216	-0.038	0.190	0.193	-0.007	0.207	0.207
τ^{02}	-0.003	0.207	0.207	-0.078	0.175	0.192	-0.006	0.197	0.197
τ^{12}	0.001	0.113	0.113	-0.040	0.110	0.117	0.000	0.111	0.111
$\gamma^{01\|0}$	0.002	0.171	0.171	-0.036	0.168	0.172	0.002	0.171	0.171
$\gamma^{02\|0}$	0.000	0.165	0.165	-0.079	0.154	0.174	-0.001	0.160	0.160
$\gamma^{10\|1}$	0.000	0.183	0.183	0.038	0.177	0.182	0.002	0.183	0.183
$\gamma^{12\|1}$	-0.001	0.111	0.111	-0.042	0.110	0.118	-0.001	0.110	0.110
$\gamma^{20\|2}$	0.004	0.241	0.241	0.077	0.190	0.205	0.009	0.224	0.224
$\gamma^{21\|2}$	-0.002	0.120	0.120	0.039	0.114	0.121	-0.002	0.117	0.117

N=2000	WEIGHTING			REGRESSION			DOUBLY ROBUST		
	BIAS	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**
μ_0	0.000	0.094	0.094	-0.060	0.079	0.099	-0.001	0.091	0.091
μ_1	0.001	0.049	0.049	-0.020	0.048	0.052	0.001	0.048	0.048
μ_2	0.000	0.029	0.029	0.018	0.028	0.034	0.000	0.029	0.029
τ^{01}	-0.001	0.106	0.106	-0.040	0.092	0.100	-0.002	0.103	0.103
τ^{02}	0.000	0.098	0.098	-0.079	0.084	0.115	-0.001	0.095	0.095
τ^{12}	0.001	0.057	0.057	-0.039	0.056	0.068	0.001	0.056	0.056
$\gamma^{01\|0}$	0.000	0.084	0.084	-0.038	0.084	0.093	0.000	0.084	0.084
$\gamma^{02\|0}$	0.000	0.080	0.080	-0.079	0.075	0.109	0.000	0.078	0.078
$\gamma^{10\|1}$	0.000	0.089	0.089	0.039	0.086	0.095	0.001	0.089	0.089
$\gamma^{12\|1}$	0.000	0.057	0.057	-0.041	0.056	0.069	0.000	0.056	0.056
$\gamma^{20\|2}$	0.000	0.114	0.114	0.079	0.090	0.120	0.001	0.109	0.109
$\gamma^{21\|2}$	-0.002	0.060	0.060	0.038	0.057	0.069	-0.002	0.058	0.058

Table B.3: Monte Carlo Results: Wrong GPS Model

N=500	WEIGHTING			REGRESSION			DOUBLY ROBUST		
	BIAS	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**
μ_0	-0.061	0.170	0.181	0.001	0.179	0.179	0.000	0.187	0.187
μ_1	-0.021	0.096	0.098	0.001	0.095	0.095	0.001	0.096	0.096
μ_2	0.019	0.058	0.061	0.000	0.058	0.058	0.000	0.058	0.058
τ^{01}	-0.040	0.193	0.197	0.000	0.200	0.200	-0.001	0.207	0.207
τ^{02}	-0.081	0.180	0.198	0.000	0.189	0.189	0.000	0.196	0.196
τ^{12}	-0.041	0.111	0.118	0.000	0.110	0.110	0.000	0.110	0.110
$\gamma^{01\|0}$	-0.036	0.168	0.172	0.003	0.169	0.169	0.002	0.170	0.170
$\gamma^{02\|0}$	-0.081	0.158	0.177	0.000	0.157	0.157	0.000	0.160	0.160
$\gamma^{10\|1}$	0.039	0.177	0.182	0.000	0.181	0.181	0.000	0.183	0.183
$\gamma^{12\|1}$	-0.042	0.111	0.118	-0.001	0.110	0.110	-0.001	0.110	0.110
$\gamma^{20\|2}$	0.080	0.198	0.214	-0.001	0.210	0.210	0.000	0.228	0.228
$\gamma^{21\|2}$	0.039	0.115	0.122	-0.002	0.115	0.115	-0.002	0.117	0.117

N=2000	WEIGHTING			REGRESSION			DOUBLY ROBUST		
	BIAS	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**	**BIAS**	**SE**	**RMSE**
μ_0	-0.060	0.081	0.102	0.000	0.085	0.085	0.000	0.089	0.089
μ_1	-0.020	0.048	0.053	0.001	0.048	0.048	0.001	0.048	0.048
μ_2	0.018	0.028	0.034	0.000	0.029	0.029	0.000	0.029	0.029
τ^{01}	-0.039	0.094	0.103	-0.001	0.097	0.097	-0.001	0.101	0.101
τ^{02}	-0.079	0.086	0.117	0.000	0.090	0.090	0.000	0.093	0.093
τ^{12}	-0.039	0.056	0.069	0.001	0.056	0.056	0.001	0.056	0.056
$\gamma^{01\|0}$	-0.039	0.084	0.093	0.000	0.084	0.084	0.000	0.084	0.084
$\gamma^{02\|0}$	-0.080	0.077	0.111	0.000	0.076	0.076	0.000	0.078	0.078
$\gamma^{10\|1}$	0.039	0.086	0.095	0.000	0.088	0.088	0.000	0.089	0.089
$\gamma^{12\|1}$	-0.041	0.056	0.070	0.000	0.056	0.056	0.000	0.056	0.056
$\gamma^{20\|2}$	0.079	0.095	0.123	0.000	0.100	0.100	0.000	0.108	0.108
$\gamma^{21\|2}$	0.038	0.058	0.070	-0.002	0.058	0.058	-0.002	0.058	0.058

C Tables: Empirical Study

Table C.1: Noncognitive Ability Scales

LAWSEQ Self-Esteem Scale	
1	Do you think that your parent usually like to hear about your new ideas?
2	Do you often feel lonely at school?
3	Do other children often break friends or fall out with you?
4	Do you like team games?*
5	Do you think that other children often say nasty things about you?
6	When you have to say things in front of teachers, do you usually feel shy?
7	Do you like writing stories or doing creative writing?*
8	Do you often feel sad because you have nobody to play with at school?
9	Are you good at mathematics?*
10	Are there lots of things about yourself you would like to change?
11	When you have to say things in front of other children, do you usually feel foolish?
12	Do you find it difficult to do things like woodwork or knitting?*
13	When you want to tell a teacher something do you usually feel foolish?
14	Do you often have to find new friends because your old friends are playing with somebody else?
15	Do you usually feel foolish when you talk to your parents?
16	Do other people often think that you tell lies?
Note: Score +2 for all numbers answering "no" except for question 1. Score +2 for question 1 answering "yes". 4,7,9 and 12 do not count. Score +1 for all answers "don't know". High scores indicate higher self-esteem.	

CARALOC Locus of Control Scale	
1	Do you feel that most of the time it is not worth trying hard because things never turn out right anyway?
2	Do you feel that wishing can make good things happen?
3	Are people good to you no matter how you act towards them?
4	Do you like taking part in plays or concerts?*
5	Do you usually feel that it is almost useless to try in school because most children are cleverer than you?
6	Is a high mark just a matter of luck for you?
7	Are you good at spelling?*
8	Are tests just a lot of guess work for you?
9	Are you often blamed for things which just aren't your fault?
10	Are you the kind of person who believes that planning ahead makes things turn out better?
11	Do you find it easy to get up in the morning?*
12	When bad things happen to you, is it usually someone else's fault?
13	When someone is very angry with you, is it impossible to make him your friend again?
14	When nice things happen to you is it only good luck?
15	Do you feel sad when it is time to leave school each day?*
16	When you get into an argument is it usually the other person's fault?
17	Are you surprised when your teacher says you've done well?
18	Do you usually get low marks, even when you study hard?
19	Do you like to read books?*
20	Do you think studying for tests is a waste of time
Note: Each "No" response counts as one point, except for the question ten where "Yes" equals one point. Questions 4, 7, 11, 15 and 19 do not count.	

Table C.2: Behavior Difficulties Scales

Rutter Parental 'A' Scale of Behavior Disorder	
1	Very restless. Often running or jumping up and down. Hardly ever still.
2	Is squirmy or fidgety.
3	Often destroys own or others' belongings.
4	Frequently fights with other children.
5	Not much liked by other children.
6	Often worried, worries about many things.
7	Tends to do things on his/her own, rather solitary.
8	Irritable. Is quick to 'fly off the handle'.
9	Often appears miserable, unhappy, tearful or distressed.
10	Sometimes takes things belonging to others.
11	Has twitches, mannerisms or tics of the face or body.
12	Frequently sucks thumb or finger.
13	Frequently bites nails or fingers.
14	Is often disobedient.
15	Cannot settle to do anything for more than a few moments.
16	Tends to be fearful or afraid of new things or new situation.
17	Is fussy or over-particular.
18	Often tells lies.
19	Bullies other children.
Conner's Hyperactivity Scale	
1	Is noticeably clumsy.
2	Trips or falls easily or bumps into objects or other children.
3	Inattentive, easily distracted.
4	Hums or makes other odd noises at inappropriate times.
5	Has difficulty picking up small objects.
6	Drops things which are being carried.
7	Becomes obsessional about unimportant things.
8	Requests must be met immediately, easily frustrated.
9	Shows restless or over-active behavior.
10	Is impulsive, excitable.
11	Interferes with the activity of other children.
12	Is sullen or sulky.
13	Fails to finish things he/she starts, short attention span.
14	Given to rhythmic tapping or kicking.
15	Cries for little cause.
16	Changes mood quickly and drastically.
17	Displays outbursts of temper, explosive or unpredictable behavior.
18	Has difficulty using scissors.
19	Has difficulty concentrating on any particular task though may return to it frequently.

Note: Mothers had to make a vertical mark through the line alongside each statement to indicate to what extent the child shows the behavior described. The line correspond to a scale from 0 to 100. 0 refers to "does not apply" and 100 refers to "certainly applies".

Table C.3: Variable Descriptions Source: BCS, own definitions

Variable	Description
lhnet	log hourly wage at age of 30
female	Dummy, =1 if Female
edulev	=0 if no education, =1 if O-level, =2 if A-level, =3 if higher education
motage	age of mother at the birth of cohort member
motedu0	Dummy, =1 if mother continued education beyond the minimum school leaving age
fatsoc10d	Dummy, =1 if father social status is classified as professional or intermediate or skilled nonmanual or skilled manual
brok	Dummy, =1 if the child did not live with both parents since birth till age of 10
nuchild10	number of children in household at age of 10
ethnic	Dummy, =1 if English
fmtsc	Sum of the correctly answered questions in Friendly Math Test
sertsc	Sum of the correctly answered questions in Shortened Edinburgh Reading Test
baswssc	Sum of the correctly answered questions in Word Similarities part of the BAS
baswdsc	Sum of the correctly answered questions in Word Definition part of the BAS
basrdsc	Sum of the correctly answered questions in Recall Digits part of the BAS
basmsc	Sum of the correctly answered questions in Word Similarities part of the BAS
carloc	Carloc locus of control score
lawseq	Lawreq self-esteem score
hyper1	Dummy, =1 if Normal behavior according to Conner's Hyperactivity Scale
hyper2	Dummy, =1 if Moderate behavior problems according to Conner's Hyperactivity Scale
hyper3	Dummy, =1 if Severe behavior problems according to Conner's Hyperactivity Scale
rutt1	Dummy, =1 if Normal behavior according to Rutter's Behavior Disorder Scale
rutt2	Dummy, =1 if Moderate behavior problems according to Rutter's Behavior Disorder Scale
rutt3	Dummy, =1 if Severe behavior problems according to Rutter's Behavior Disorder Scale
inc1	Dummy, =1 if income per week is under 35 £
inc2	Dummy, =1 if income per week is between 35-49 £
inc3	Dummy, =1 if income per week is between 50-99 £
inc4	Dummy, =1 if income per week is between 100-149 £
inc5	Dummy, =1 if income per week is between 150-199 £
inc6	Dummy, =1 if income per week is between 200-249 £
inc7	Dummy, =1 if income per week is above 249 £

Table C.4: Summary statistics Data Source: BCS, own calculations

	Entire Sample		Female		Male	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
lhnet	1.78	0.44	1.72	0.42	1.84	0.45
female	0.48	0.5	1	0	0	0
dseq	1.71	1.13	1.69	1.15	1.73	1.12
motage	25.9	5.2	25.9	5.22	25.91	5.19
motedu0	0.37	0.48	0.37	0.48	0.36	0.48
fatsoc10d	0.38	0.49	0.38	0.49	0.38	0.49
brok	0.1	0.3	0.11	0.31	0.09	0.29
nuchild10	2.47	0.96	2.44	0.96	2.49	0.96
etnic	0.98	0.14	0.98	0.15	0.98	0.14
fmtsc	45.85	11.59	45.32	10.88	46.35	12.19
sertsc	33.98	10.68	34.89	9.98	33.14	11.22
baswssc	12.31	2.77	12.18	2.55	12.43	2.96
baswdsc	10.81	5.10	10.3	4.81	11.28	5.31
basrdsc	22.46	4.69	22.69	4.38	22.25	4.96
basmsc	15.94	5.47	16.26	5.33	15.64	5.58
carloc	7.41	2.91	7.23	2.94	7.57	2.88
lawseq	16.46	4.58	15.99	4.75	16.89	4.37
hyper1	0.82	0.38	0.85	0.35	0.8	0.4
hyper2	0.14	0.35	0.12	0.32	0.16	0.37
hyper3	0.04	0.19	0.03	0.17	0.04	0.21
rutt1	0.83	0.38	0.85	0.35	0.8	0.4
rutt2	0.14	0.35	0.11	0.32	0.16	0.37
rutt3	0.03	0.18	0.03	0.18	0.04	0.19
inc1	0.01	0.11	0.01	0.11	0.01	0.11
inc2	0.03	0.18	0.03	0.18	0.03	0.18
inc3	0.27	0.44	0.26	0.44	0.27	0.44
inc4	0.38	0.49	0.38	0.48	0.38	0.49
inc5	0.18	0.38	0.18	0.38	0.19	0.39
inc6	0.07	0.25	0.07	0.26	0.06	0.24
inc7	0.06	0.24	0.06	0.24	0.06	0.23
N	4697		2266		2431	

Table C.5: Generalized Propensity Score Estimation Results

	Female		**Male**	
lawseq	-0.004	(0.01)	-0.007	(0.01)
caraloc	0.054***	(0.02)	0.049***	(0.02)
hyper2	-0.251*	(0.14)	-0.139	(0.12)
hyper3	-0.145	(0.28)	-0.118	(0.23)
rutt2	0.024	(0.14)	-0.095	(0.12)
rutt3	-0.359	(0.27)	-0.041	(0.25)
sertsc	0.036***	(0.01)	0.023***	(0.01)
fmtsc	0.013**	(0.01)	0.026***	(0.01)
nuchild10	-0.111**	(0.04)	-0.117***	(0.04)
inc1	-0.351	(0.38)	0.398	(0.36)
inc2	-0.419**	(0.24)	-0.158	(0.22)
inc3	-0.131	(0.10)	-0.066	(0.10)
inc5	0.029	(0.12)	0.289***	(0.11)
inc6	0.072	(0.17)	0.098	(0.18)
inc7	0.217	(0.19)	0.281	(0.19)
brok	-0.072	(0.14)	-0.118	(0.14)
motedu0	0.294***	(0.09)	0.403***	(0.09)
motage	0.020***	(0.01)	0.004	(0.01)
etnic	-1.389***	(0.31)	-1.002***	(0.30)
fatsoc10d	0.541***	(0.09)	0.340***	(0.09)
baswdsc	0.023**	(0.01)	0.033***	(0.01)
baswssc	0.048**	(0.02)	0.016	(0.02)
basrdsc	0.002	(0.01)	-0.011	(0.01)
basmsc	0.013	(0.01)	0.018	(0.01)
threshold1	0.647	(0.48)	0.112	(0.46)
threshold2	2.315***	(0.48)	1.547***	(0.46)
threshold3	3.032***	(0.48)	2.700***	(0.47)
N	2266		2431	
LR $\chi^2(24)$	598.59		667.68	
p-value	0.00		0.00	

Note: *** 1% significance level, ** 5% significance level,

* 10% Significance level.

Table C.6: Average Partial Effect after Ordered Logit

	Female			
	$\Pr[T=0\,\vert X]$	$\Pr[T=1\,\vert X]$	$\Pr[T=2\,\vert X]$	$\Pr[T=3\,\vert X]$
lawseq	0.0002	-0.0002	0.0000	0.0000
carloc	-0.0032***	0.0024***	0.0004***	0.0004***
hyper2	0.0161*	-0.0122*	-0.0019	-0.0020
hyper3	0.0092	-0.0069	-0.0011	-0.0011
rutt2	-0.0014	0.0011	0.0002	0.0002
rutt3	0.0243	-0.0183	-0.0029	-0.0031
sertsc	-0.0021***	0.0016***	0.0003***	0.0003***
fmtsc	-0.0008**	0.0006**	0.0001**	0.0001**
nuchild10	0.0067**	-0.0051**	-0.0008**	-0.0008**
inc1	0.0239	-0.0179	-0.0029	-0.0030
inc2	0.0290	-0.0218	-0.0035	-0.0037
inc3	0.0080	-0.0061	-0.0010	-0.0010
inc5	-0.0018	0.0013	0.0002	0.0002
inc6	-0.0042	0.0032	0.0005	0.0005
inc7	-0.0121	0.0092	0.0014	0.0014
brok	0.0044	-0.0034	-0.0005	-0.0005
motedu0	-0.0167***	0.0128***	0.0019***	0.0020***
motage	-0.0012**	0.0009**	0.0001**	0.0001**
etnic	0.0515***	-0.0400***	-0.0057**	-0.0058**
fatsoc10d	-0.0294***	0.0226***	0.0034***	0.0034***
baswdsc	-0.0014*	0.0010*	0.0002*	0.0002*
baswssc	-0.0029**	0.0022**	0.0003**	0.0003**
basrdsc	-0.0001	0.0001	0.0000	0.0000
basmsc	-0.0008	0.0006	0.0001	0.0001

	Male			
	$\Pr[T=0\,\vert X]$	$\Pr[T=1\,\vert X]$	$\Pr[T=2\,\vert X]$	$\Pr[T=3\,\vert X]$
lawseq	0.0008	-0.0005	-0.0002	-0.0001
carloc	-0.0056***	0.0036***	0.0013***	0.0008***
hyper2	0.0165	-0.0104	-0.0039	-0.0022
hyper3	0.0141	-0.0088	-0.0034	-0.0019
rutt2	0.0112	-0.0070	-0.0026	-0.0015
rutt3	0.0048	-0.0030	-0.0011	-0.0006
sertsc	-0.0026***	0.0017***	0.0006***	0.0004***
fmtsc	-0.0030***	0.0019***	0.0007***	0.0004***
nuchild10	0.0135***	-0.0085***	-0.0032***	-0.0018***
inc1	-0.0413	0.0268	0.0094	0.0052
inc2	0.0190	-0.0118	-0.0046	-0.0026
inc3	0.0077	-0.0048	-0.0018	-0.0010
inc5	-0.0317***	0.0204***	0.0073***	0.0040***
inc6	-0.0111	0.0070	0.0026	0.0015
inc7	-0.0302	0.0195	0.0069*	0.0038*
brok	0.0140	-0.0088	-0.0033	-0.0019
motedu0	-0.0446***	0.0288***	0.0102***	0.0056***
motage	-0.0004	0.0003	0.0001	0.0001
etnic	0.0878**	-0.0585***	-0.0191**	-0.0102**
fatsoc10d	-0.0380***	0.0245***	0.0087***	0.0048***
baswdsc	-0.0038***	0.0024***	0.0009***	0.0005***
baswssc	-0.0018	0.0012	0.0004	0.0002
basrdsc	0.0012	-0.0008	-0.0003	-0.0002
basmsc	-0.0020**	0.0013**	0.0005**	0.0003*

Note: *** 1% significance level, ** 5% significance level, * 10% Significance level.

D Figures

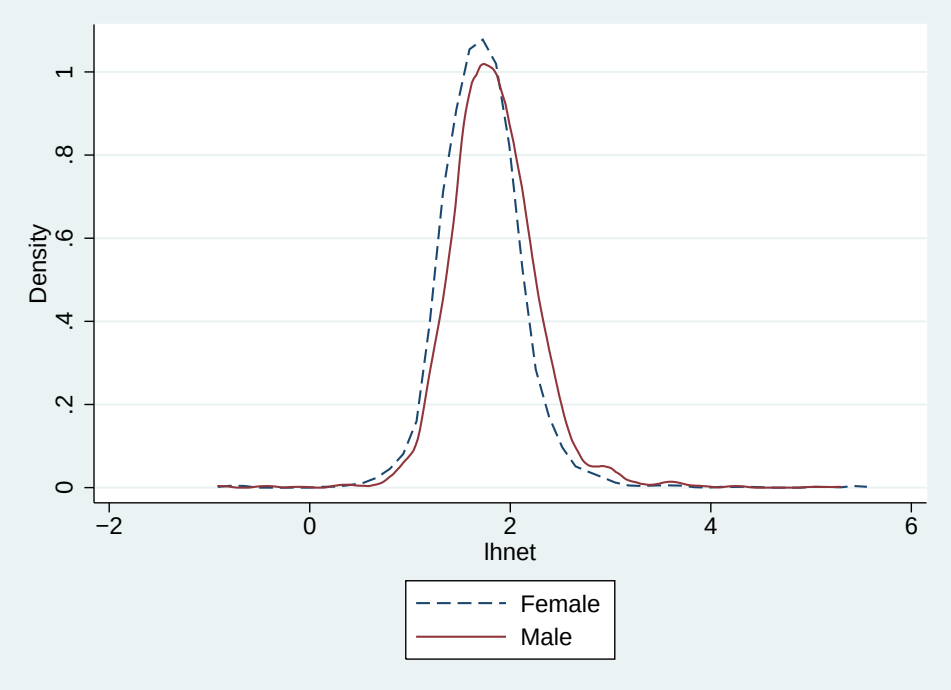

Figure D.1: Distribution of log hourly wages by gender (Epanechnikov Kernel is used and the bandwidth is chosen by Silverman's rule of thumb)

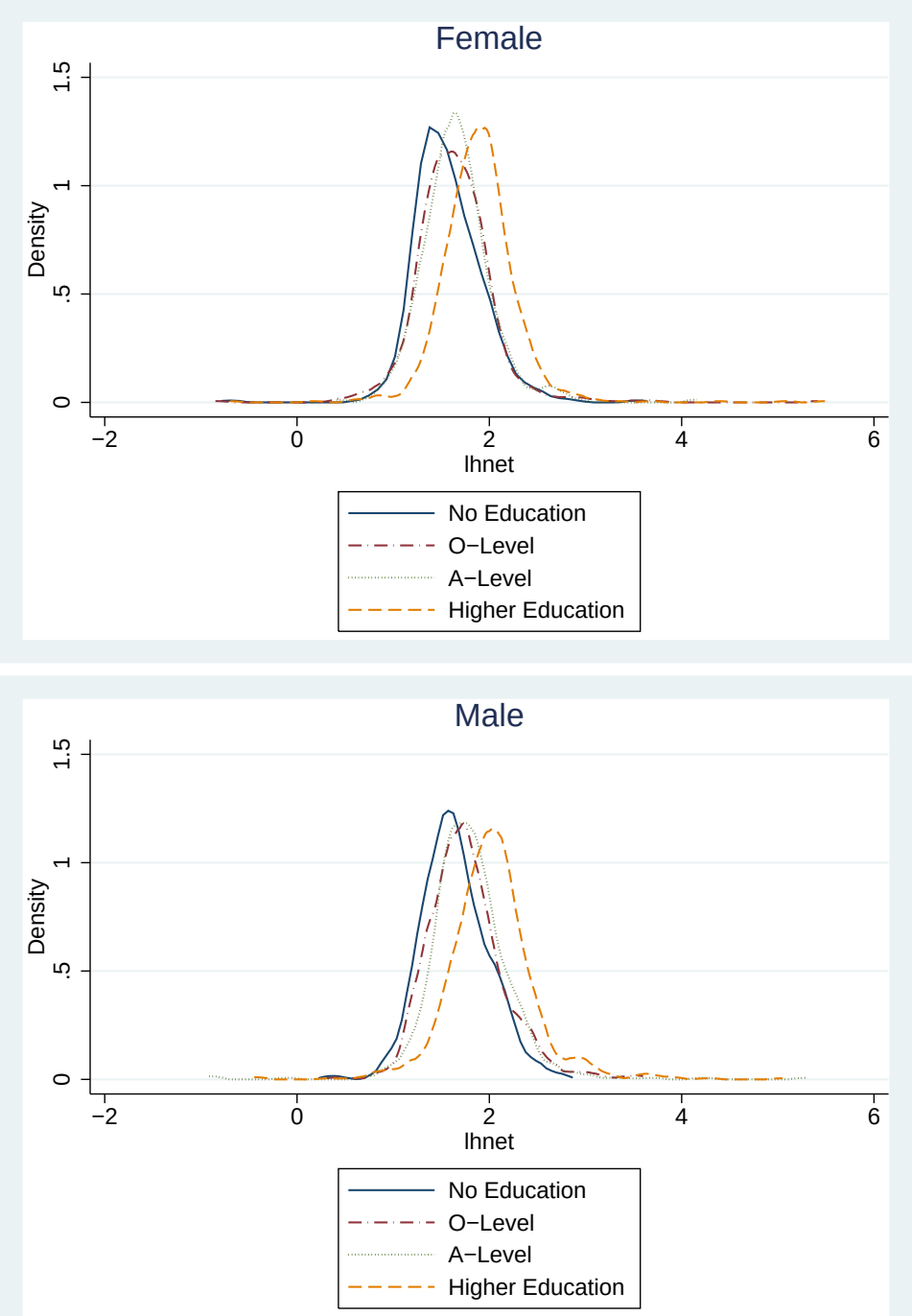

Figure D.2: Distribution of log hourly wages by education level (Epanechnikov Kernel is used and the bandwidth is chosen by Silverman's rule of thumb)

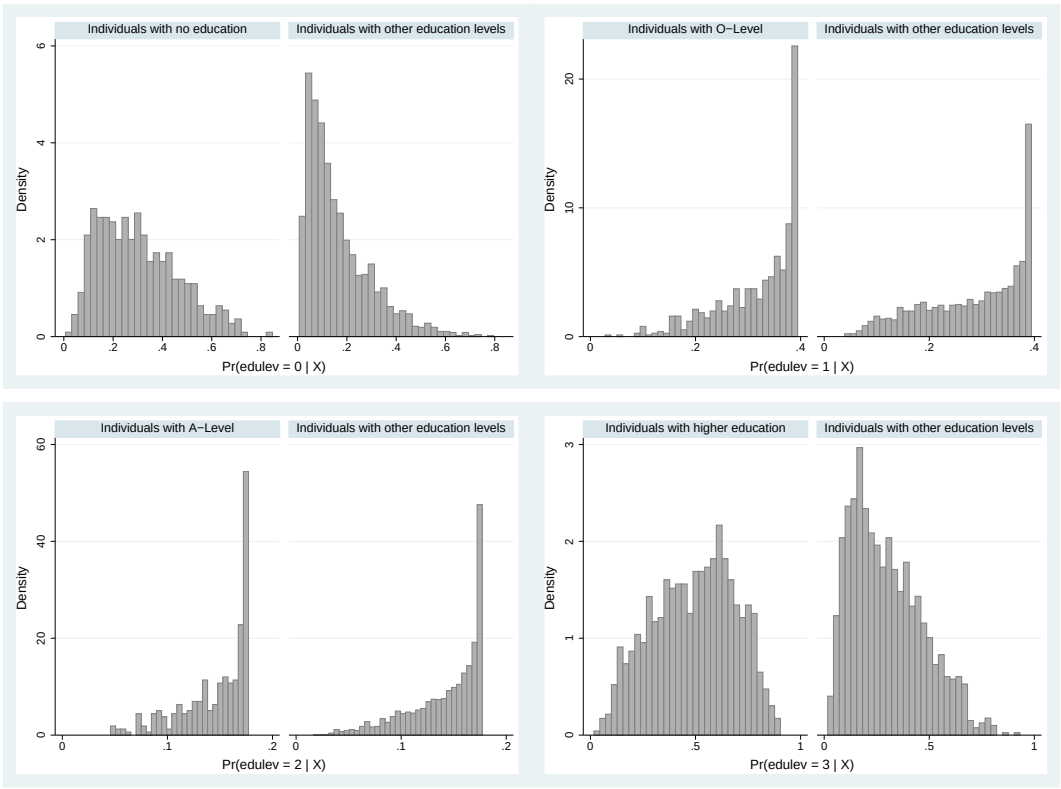

Figure D.3: Histogram estimates of the GPS for individuals with $T_i = t$ and $T_i \neq t$ for each $t = \{0, 1, 2, 3\}$ for females.

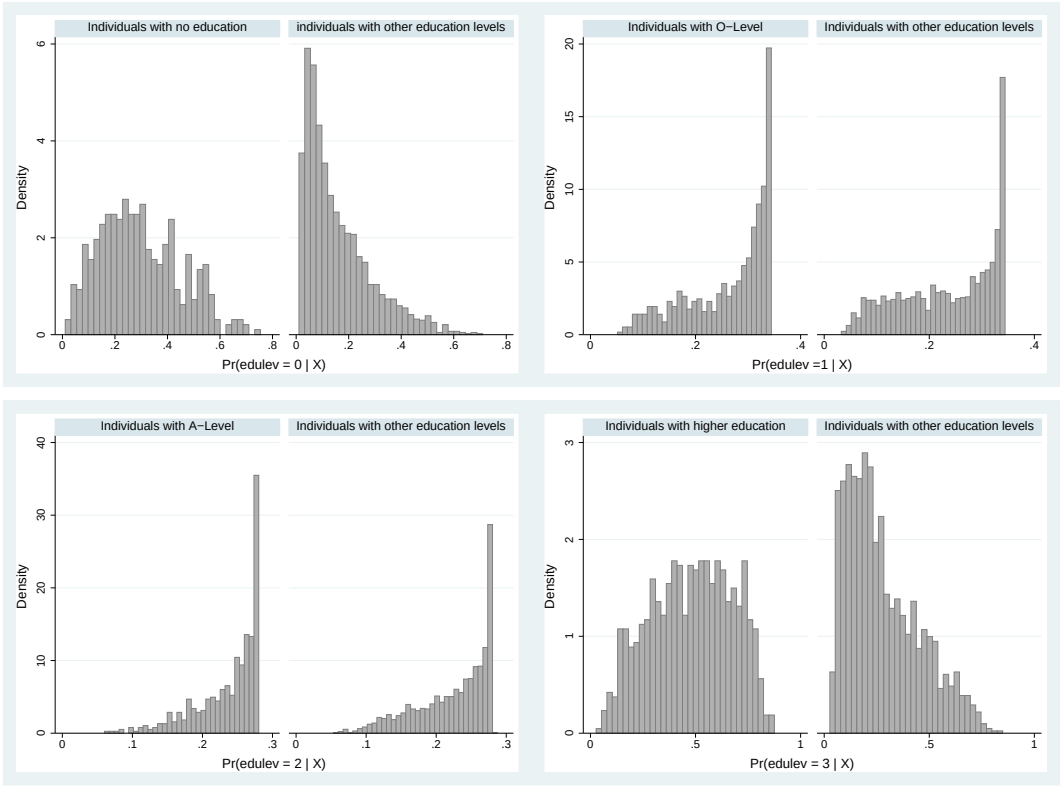

Figure D.4: Histogram estimates of the GPS for individuals with $T_i = t$ and $T_i \neq t$ for each $t = \{0, 1, 2, 3\}$ for male sample.

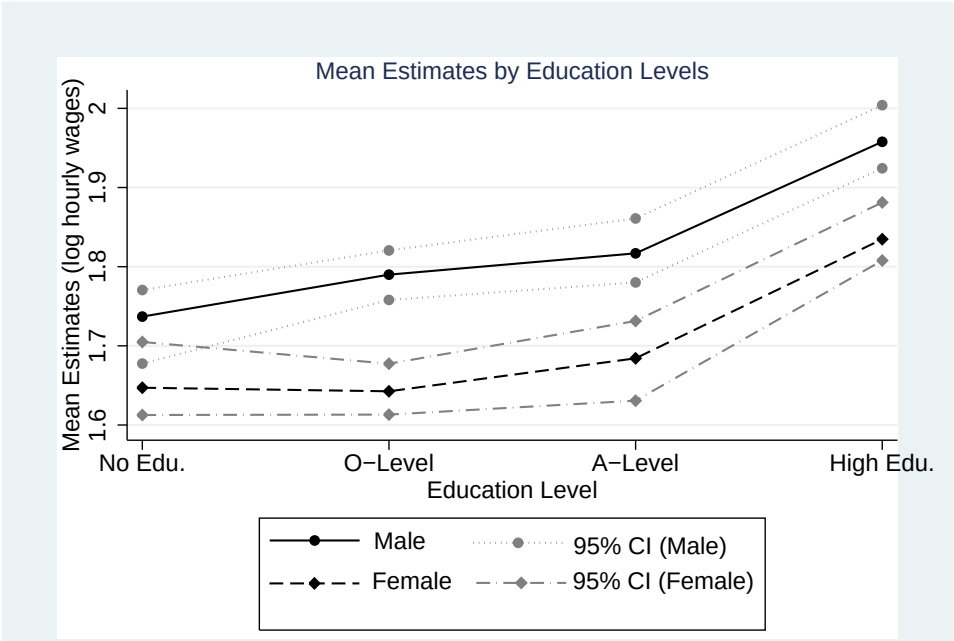

Figure D.5: Estimated mean log hourly wages by education level for female and male samples.

Author: S. Derya Uysal

Title: Doubly Robust Estimation of Causal Effects with Multivalued Treatments

Reihe Ökonomie / Economics Series 297

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Michael Reiter (Macroeconomics), Selver Derya Uysal (Microeconomics)

ISSN: 1605-7996
© 2013 by the Department of Economics and Finance, Institute for Advanced Studies (IHS),
Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996