Web Appendix
Doubly Robust Estimation of Causal Effects with Multivalued Treatments: An Application to the Returns to Schooling

S. Derya Uysal *
IHS, Vienna

This draft: January 7, 2014

*Institute for Advanced Studies, Department of Economics and Finance, Stumpergasse 56 1060 Vienna, Austria. Phone +43 1 59991 156. email: uysal@ihs.ac.at. Supported by funds of the Oesterreichische Nationalbank (Anniversary Fund, project number: 14986)

1 Additional Derivations

Derivation of Equation (2.8)

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

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

Equation (2.8) can be derived by combining (2.1) and WA.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.

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

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.

2 Asymptotic Distribution

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:

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

where $\widetilde{\varepsilon}_i$ 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.[1] Define $s_i = s(Y_i, X_i, T_i; \theta, \psi) \equiv \omega_i \frac{\partial \tilde{\varepsilon}_i^2}{\partial \theta}$ as the $P \times 1$ weighted score of the (unweighted) objective function $q(\cdot)$, $H(Y_i, X_i; \theta) = \frac{\partial^2 \tilde{\varepsilon}_i^2}{\partial \theta \partial \theta'}$ as the $P \times P$ Hessian of the objective function $q(\cdot)$. Under standard regularity conditions,

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, A^{-1} D A^{-1}\right), \tag{WA.2}$$

where $A \equiv \mathrm{E}\left[H(Y_i, X_i; \theta)\right]$, $D \equiv \mathrm{E}\left[e_i e_i'\right]$, $e_i \equiv s_i - \mathrm{E}\left[s_i c_i'\right]\left[\mathrm{E}\left[c_i c_i'\right]\right]^{-1} c_i$, $c_i \equiv c_i(\boldsymbol{\psi}) = [c_{1i}' \ldots c_{Ki}']'$ 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:

$$\hat{A} \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{\omega}_i H(Y_i, X_i; \hat{\theta}) \tag{WA.3}$$

and

$$\hat{D} \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{e}_i \hat{e}_i' \tag{WA.4}$$

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 $\hat{\boldsymbol{\psi}}$. Wooldridge (2007, p. 1298) states that $\hat{A}$ which is estimated as the weighted sample mean of Hessian is consistent for A even if the propensity score model is misspecified. $\hat{D}$ is a consistent estimator for D as long as outcome model or generalized probability model is correctly specified (Wooldridge, 2007, p. 1298). Since the treatment effects τ_{ml} and $\gamma_{ml|m}$ are estimated as differences of regression parameters (Equations (2.20)

[1] 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).

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 $\hat{\theta}$.

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.[2] 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 WA3.1: 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:

$$T_i = \arg\max_{t\in\mathfrak{T}} \{D_i^*(t)\} \quad \text{(WA.5)}$$

$$D_{it}(T_i) = \mathbb{1}\{T_i = t\} \quad \text{(WA.6)}$$

$$Y_i = \sum_{t=0}^{K=2} D_{it}(T_i)Y_{it}. \quad \text{(WA.7)}$$

[2] 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.

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 WA3.2: 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.

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:

$$r(t, x_i) \equiv \Pr[T_i = t \,|X_i] = \frac{\exp(\psi_{0t} + \psi_{1t}X_{i1} + \psi_{2t}X_{i2} + \psi_{3t}X_{i3})}{\sum_{j=0}^{2} \exp(\psi_{0j} + \psi_{1j}X_{i1} + \psi_{2j}X_{i2} + \psi_{3j}X_{i3})}, \quad \text{(WA.8)}$$

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

$$\mathrm{E}\left[Y_{it}|\, X_i\right] = \beta_{0t} + \beta_{1t} X_{i1} + \beta_{2t} X_{i2} + \beta_{3t} X_{i3}. \tag{WA.9}$$

The model specification given in Equation (WA.8) 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 (WA.9) 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 WA3.3: 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 3.4-3.6 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.

Table WA3.4: 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 WA3.5: 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 WA3.6: 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

This 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.

4 Additional Tables

Table WA4.7: Average Partial Effect after Ordered Logit

	$\Pr[T=0 \mid X]$		$\Pr[T=1 \mid X]$		$\Pr[T=2 \mid X]$		$\Pr[T=3 \mid X]$	
	Female	Male	Female	Male	Female	Male	Female	Male
motage	-0.003*	-0.000	-0.001*	-0.000	0.000*	0.000	0.004*	0.000
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
motedu0	-0.041‡	-0.052‡	-0.024‡	-0.032‡	0.003‡	0.004‡	0.062‡	0.081‡
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
fatss1	-0.066‡	-0.042†	-0.046‡	-0.026†	0.002‡	0.002†	0.110‡	0.067†
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
fatss3	0.035*	0.013	0.016*	0.007	-0.004*	-0.001	-0.047*	-0.018
	(0.11)	(0.10)	(0.11)	(0.10)	(0.11)	(0.10)	(0.11)	(0.10)
fatss4	-0.025	0.006	-0.014	0.003	0.002	-0.001	0.037	-0.008
	(0.18)	(0.19)	(0.18)	(0.19)	(0.18)	(0.19)	(0.18)	(0.19)
brok	0.017	0.012	0.008	0.006	-0.002	-0.001	-0.023	-0.017
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
nosib	-0.021	-0.007	-0.012	-0.004	0.002	0.001	0.031	0.010
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
twokid	0.008	0.005	0.004	0.003	-0.001	-0.001	-0.011	-0.007
	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)
more3kid	0.047*	0.049†	0.020*	0.022†	-0.006*	-0.008†	-0.062*	-0.063†
	(0.14)	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)
inc1	0.073	-0.022	0.026	-0.013	-0.010	0.001	-0.088	0.033
	(0.40)	(0.36)	(0.40)	(0.36)	(0.40)	(0.36)	(0.40)	(0.36)
inc2	0.070*	0.043	0.026*	0.019	-0.010*	-0.007	-0.086*	-0.055
	(0.24)	(0.23)	(0.24)	(0.23)	(0.24)	(0.23)	(0.24)	(0.23)
inc3	0.024	0.012	0.012	0.006	-0.002	-0.001	-0.033	-0.017
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
inc5	-0.007	-0.035*	-0.004	-0.021*	0.001	0.002*	0.011	0.054*
	(0.12)	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)
inc6	-0.022	-0.016	-0.013	-0.009	0.002	0.001	0.033	0.024
	(0.17)	(0.18)	(0.17)	(0.18)	(0.17)	(0.18)	(0.17)	(0.18)
inc7	-0.031	-0.037	-0.019	-0.023	0.002	0.001	0.048	0.058
	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)	(0.19)
fmtsc	-0.002*	-0.003‡	-0.001*	-0.002‡	0.000*	0.000‡	0.003*	0.005‡
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
sertsc	-0.005‡	-0.003‡	-0.003‡	-0.002‡	0.000‡	0.000‡	0.007‡	0.005‡
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
baswssc	-0.007*	-0.002	-0.004*	-0.001	0.001*	0.000	0.010*	0.003
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
baswdsc	-0.003	-0.004†	-0.001	-0.002†	0.000	0.000†	0.004	0.006†
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
basrdsc	0.000	0.001	0.000	0.001	-0.000	-0.000	-0.000	-0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
basmsc	-0.002	-0.002*	-0.001	-0.001*	0.000	0.000*	0.002	0.004*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
carloc	-0.007†	-0.006†	-0.004†	-0.003†	0.001†	0.001†	0.010†	0.009†
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
lawseq	0.001	0.001	0.000	0.001	-0.000	-0.000	-0.001	-0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
rutt2	-0.004	0.011	-0.002	0.006	0.000	-0.001	0.006	-0.016
	(0.14)	(0.12)	(0.14)	(0.12)	(0.14)	(0.12)	(0.14)	(0.12)
rutt3	0.064	0.004	0.025	0.002	-0.009	-0.000	-0.080	-0.005
	(0.27)	(0.25)	(0.27)	(0.25)	(0.27)	(0.25)	(0.27)	(0.25)
hyper2	0.041*	0.021	0.018*	0.010	-0.005*	-0.003	-0.054*	-0.029
	(0.14)	(0.12)	(0.14)	(0.12)	(0.14)	(0.12)	(0.14)	(0.12)
hyper3	0.020	0.014	0.009	0.007	-0.002	-0.002	-0.027	-0.019
	(0.28)	(0.23)	(0.28)	(0.23)	(0.28)	(0.23)	(0.28)	(0.23)
N	2208	2385	2208	2385	2208	2385	2208	2385

Data Source: BCS. ‡ 1% significance level, † 5% significance level, * 10% significance level. Standard errors are in parentheses.

References

Newey WK. 1985. Maximum likelihood specification testing and conditional moment tests. *Econometrica* **53**: 1047–1070. ISSN 00129682. URL `http://www.jstor.org/stable/1911011`

Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wooldridge JM. 2007. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* **141**: 1281 – 1301. ISSN 0304-4076. URL `http://www.sciencedirect.com/science/article/pii/S0304407607000437`