Doubly Robust Estimation of Local Average Treatment Effects Using Inverse Probability Weighted Regression Adjustment*

Tymon Słoczyński
Brandeis University

S. Derya Uysal
Ludwig Maximilian University of Munich

Jeffrey M. Wooldridge
Michigan State University

This version: November 14, 2022

Abstract: We revisit the problem of estimating the local average treatment effect (LATE) and the local average treatment effect on the treated (LATT) when control variables are available, either to render the instrumental variable (IV) suitably exogenous or to improve precision. Unlike previous approaches, our doubly robust (DR) estimation procedures use quasi-likelihood methods weighted by the inverse of the IV propensity score – so-called inverse probability weighted regression adjustment (IPWRA) estimators. By properly choosing models for the propensity score and outcome models, fitted values are ensured to be in the logical range determined by the response variable, producing DR estimators of LATE and LATT with appealing small sample properties. Inference is relatively straightforward both analytically and using the nonparametric bootstrap. Our DR LATE and DR LATT estimators work well in simulations. We also propose a DR version of the Hausman test that can be used to assess the unconfoundedness assumption through a comparison of different estimates of the average treatment effect on the treated (ATT) under one-sided noncompliance. Unlike the usual test that compares OLS and IV estimates, this procedure is robust to treatment effect heterogeneity.

*This article supersedes "Doubly Robust IV Estimation of Local Average Treatment Effects" by S. Derya Uysal.

1 Introduction

Instrumental variables estimation of causal effects has a long history in applied econometrics. In introductory econometrics courses, the properties of the instrumental variables (IV) estimator are often taught within the framework of a linear model with a constant coefficient. When applied in a treatment effects setting, the constant coefficient assumption is equivalent to assuming a constant treatment effect in the population. In their pioneering work, Imbens and Angrist (1994) [IA (1994)] used a potential outcomes framework to study the probability limit of the simple IV estimator in the setting of a binary treatment and binary instrumental variable. IA (1994) showed that, under reasonable assumptions, the IV estimator consistently estimates a parameter now known as the *local average treatment effect (LATE)*, which is the average treatment effect over the subpopulation of units that comply with their randomized eligibility. Angrist, Imbens and Rubin (1996) explicitly embedded the LATE setup within the setting of the Rubin Causal Model, showed how the IV estimand identifies a causal parameter under certain assumptions, and discussed the consequences of violations of those assumptions. Vytlacil (2002) demonstrated that the LATE framework is equivalent to a nonparametric selection model with a weakly separable selection equation.

In many applications of IV, the instrument is not randomly assigned, in which case the simple IV estimator – also known as the *Wald estimator* – is no longer consistent for LATE. In some cases, conditioning on observed covariates, or controls, can render the IV as good as randomly assigned within subpopulations defined by the covariate values. In effect, the IV is assumed to satisfy an unconfoundedness assumption conditional on observables. In most textbook treatments of instrumental variables that include control variables X, these are added linearly and then they act as their own instruments. In the treatment effects context, it may seem appealing to include interactions of the binary treatment variable, W, with the (suitably centered) covariates X. If Z is the instrument for W, a natural instrument for $W \cdot X$ is $Z \cdot X$. Wooldridge (2010, Procedure 21.2) describes an IV procedure that exploits the binary nature of W by using a binary response model for $p(Z, X) \equiv \mathbb{P}(W = 1|Z, X)$ and then using $\hat{p}(Z_i, X_i)$ and $\hat{p}(Z_i, X_i) \cdot X_i$ as instruments in the linear equation that includes W_i and $W_i \cdot X_i$.

Adding covariates to a linear equation and interacting them with the treatment indicator seems like a natural way to account for nonrandom assignment of the instrument while allowing for heterogeneous treatment effects. Unfortunately, no results imply that this procedure generally uncovers the LATE. By contrast, Tan (2006) and Frölich (2007) independently obtained a useful identification result for LATE when covariates are needed in order to render the IVs ignorable. Frölich (2007) used his identification result to obtain consistent, asymptotically normal estimators of LATE. As a practical matter, however, the need to estimate four conditional mean functions nonparametrically makes Frölich's estimator difficult to implement, even with just a small number of covariates. Plus, issues of how to handle discrete, continuous, and mixed control variables need to be addressed.

On the other hand, the estimation approach proposed by Tan (2006) is based on so-called *augmented inverse probability weighting (AIPW)* estimators. AIPW is a standard class of doubly robust (DR) estimators, that is, estimators that remain consistent under misspecification of either of the two (sets of) parametric working models on which they are based. However, as discussed in Kang and Schafer (2007), AIPW estimators, as commonly applied, are often unstable in practice, as is standard *inverse probability weighting (IPW)*. One reason is that these estimators are often based on weights in the weighted averages that do not sum to unity; in other words, the weights are not normalized.

Following Tan (2006) and Frölich (2007), many other estimation approaches for LATE have been proposed, some of which are doubly robust and some are not. For example, Donald, Hsu and Lieli (2014) [DHL (2014)] studied the aforementioned IPW estimators, which are consistent when the instrument propensity score is correctly specified but not otherwise. Admittedly, DHL (2014) suggested nonparametric series estimators, which in theory resolves the issues of misspecification, but in practice their approach would be applied in a flexible parametric framework by most practitioners. Similar to Tan (2006), other DR estimators have also been based on the AIPW approach, and this includes both those proposed in Ogburn, Rotnitzky and Robins (2015) and several estimators that employ high-dimensional selection, including those in Belloni et al. (2017),

Chernozhukov et al. (2018), and Sun and Tan (2022), which often also rely on sample splitting to allow for high-dimensional covariates. In recent work, Heiler (2022) discussed a DR extension of a particular balancing estimator of LATE while Singh and Sun (2022) combined "kappa weighting" (Abadie, 2003) and high-dimensional selection to obtain DR estimators of LATE and related parameters.

Our primary purpose in this paper is to propose a new class of doubly robust estimators of LATE that are simple to implement and avoid the shortcomings of nonparametric conditional mean estimation and AIPW methods. In particular, using the identification result in Frölich (2007) and building also on Wooldridge (2007) and Słoczyński and Wooldridge (2018), we show how estimators that use the inverse of the instrument propensity score to weight the objective functions for estimating the treatment propensity score and the conditional mean of the response allow consistent estimation of LATE. These estimators, now commonly labeled *inverse probability weighted regression adjustment (IPWRA)* estimators, have the same double robustness property of AIPW estimators. An advantage of IPWRA estimators is that one can choose functional forms so that the estimated conditional probability and conditional mean functions are guaranteed to produce predictions within the logical range of the outcomes. This feature of IPWRA makes the resulting estimators of LATE have good finite sample properties. Moreover, the estimators are easy to obtain and inference is relatively straightforward. Using a similar approach, we also propose DR estimators for the local average treatment effect on the treated (LATT).

In Section 2 we provide the setting, define the LATE parameter, and summarize the identification result in Frölich (2007). We also study identification of LATT, beginning with a result due to Frölich and Lechner (2010) but modifying it to obtain a simple representation of this parameter that leads naturally to DR estimation.

Section 3 shows how the IPWRA approach can be used to identify the four expectations appearing in LATE when conditioning on covariates. We modify existing arguments to account for the fact that the conditional means we need to estimate for the outcome are not of the potential outcomes. Nevertheless, the IPWRA approach still identifies the required unconditional means.

This section carries out a similar analysis for LATT where we are able to relax the assumptions used to identify LATE.

Section 4 shows how to obtain standard errors for the DR LATE and DR LATT estimators that account for the sampling error in all estimation steps. Section 5 shows how to modify the Hausman-type test proposed by DHL (2014) to allow for DR estimation. In the case with one-sided noncompliance, if assignment is unconfounded then LATT is the same as the average treatment effect on the treated (ATT). Therefore, we can obtain two DR estimators using IPWRA estimation schemes: one that uses an instrumental variable and another that employs unconfoundedness conditional on covariates. We show how to test the null hypothesis that the two estimators consistently estimate the same parameter.

Section 6 revisits two empirical studies. First, we use the data in Abadie (2003) to produce LATE and LATT estimates for the effect of participating in a 401(k) pension plan on net financial wealth. We also look at the causal effect of participating in a 401(k) plan on participation in individual retirement accounts (IRAs). In this case, we compare IV estimates of a linear probability model with our DR estimates that recognize the binary nature of the IRA participation decision. Using our proposed method, we find that 401(k) participation has a positive effect on net financial assets and the probability of IRA participation. This is despite the fact that the corresponding AIPW (for net financial assets) and IPW estimates (for both outcomes) are much smaller and imprecise. Even though both AIPW and IPWRA are doubly robust, they lead to different conclusions about the LATE on net financial assets.

Second, we also revisit Finkelstein et al. (2012) and Taubman et al. (2014), and use the data from the Oregon Health Insurance Experiment to study the effects of Medicaid on emergency room visits. Like in previous work, our estimates are positive, which suggests, perhaps counterintuitively, that access to health insurance may increase the utilization of emergency rooms. Our novel empirical contribution is that LATT, the effect on the treated compliers, is larger than the usual LATE, at least along the extensive margin. This is because treatment effects appear to be more pronounced in larger households (cf. Denteh and Liebert, 2022), which are also more likely

to be treated.

Section 7 presents simulation evidence on the performance of several estimators of LATE, including IV, regression adjustment (RA), IPW, AIPW, and IPWRA. The performance of our proposed method, IPWRA, is very satisfying. It is never substantially more biased than the competing estimators while its precision is better than that of AIPW, which is the only alternative that shares the double robustness property of IPWRA. Finally, Section 8 concludes.

2 Identification of LATE and LATT

The potential outcomes setting in this paper is the one pioneered by IA (1994). Eventually, we will assume access to a random sample from the population, and so all assumptions can be stated in terms of random variables representing the population of interest.

For a binary intervention, let $Y(0)$ be the potential outcome in the control state and $Y(1)$ the potential outcome in the treated state. The observed binary treatment indicator is W, where $W = 1$ denotes treatment and $W = 0$ denotes control. We have access to a binary instrumental variable, Z. As in IA (1994), there are potential treatment statuses based on the assignment of the instrument (which is often eligibility): $W(1)$ is participation status when a unit is made "eligible" and $W(0)$ is participation status in the "ineligible" state. This framework allows for the possibility that units do not comply with their assigned "eligibility." For example, some workers, if selected to participate in a job training program ($Z = 1$), may choose not to participate [$W(1) = 0$].

The observed outcome Y is a function of the observed treatment variable and the potential outcomes corresponding to the treatment and control status:

$$Y = WY(1) + (1 - W)Y(0).$$

Further, the realized treatment status can be written in terms of the instrument Z and the potential treatment statuses:

$$W = ZW(1) + (1 - Z)W(0). \tag{1}$$

According to the relationship between the potential treatment status and the binary instrument, the population can be divided into four subpopulations: compliers, always-takers, never-takers, and defiers. From the observed dataset one cannot identify the group to which an individual belongs since only the pair (W, Z) is observed. For example, if $Z = 1$ and $W = 1$, the individual is either a complier or an always-taker. Always-takers and never-takers do not change their treatment behavior when the assignment of the IV changes. The only subpopulations that can be induced into changing W through a variation in Z are the defiers and compliers.

Generally, the treatment effect of interest can be defined either as the impact of the treatment on the outcome for the defiers $[W(1) < W(0)]$ or for the compliers $[W(1) > W(0)]$. Following the literature, we focus on average treatment effects on compliers under the assumption that defiers do not exist. The local average treatment effect (LATE) is the expected difference between the potential outcomes for the subpopulation of compliers:

$$\tau_{LATE} = \mathbb{E}[Y(1) - Y(0)|W(1) > W(0)]. \tag{2}$$

Compliers are members of a hypothetically defined subpopulation and cannot be identified from observed data without further assumptions.

As in much of the literature since IA (1994) – including several papers discussed in the introduction – we assume that we have (pre-treatment) covariates, X, that render the instrumental variable suitably exogenous when conditioned on. The support of X is indicated by $\mathcal{X}$. With these covariates, identification of τ_{LATE} is possible if certain assumptions are met.

The first assumption is that, conditional on X, the instrumental variable has no direct effect on the potential outcomes; its effect can come only through the treatment assignment. The formal statement requires indicating two arguments in the potential outcomes, $Y(w, z)$ for $w, z \in \{0, 1\}$.

Assumption 1 (Exclusion Restriction)**.** *For $w \in \{0, 1\}$ and almost all $x \in \mathcal{X}$,*

$$\mathbb{P}\left[Y(w, 1) = Y(w, 0) \mid X = x\right] = 1. \square \tag{3}$$

Assumption 1 justifies labeling the potential outcomes using a single index that indicates actual

treatment status because we condition on X in stating ignorability of the instruments. In what follows, we use $Y(w)$ as the potential outcome for treatment status $w \in \{0, 1\}$.

Assumption 2 (Ignorability of Instrument)**.** *Conditional on X, the potential outcomes are jointly independent of Z:*

$$[Y(0), Y(1), W(0), W(1)] \perp Z \mid X. \square \tag{4}$$

Assumption 2 requires that, conditional on observed confounders, the instrument can be regarded as random.

Assumption 3 (Monotonicity)**.**

$$\mathbb{P}[W(1) \geq W(0)] = 1. \square$$

This monotonicity assumption is standard in the literature: it says that there are no defiers in the population (or that the group is so small it has probability zero). It is equivalent to a conditional statement, namely, $\mathbb{P}[W(1) \geq W(0)|X = x] = 1$ for almost all $x \in \mathcal{X}$. In other words, if Assumption 3 holds, $\mathbb{P}[W(1) \geq W(0)|X = x] < 1$ is possible only on a subset of $x \in \mathcal{X}$ with measure zero. Formally, this claim is equivalent to the proposition that for a random variable $R \geq 0$, $\mathbb{E}(R) = 0$ if and only if $\mathbb{P}(R = 0) = 1$.

The next assumption requires the existence of compliers in the population.

Assumption 4 (Existence of Compliers)**.**

$$\mathbb{P}[W(1) > W(0)] > 0. \square$$

When we partition the population on the basis of the covariates X, Assumption 4 implies that, for some subset $\mathcal{C} \subset \mathcal{X}$ with $\mathbb{P}(\mathcal{C}) > 0$, $\mathbb{P}[W(1) > W(0)|X = x] > 0$ if $x \in \mathcal{C}$. This follows by iterated expectations: If $\mathbb{1}[W(1) > W(0)]$ has positive expectation then its expectation conditional on X must be positive with nonzero probability. The subset $\mathcal{C}$ defines the subpopulation of compliers based on the values of X.

The requisite overlap assumption is stated in terms of the propensity score involving the instrumental variable, sometimes referred to as the *instrument propensity score*.

Assumption 5 (Overlap for LATE)**.** *For almost all $x \in \mathcal{X}$,*

$$0 < \mathbb{P}\left(Z = 1 | X = x\right) < 1. \square$$

IA (1994) show that if Assumptions 1–5 hold without conditioning on X, then τ_{LATE} is identified as

$$\tau_{LATE} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[W|Z=1] - \mathbb{E}[W|Z=0]}. \tag{5}$$

In that case, given a random sample from the population, τ_{LATE} can be consistently estimated by replacing the expectations in (5) with the corresponding sample averages. This simple estimator is the well-known Wald estimator; it is also obtained by estimating the simple linear equation $Y = \alpha + \beta W + U$ by instrumental variables using instruments $(1, Z)$. In other words, $\hat{\beta}_{IV} = \hat{\tau}_{LATE}$. In some cases, one cannot participate ($W = 1$) unless assigned to the treatment ($Z = 1$), in which case the second term in the denominator of τ_{LATE} is zero: $\mathbb{E}[W|Z=0] = \mathbb{P}[W=1|Z=0] = 0$. This is the case in the application in Abadie (2003) (also used in several subsequent studies), where an employee cannot participate in an employer-sponsored pension plan unless the employer offers such a plan. In other applications, $Z = 1$ implies $W = 1$ (no never-takers), in which case the first term in the denominator of τ_{LATE} is unity. This situation arises in Angrist and Evans (1998) when, in a population of women with at least two children, the "treatment" is having more than two children and the binary instrument indicates whether the second birth was a multiple birth.

In many cases the instrumental variable candidate, Z, is not truly randomized, but we might be willing to assume it is as good as randomized conditional on X. Then, Assumptions 1–5 imply that we can identify τ_{LATE}. The following theorem is due to Frölich (2007, Theorem 1). Frölich and Lechner (2010) relax the assumptions somewhat but not in a way that makes the theorem clearly more applicable.

Theorem 1 (Identification of LATE)**.** *Under Assumptions 1–5,*

$$\tau_{LATE} = \frac{\mathbb{E}\left[\mathbb{E}\left(Y|X, Z=1\right) - \mathbb{E}\left(Y|X, Z=0\right)\right]}{\mathbb{E}\left[\mathbb{E}\left(W|X, Z=1\right) - \mathbb{E}\left(W|X, Z=0\right)\right]} = \frac{\mathbb{E}[\mu_1(X) - \mu_0(X)]}{\mathbb{E}[\rho_1(X) - \rho_0(X)]}, \tag{6}$$

where

$$\mu_0(X) \equiv \mathbb{E}\left(Y|X, Z=0\right) \tag{7}$$

$$\mu_1(X) \equiv \mathbb{E}\left(Y|X, Z=1\right) \tag{8}$$

and

$$\mathbb{E}\left(W|X, Z=0\right) = \mathbb{E}[W(0)|X] = \rho_0(X) \tag{9}$$

$$\mathbb{E}\left(W|X, Z=1\right) = \mathbb{E}[W(1)|X] = \rho_1(X). \square \tag{10}$$

As discussed in Frölich (2007), the result in equation (6) suggests that one can estimate each of the four conditional mean functions, $\mathbb{E}\left(Y|X, Z=0\right)$, $\mathbb{E}\left(Y|X, Z=1\right)$, $\mathbb{E}\left(W|X, Z=0\right)$, and $\mathbb{E}\left(W|X, Z=1\right)$, using nonparametric methods, and then average the estimates across i to estimate the unconditional means. Especially when the dimension of X is large, nonparametric estimation is not attractive, and inference is also complicated. One of the estimators considered by Tan (2006) estimates the numerator and denominator in (6) using augmented inverse probability weighting (AIPW) estimators. AIPW estimators are popular in the case of unconfounded assignment but simulations show they are not always well behaved even in somewhat large samples (e.g., Kang and Schafer, 2007). Słoczyński and Wooldridge (2018) provide a recent overview of the debate on the merits of AIPW approaches. One issue that apparently has not been noted is that commonly used AIPW estimators – like those in Tan (2006) – implicitly use weights in the weighted averages that do not sum to unity; in other words, the weights are not normalized. Because many of the estimators summarized in the introduction have an AIPW flavor, they suffer from the same problem – whether they are based on parametric approaches or machine learning algorithms. In the next section we show how to use a class of DR estimators based on weighted quasi-maximum likelihood estimation (QMLE) to estimate the four means that appear in (6). This possibility was already

indicated by Słoczyński and Wooldridge (2018) but without any of the details that we consider here.

There is also some interest in estimating the local average treatment effect on the treated (LATT), formally defined as

$$\tau_{LATT} = \mathbb{E}[Y(1) - Y(0) \mid W(1) > W(0), W = 1]. \tag{11}$$

Frölich and Lechner (2010) study identification of this parameter and show that, under the assumptions in Theorem 1,

$$\tau_{LATT} = \frac{\mathbb{E}\left\{[\mu_1(X) - \mu_0(X)]\eta(X)\right\}}{\mathbb{E}\left\{[\rho_1(X) - \rho_0(X)]\eta(X)\right\}}, \tag{12}$$

where $\eta(x)$ is the instrument propensity score:

$$\eta(x) \equiv \mathbb{P}\left(Z = 1 | X = x\right), x \in \mathcal{X}. \tag{13}$$

For our purposes, we use a different representation of τ_{LATT}. We state this theorem under the exclusion and ignorability assumptions in Theorem 1 even though we could relax some of the assumptions. It is useful to explicitly relax the overlap assumption.

Assumption 6 (Overlap for LATT)**.** *For almost all $x \in \mathcal{X}$,*

$$\mathbb{P}\left(Z = 1 | X = x\right) < 1. \square$$

The overlap assumption for LATT means that there can be subsets of the population, based on the values of the control variables X, where units are not eligible for the treatment.

Theorem 2 (Identification of LATT)**.** *Under Assumptions 1–4 and Assumption 6,*

$$\tau_{LATT} = \frac{\mathbb{E}\left(Y | Z = 1\right) - \mathbb{E}[\mu_0(X) | Z = 1]}{\mathbb{E}\left(W | Z = 1\right) - \mathbb{E}[\rho_0(X) | Z = 1]}. \square \tag{14}$$

Proof. By iterated expectations and (12),

$$\tau_{LATT} = \frac{\mathbb{E}\left\{[\mu_1(X) - \mu_0(X)]Z\right\}}{\mathbb{E}\left\{[\rho_1(X) - \rho_0(X)]Z\right\}}$$

and so, dividing the numerator and the denominator by $\mathbb{P}(Z=1)>0$,

$$\tau_{LATT}=\frac{\mathbb{E}\left\{[\mu_1(X)-\mu_0(X)]|Z=1\right\}}{\mathbb{E}\left\{[\rho_1(X)-\rho_0(X)]|Z=1\right\}}.$$

By definition of $\mu_0(x)$ and $\mu_1(x)$, we can write

$$Y=(1-Z)\mu_0(X)+Z\mu_1(X)+U,\ \ \mathbb{E}(U|X,Z)=0.$$

It follows that

$$\mathbb{E}(Y|Z=1)=\mathbb{E}\left[\mu_1(X)|Z=1\right].$$

Also, $W=(1-Z)W(0)+ZW(1)$ and so, by ignorability,

$$\mathbb{E}(W|X,Z=1)=\mathbb{E}\left[W(1)|X,Z=1\right]=\mathbb{E}\left[W(1)|X\right]=\rho_1(X).$$

Iterated expectations implies $\mathbb{E}(W|Z=1)=\mathbb{E}\left[\rho_1(X)|Z=1\right]$. Therefore, we can write τ_{LATT} as in (14). The overlap assumption ensures that $\mathbb{E}[\mu_0(X)|Z=1]$ is identified and ignorability and overlap ensure $\mathbb{E}[\rho_0(X)|Z=1]$ is identified. □

As we show in the next section, the representations in (6) and (14) permit doubly robust estimation of τ_{LATE} and τ_{LATT} using IPWRA estimators, providing a unification that makes the estimation approaches transparent and simple.

3 Doubly Robust Estimation of LATE and LATT

We now turn to estimation of τ_{LATE} and τ_{LATT} using a particular class of doubly robust (DR) estimators, starting with the former. The approach we take allows us to tailor the analysis based on the nature of the observed outcome, Y, by choosing a suitable conditional mean function. In particular, using the identification results in Słoczyński and Wooldridge (2018), we extend Wooldridge (2007)'s approach of combining inverse probability weighting (IPW) and regression adjustment (RA) using a particular quasi-maximum likelihood estimator. These estimators are commonly

referred to as IPWRA estimators.

3.1 Estimation of LATE

The identification result in equation (6) shows that, in order to consistently estimate τ_{LATE}, we need to consistently estimate the following four quantities:

$$\theta_1 = \mathbb{E}[\mu_1(X)],\ \theta_0 = \mathbb{E}[\mu_0(X)], \tag{15}$$

$$\pi_1 = \mathbb{E}[\rho_1(X)],\ \pi_0 = \mathbb{E}[\rho_0(X)]. \tag{16}$$

Because of the representation of W in equation (1), we can immediately apply the DR results on IPWRA estimation from Wooldridge (2007) and Słoczyński and Wooldridge (2018). The approach first requires estimating a binary response model for the instrument propensity score defined in (13). By the overlap assumption (Assumption 5), $0 < \eta(x) < 1$ for all $x \in \mathcal{X}$. The proposal here is to use a standard parametric model for $\eta(x)$, as is common in the literature when estimating a propensity score function. Probably most popular is a flexible logit model, but it could be a probit model, heteroskedastic probit model, or something else. Let $G(x, \gamma)$ denote the parametric model for $\eta(x)$. Under very general assumptions, the Bernoulli quasi-maximum likelihood estimator, $\hat{\gamma}$, converges in probability to some value, γ^*, which is sometimes called the quasi-true value or pseudo-true value. If the model for $\eta(x)$ is correctly specified then $G(x, \gamma^*) = \mathbb{P}(Z = 1|X = x)$. At this point, one would use the fitted probabilities, $G(X_i, \hat{\gamma})$, to study the LATE overlap condition using standard methods; for a detailed discussion, see Imbens and Rubin (2015, Chapter 14).

After estimating the model for $\mathbb{P}(Z = 1|X)$, next we estimate models for $\rho_0(x)$ and $\rho_1(x)$, as defined in (9) and (10). These are estimated by separate logit models for W for the $Z_i = 0$ and $Z_i = 1$ subgroups, applying the inverse probability weights $1/\left[1 - G(X_i, \hat{\gamma})\right]$ and $1/G(X_i, \hat{\gamma})$, respectively. The reason for using logit models for $\rho_0(x)$ and $\rho_1(x)$ is to ensure that the resulting estimators of the expected probabilities, π_0 and π_1, are doubly robust, as discussed in Wooldridge (2007) and Słoczyński and Wooldridge (2018). The logit function is the canonical link function for the Bernoulli distribution, and that ensures the DR property. Let $\Lambda(\hat{\omega}_0 + X_i\hat{\delta}_0)$ and $\Lambda(\hat{\omega}_1 + X_i\hat{\delta}_1)$

be the logit fitted values, where the estimated parameters are obtained from the $Z_i = 0$ and $Z_i = 1$ subsamples, respectively. For notational ease we show the indexes as linear functions of X_i but, naturally, any functions of the covariates may appear in the logit models. In principle, one could use different functions of X_i, say $h_0(X_i)$ and $h_1(X_i)$, inside the logistic function, but that seems to be rare in practice.

Having estimated the separate logit models by weighted Bernoulli QMLE, the DR estimates of π_0 and π_1 are

$$\hat{\pi}_0 = N^{-1} \sum_{i=1}^{N} \Lambda(\hat{\omega}_0 + X_i \hat{\delta}_0), \ \hat{\pi}_1 = N^{-1} \sum_{i=1}^{N} \Lambda(\hat{\omega}_1 + X_i \hat{\delta}_1).$$

From Wooldridge (2007), under standard regularity conditions, $\hat{\pi}_z$ is consistent for π_z if the model for $\mathbb{P}(Z = 1|X)$ is correct or if the models for $\mathbb{P}(W = 1|X, Z = 0)$ and $\mathbb{P}(W = 1|X, Z = 1)$ are correct. Naturally, if we know $\mathbb{P}(W = 1|Z = 1) = 1$ then $\hat{\pi}_1$ is replaced with one and if we know $\mathbb{P}(W = 1|Z = 0) = 0$ then $\hat{\pi}_0$ is replaced with zero (the more likely scenario when Z is eligibility and W is participation).

Next, we show how to obtain DR estimators of θ_0 and θ_1. In doing so, it is useful to write Y with a zero conditional mean error term:

$$Y = (1 - Z)\mu_0(X) + Z\mu_1(X) + U, \ \ \mathbb{E}\left(U|X, Z\right) = 0,$$

where $\mu_0(X)$ and $\mu_1(X)$ are defined in (7) and (8). Note that this is not the usual representation that leads to DR estimation because $\mu_0(X)$ and $\mu_1(X)$ are not the potential outcome conditional means; rather, these are the conditional mean functions for the (observed) $Z = 0$ and $Z = 1$ subpopulations, respectively. Therefore, we must modify the usual double robustness argument.

Let $m(\alpha_0 + X\beta_0)$ and $m(\alpha_1 + X\beta_1)$ be the parametric models for $\mu_0(X)$ and $\mu_1(X)$. Again, for notational ease we show these depending on an index linear in X, whereas they could depend on (different) transformations of X inside the function $m\left(\cdot\right)$. We assume that the function $m\left(\cdot\right)$ is based on the canonical link function for the chosen quasi-log likelihood (QLL) in the linear exponential family – which is why we show the mean function to have the index form. If Y is a

binary or fractional response then we couple the Bernoulli QLL with the logistic mean function – just as when we estimate $\rho_0(x)$ and $\rho_1(x)$. If $Y \geq 0$, the appropriate combination is $m(\cdot) = \exp(\cdot)$ with the Poisson QLL. When Y has no particular features worth exploiting, one commonly uses $m(\alpha + x\beta) = \alpha + x\beta$ and the least squares objective function (which corresponds to the normal QLL).

As in the case of estimating the parametric models for $\rho_0(x)$ and $\rho_1(x)$, the objective functions for estimating (α_0, β_0) and (α_1, β_1) are weighted by $1/[1 - G(X_i, \hat{\gamma})]$ and $1/G(X_i, \hat{\gamma})$ for the $Z_i = 0$ and $Z_i = 1$ subsamples, respectively. When the mean function corresponds to the canonical link function in the chosen linear exponential family (LEF), the first-order conditions for $(\hat{\alpha}_1, \hat{\beta}_1)$ can be written as

$$\sum_{i=1}^{N} \left\{ \frac{Z_i}{G(X_i, \hat{\gamma})} \left[Y_i - m(\hat{\alpha}_1 + X_i\hat{\beta}_1) \right] \right\} = 0 \tag{17}$$

$$\sum_{i=1}^{N} \left\{ \frac{Z_i}{G(X_i, \hat{\gamma})} X_i' \left[Y_i - m(\hat{\alpha}_1 + X_i\hat{\beta}_1) \right] \right\} = 0. \tag{18}$$

Under general conditions, $(\hat{\alpha}_1, \hat{\beta}_1)$ converge in probability to the (unique) solutions (α_1^*, β_1^*) to the weighted population moment conditions

$$\mathbb{E}\left\{ \frac{Z}{G(X, \gamma^*)} \left[Y - m(\alpha_1^* + X\beta_1^*) \right] \right\} = 0 \tag{19}$$

$$\mathbb{E}\left\{ \frac{Z}{G(X, \gamma^*)} X' \left[Y - m(\alpha_1^* + X\beta_1^*) \right] \right\} = 0. \tag{20}$$

We now show that the solutions to these FOCs result in doubly robust estimators of θ_0 and θ_1. We show the argument for the latter with an almost identical argument for θ_0.

As discussed in Wooldridge (2007), when the weights depend on conditioning variables – in this case, X – and the relevant feature of the conditional distribution is correctly specified – in this case, the conditional mean $\mu_1(X) \equiv \mathbb{E}(Y|X, Z = 1)$ – weighting a suitably chosen objective function does not alter consistency of the estimators. We can see this directly from the population

FOCs. Assume there are values (α_1^*, β_1^*) such that

$$\mathbb{E}\left(Y|X, Z=1\right) = m(\alpha_1^* + X\beta_1^*),$$

so that the conditional mean is correctly specified. Then $ZY = Zm(\alpha_1^* + X\beta_1^*) + ZU$ and, since $\mathbb{E}\left(ZU|X, Z\right) = 0$, it follows immediately that

$$\mathbb{E}\left\{Z\left[Y - m(\alpha_1^* + X\beta_1^*)\right]|X\right\} = 0.$$

Because $G(X, \gamma^*) > 0$ is a function of X, it follows that

$$\mathbb{E}\left\{ \frac{Z}{G(X,\gamma^*)}\left[Y - m(\alpha_1^* + X\beta_1^*)\right] \middle| X \right\} = 0.$$

Given $G(X, \gamma^*) > 0$ and sufficient variability in X when $Z = 1$, the solutions to (19) and (20), (α_1^*, β_1^*), are unique. By iterated expectations,

$$\theta_1 = \mathbb{E}\left[m(\alpha_1^* + X\beta_1^*)\right].$$

Similarly, $\theta_0 = \mathbb{E}[m(\alpha_0^* + X\beta_0^*)]$ when $\mathbb{E}\left(Y|X, Z=0\right) = m(\alpha_0^* + X\beta_0^*)$. This is the first half of the double robustness result, which does not actually use the assumption of a canonical link in the linear exponential family.

For the other part of DR, it is useful to express the population FOCs somewhat differently. Plug in for Y and use $ZY = Z\mu_1(X) + ZU$ to get

$$\begin{aligned}
\mathbb{E}\left[\frac{Z}{G(X,\gamma^*)}[\mu_1(X) + U - m(\alpha_1^* + X\beta_1^*)]\right] &= 0 \\
\mathbb{E}\left[\frac{Z}{G(X,\gamma^*)}X'[\mu_1(X) + U - m(\alpha_1^* + X\beta_1^*)]\right] &= 0
\end{aligned}$$

or, because $\mathbb{E}\left(U|X, Z\right) = 0$,

$$\mathbb{E}\left[\frac{Z}{G(X,\gamma^*)}[\mu_1(X) - m(\alpha_1^* + X\beta_1^*)]\right] = 0 \tag{21}$$

$$\mathbb{E}\left[\frac{Z}{G(X,\gamma^*)}X'[\mu_1(X) - m(\alpha_1^* + X\beta_1^*)]\right] = 0. \tag{22}$$

By iterated expectations, these equations are equivalent to

$$\mathbb{E}\left[\frac{\eta(X)}{G(X,\gamma^*)}[\mu_1(X) - m(\alpha_1^* + X\beta_1^*)]\right] = 0 \tag{23}$$

$$\mathbb{E}\left[\frac{\eta(X)}{G(X,\gamma^*)}X'[\mu_1(X) - m(\alpha_1^* + X\beta_1^*)]\right] = 0. \tag{24}$$

When $G(x,\gamma)$ is correctly specified, $\eta(X) = G(X,\gamma^*)$, and the first population moment condition becomes

$$\mathbb{E}\left[\mu_1(X) - m(\alpha_1^* + X\beta_1^*)\right] = 0.$$

It follows immediately that $\theta_1 = \mathbb{E}[m(\alpha_1^* + X\beta_1^*)]$, even though the conditional mean function need not be correctly specified. This part of the DR result uses the assumption that we have chosen the canonical link function for the chosen LEF density. The same argument holds for θ_0. Except for adding standard regularity conditions, we have established consistency of the DR estimators that combine inverse probability weighting (IPW) and regression adjustment (RA), where RA is defined generally to include QMLEs in the LEF with a canonical link function.

Because the LEF/canonical link combinations play an important role in DR estimation, we summarize the common choices for the quasi-likelihoods and mean functions in Table 1.

The first entry in Table 1 simply means using weighted least squares with linear conditional mean functions, but the weights here are based on the instrument propensity score, chosen to achieve double robustness, and have nothing to do with heteroskedasticity. When Y is binary or fractional (the second entry), a logistic conditional mean function is more attractive because it ensures fitted values are in the unit interval. For example, Y_i could be the fraction of retirement savings held in the stock market or the fraction of students passing a standardized test. The third entry allows for corners at zero and some unit-specific, known upper bound, B_i. This B_i should be a conditioning variable – like the elements of X_i. For example, Y_i could be the amount of income put into retirement with B_i being an individual-specific bound determined by legal restrictions. The final entry is important across many kinds of response variables that are nonnegative but have no natural upper bound. These outcomes could be count variables but they could be roughly

continuous or have an atom at zero.

In what follows, we summarize the steps for doubly robust estimation of τ_{LATE} using IPWRA.

Procedure DR LATE.

1. Estimate a flexible binary response model for the instrument propensity score, $\eta(x) = \mathbb{P}(Z = 1|X = x)$; denote the fitted probabilities $G(X_i, \hat{\gamma})$. In many cases, one would use a flexible logit. Overlap needs to be studied at this step.

2. Use weighted Bernoulli QMLE to estimate separate (flexible) logit models for $\mathbb{P}(W = 1|X, Z = 0)$ and $\mathbb{P}(W = 1|X, Z = 1)$ (i.e., only using the units with $Z_i = 0$ and $Z_i = 1$, respectively), where the weights in the former case are $1/[1 - G(X_i, \hat{\gamma})]$, and in the latter case, $1/G(X_i, \hat{\gamma})$. These produce $\left(\hat{\omega}_0, \hat{\delta}_0\right)$, $\left(\hat{\omega}_1, \hat{\delta}_1\right)$, and the fitted probabilities $\Lambda(\hat{\omega}_0 + X_i\hat{\delta}_0)$ and $\Lambda(\hat{\omega}_1 + X_i\hat{\delta}_1)$.

3. Choose conditional mean models for $\mathbb{E}(Y|X, Z = 0)$ and $\mathbb{E}(Y|X, Z = 1)$ that reflect the nature of Y. These should correspond to the canonical link functions for the chosen LEF quasi-log likelihood. Use weights $1/[1 - G(X_i, \hat{\gamma})]$ to obtain the weighted QMLEs of (α_0, β_0) and weights $1/G(X_i, \hat{\gamma})$ to obtain the weighted QMLEs of (α_1, β_1). (As above, this only uses the units with $Z_i = 0$ and then $Z_i = 1$, respectively.) These produce the fitted mean functions $m(\hat{\alpha}_0 + X_i\hat{\beta}_0)$ and $m(\hat{\alpha}_1 + X_i\hat{\beta}_1)$.

4. Obtain the DR estimator of τ_{LATE} as

$$\hat{\tau}_{DRLATE} = \frac{N^{-1}\sum_{i=1}^{N}\left[m(\hat{\alpha}_1 + X_i\hat{\beta}_1) - m(\hat{\alpha}_0 + X_i\hat{\beta}_0)\right]}{N^{-1}\sum_{i=1}^{N}\left[\Lambda(\hat{\omega}_1 + X_i\hat{\delta}_1) - \Lambda(\hat{\omega}_0 + X_i\hat{\delta}_0)\right]}. \square \tag{25}$$

The DR LATE estimator has the same form as Frölich (2007), but we use parametric models that can exploit the nature of Y and we estimate the parameters in the numerator and denominator using inverse probability weighting in order to achieve double robustness. Consequently, the numerator of $\hat{\tau}_{DRLATE}$ is a DR ATE estimator where Z (the instrument) is taken as the "treatment" and the outcome Y is the response. The denominator is a DR ATE estimator where, again, Z is the

"treatment" and the actual treatment indicator, W, is the response. Obtaining the estimate for a given sample is very easy using software packages that support IPWRA estimation.

3.2 Estimation of LATT

The IPWRA doubly robust estimators of τ_{LATT} require a different weighting scheme. First, there is no need to model $\mu_1(X) = \mathbb{E}(Y|X, Z = 1)$ or $\rho_1(X) = \mathbb{E}(W|X, Z = 1)$ because, as shown in (14), we only need to estimate $\mathbb{E}(Y|Z = 1)$ and $\mathbb{E}(W|Z = 1)$. But we need DR estimators of $\mathbb{E}[\mu_0(X)|Z = 1]$ and $\mathbb{E}[\rho_0(X)|Z = 1]$. Following, for example, Słoczyński and Wooldridge (2018), we now show that the following population FOC provides DR estimators of $\mathbb{E}[\mu_0(X)|Z = 1]$:

$$\mathbb{E}\left\{\frac{(1-Z)\,G(X,\gamma^*)}{[1-G(X,\gamma^*)]}\left[Y - m(\alpha_0^* + X\beta_0^*)\right]\right\} = 0. \tag{26}$$

Using the same argument as for τ_{LATE}, $\mathbb{E}\left(Y|X, Z = 0\right) = m(\alpha_0^* + X\beta_0^*)$ ensures that

$$\mathbb{E}\left\{(1-Z)\left[Y - m(\alpha_0^* + X\beta_0^*)\right]|X\right\} = 0,$$

and then (26) holds by iterated expectations. Again, the weights are nonnegative functions of X and so this does not change that the solutions to the FOCs are the conditional mean parameters.

For the second half of DR, we use an argument similar to the case of LATE and write the FOC as

$$\mathbb{E}\left\{\frac{(1-Z)\,G(X,\gamma^*)}{[1-G(X,\gamma^*)]}\left[\mu_0\left(X\right) - m(\alpha_0^* + X\beta_0^*)\right]\right\} = 0.$$

By iterated expectations, this FOC is equivalent to

$$\mathbb{E}\left\{\frac{[1-\eta\left(X\right)]\,G(X,\gamma^*)}{[1-G(X,\gamma^*)]}\left[\mu_0\left(X\right) - m(\alpha_0^* + X\beta_0^*)\right]\right\} = 0,$$

where $\eta\left(X\right) = \mathbb{P}\left(Z = 1|X\right)$. When the instrument propensity score is correctly specified, $\eta\left(X\right) = G(X, \gamma^*)$, this equation becomes

$$\mathbb{E}\left\{\eta\left(X\right)\left[\mu_0\left(X\right) - m(\alpha_0^* + X\beta_0^*)\right]\right\} = 0,$$

and by iterated expectations, this is equivalent to

$$\mathbb{E}\left\{Z\left[\mu_0\left(X\right) - m(\alpha_0^* + X\beta_0^*)\right]\right\} = 0.$$

It now follows that

$$\mathbb{E}\left[\mu_0\left(X\right)|Z=1\right] = \mathbb{E}\left[m(\alpha_0^* + X\beta_0^*)|Z=1\right]$$

even though the mean function is arbitrarily misspecified. This is the second half of the DR result for τ_{LATT}.

Procedure DR LATT.

1. Using all of the data, estimate a flexible binary response model for the instrument propensity score, $\eta\left(x\right) = \mathbb{P}\left(Z=1|X=x\right)$; denote the fitted probabilities $G\left(X_i, \hat{\gamma}\right)$. In many cases, one would use a flexible logit. The LATT overlap assumption needs to be studied at this step.

2. Use the units with $Z_i = 0$ and weighted Bernoulli QMLE to estimate a (flexible) logit model for $\rho_0(X) = \mathbb{P}(W=1|X, Z=0)$, where the weights are $G(X_i, \hat{\gamma})/\left[1 - G\left(X_i, \hat{\gamma}\right)\right]$. This produces $\left(\hat{\omega}_0, \hat{\delta}_0\right)$ and the fitted probabilities $\Lambda(\hat{\omega}_0 + X_i\hat{\delta}_0)$.

3. Choose a conditional mean model for $\mu_0(X) = \mathbb{E}(Y|X, Z=0)$ that reflects the nature of Y. This should correspond to the canonical link function for the chosen LEF quasi-log likelihood. Use the units with $Z_i = 0$ and weights $G(X_i, \hat{\gamma})/\left[1 - G\left(X_i, \hat{\gamma}\right)\right]$ to obtain the weighted QMLEs of (α_0, β_0). This produces the fitted mean, $m(\hat{\alpha}_0 + X_i\hat{\beta}_0)$.

4. Obtain the DR estimator of τ_{LATT} as

$$\hat{\tau}_{DRLATT} = \frac{\bar{Y}_1 - N_1^{-1}\sum_{i=1}^N Z_i m(\hat{\alpha}_0 + X_i\hat{\beta}_0)}{\bar{W}_1 - N_1^{-1}\sum_{i=1}^N Z_i \Lambda(\hat{\omega}_0 + X_i\hat{\delta}_0)}, \tag{27}$$

   where $\bar{Y}_1 = N_1^{-1}\sum_{i=1}^N Z_i Y_i$ and $\bar{W}_1 = N_1^{-1}\sum_{i=1}^N Z_i W_i$. □

The numerator of $\hat{\tau}_{DRLATT}$ is a DR estimator of ATT where Z is taken as the "treatment" and

Y is the outcome. Similarly, the denominator is a DR estimator of ATT where Z is taken as the treatment and the actual treatment status, W, is the outcome.

4 Inference

To perform valid inference on τ_{LATE} and τ_{LATT}, such as obtaining confidence intervals, we need to obtain standard errors for $\hat{\tau}_{DRLATE}$ and $\hat{\tau}_{DRLATT}$ that account for the sampling error in all of the estimators and also the sample averages in (25) and (27). One possibility is to use a resampling scheme. The most convenient is the nonparametric bootstrap, which resamples all variables (and so accounts for sampling error in the estimators and in the averages). Given a bootstrapped standard error we can easily obtain asymptotically valid confidence intervals for τ_{LATE} and τ_{LATT}.

Because bootstrapping is not always desirable, we summarize a method of obtaining a valid standard error that stacks the first-order conditions for all estimation problems and then obtains a proper standard error from the resulting generalized method of moments framework. We explicitly consider how to do this for $\hat{\tau}_{DRLATE}$.

To allow one to choose the treatment binary response models and the conditional mean models in a way that does not (theoretically) lead to DR estimation, let $m_0(X, \alpha_0, \beta_0)$ and $m_1(X, \alpha_1, \beta_1)$ be the parametric models for $\mu_0(X)$ and $\mu_1(X)$, respectively, and let $p_0(X, \omega_0, \delta_0)$ and $p_1(X, \omega_1, \delta_1)$ be the parametric models for $\rho_0(X)$ and $\rho_1(X)$, respectively. $G(X, \gamma)$ is the parametric model for $\mathbb{P}(Z = 1|X)$. Let $\tau_{Y|Z}$ and $\tau_{W|Z}$ be the numerator and denominator of the LATE, respectively, that is,

$$\begin{aligned} \tau_{Y|Z} &= \theta_1 - \theta_0, \\ \tau_{W|Z} &= \pi_1 - \pi_0. \end{aligned}$$

Let $\phi = (\alpha_0, \beta_0, \alpha_1, \beta_1, \omega_0, \delta_0, \omega_1, \delta_1, \gamma, \tau_{Y|Z}, \tau_{W|Z})$ and $S = (Y, X, W, Z)$. The estimators can be

defined as a solution for the following sample moment equation:

$$\sum_{i=1}^{N} \psi(S_i, \hat{\phi}) = 0. \tag{28}$$

By standard results for estimators that solve a first-order condition, it follows that:

$$\sqrt{N}(\hat{\phi} - \phi) \overset{a}{\sim} \text{Normal}\left(0, A^{-1}VA^{-1}\right), \tag{29}$$

where

$$\begin{aligned} A &\equiv \mathbb{E}\left[\frac{\partial \psi(S_i, \phi)}{\partial \phi'}\right], \\ V &\equiv \mathbb{V}[\psi(S_i, \phi)] = \mathbb{E}[\psi(S_i, \phi)\psi(S_i, \phi)']. \end{aligned}$$

Using the moment functions related to each parameter, the moment function $\psi(S_i, \hat{\phi})$ in equation (28) can be written explicitly in the following way:

$$\psi(S_i, \phi) = \begin{pmatrix} \psi_1(S_i, \phi) \\ \psi_2(S_i, \phi) \\ \psi_3(S_i, \phi) \\ \psi_4(S_i, \phi) \\ \psi_5(S_i, \phi) \\ \psi_6(S_i, \phi) \\ \psi_7(S_i, \phi) \end{pmatrix} = \begin{pmatrix} \frac{Z_i}{G(X_i,\gamma)} \frac{\partial q_1^y(Y_i, X_i; \alpha_1, \beta_1)}{\partial(\alpha_1, \beta_1)} \\ \frac{1-Z_i}{1-G(X_i,\gamma)} \frac{\partial q_0^y(Y_i, X_i; \alpha_0, \beta_0)}{\partial(\alpha_0, \beta_0)} \\ \frac{Z_i}{G(X_i,\gamma)} \frac{\partial q_1^w(W_i, X_i; \omega_1, \delta_1)}{\partial(\omega_1, \delta_1)} \\ \frac{1-Z_i}{1-G(X_i,\gamma)} \frac{\partial q_0^w(W_i, X_i; \omega_0, \delta_0)}{\partial(\omega_0, \delta_0)} \\ \frac{Z_i - G(X_i,\gamma)}{G(X_i,\gamma)(1-G(X_i,\gamma))} \frac{\partial G(X_i,\gamma)}{\partial \gamma} \\ m_1(X_i, \alpha_1, \beta_1) - m_0(X_i, \alpha_0, \beta_0) - \tau_{Y|Z} \\ p_1(X_i, \omega_1, \delta_1) - p_0(X_i, \omega_0, \delta_0) - \tau_{W|Z} \end{pmatrix},$$

where $q_z^y(\cdot)$ and $q_z^w(\cdot)$ are the objective functions for the estimation problems involving Y and W, respectively. The moment condition $\psi_1(S_i, \phi)$ corresponds to the FOCs given in equations (19) and (20). Similarly, $\psi_2(S_i, \phi)$ is the FOC for the estimation of (α_0, β_0), and so on. The moment conditions $\psi_6(S_i, \phi)$ and $\psi_7(S_i, \phi)$ account for the sampling variation in X_i in obtaining $\hat{\tau}_{Y|Z}$ and $\hat{\tau}_{W|Z}$.

The asymptotic distribution of any parametric LATE estimator that uses consistent estimators of $\tau_{Y|Z}$ and $\tau_{W|Z}$ can be derived for a known joint asymptotic distribution of these estimators $\hat{\tau}_{Y|Z}$

and $\hat{\tau}_{W|Z}$, which satisfy:

$$\sqrt{N}\left[\begin{pmatrix}\hat{\tau}_{Y|Z}\\ \hat{\tau}_{W|Z}\end{pmatrix}-\begin{pmatrix}\tau_{Y|Z}\\ \tau_{W|Z}\end{pmatrix}\right]\xrightarrow{d}\text{Normal}(0,\Omega),$$

where Ω is the 2×2 variance-covariance matrix corresponding to the lower right block of $A^{-1}VA^{-1}$. Given $\hat{\tau}_{LATE}=\hat{\tau}_{Y|Z}/\hat{\tau}_{W|Z}$, we can apply the delta method to obtain

$$\begin{aligned}\mathbb{AVAR}\left(\hat{\tau}_{DRLATE}\right) &= \left(\frac{1}{\tau_{W|Z}^2}\right)\mathbb{AVAR}\left(\hat{\tau}_{Y|Z}\right)+\left(\frac{\tau_{Y|Z}}{(\tau_{W|Z})^2}\right)^2\mathbb{AVAR}\left(\hat{\tau}_{Y|Z}\right)\\ &\quad-\left(\frac{2\tau_{Y|Z}}{(\tau_{W|Z})^3}\right)\mathbb{ACOV}\left(\hat{\tau}_{Y|Z},\hat{\tau}_{W|Z}\right).\end{aligned}\tag{30}$$

The three asymptotic variance terms are available from $\hat{\Omega}/N$, and the other terms are easily estimated by plugging in $\hat{\tau}_{Y|Z}$ and $\hat{\tau}_{W|Z}$.

5 A Test Comparing LATT and ATT Estimators

In textbook treatment of instrumental variables, where the treatment effect is taken to be constant, it is fairly common to construct a Hausman (1978) test for comparing the IV estimator with the OLS estimator of the coefficients on the endogenous explanatory variable, W. The idea is that, with good controls in X, maybe W is unconfounded conditional on X, and the instrumental variables are not needed. In practice, with cross-sectional data one uses a heteroskedasticity-robust version of the Hausman test that is easily implemented using a control function regression; see, for example, Wooldridge (2010, Section 6.3.1). In the traditional setting, efficiency considerations are the primary reason for preferring OLS if the Hausman test does not reject the null that W is unconfounded: the OLS estimator is typically much more precise than the IV estimator.

Efficiency remains a valid consideration when treatment effects are heterogeneous, as in the current setting, but the usual Hausman test is no longer valid because OLS and IV estimands constitute different weighted averages of heterogeneous treatment effects even under the null. Instead, if W is unconfounded conditional on X then, with sufficient overlap, one can identify the aver-

age treatment effect on the treated (ATT) without requiring an instrumental variable. As shown by DHL (2014), under one-sided noncompliance, LATT is the same as ATT. Therefore, it makes sense to use doubly robust estimators of the ATT (DR ATT) that do not use an instrument and compare that with the DR LATT estimates.

The ATT parameter is

$$\tau_{ATT} = \mathbb{E}\left[Y(1) - Y(0) \mid W = 1\right]. \tag{31}$$

Following Słoczyński and Wooldridge (2018), we use DR estimators of τ_{ATT} that are natural given the form of the DR LATT estimators in Section 3. We no longer need an instrument propensity score. Instead, let $F\left(x, \gamma\right)$ be a model of the treatment propensity score, $\mathbb{P}\left(W = 1|X = x\right)$. Given a random sample of size N, let $\hat{\gamma}$ be the (quasi-) MLE based on the Bernoulli log likelihood. As before, a typical choice of $F\left(x, \gamma\right)$ is a flexible logistic function.

Under the assumption that W is independent of $Y(0)$ conditional on X and the overlap assumption

$$\mathbb{P}\left(W = 1|X = x\right) < 1 \text{ for almost all } x \in \mathcal{X}, \tag{32}$$

we can obtain DR estimators of ATT using quasi-MLE in the LEF with a canonical link function.

The conditional mean we need to estimate is $\mathbb{E}\left[Y(0)|X = x\right]$, and we again take the model to have the index form, $m(\alpha_0 + x\beta_0)$ (reusing earlier notation). As before, we can choose $m(\alpha_0 + x\beta_0)$ to reflect the nature of the outcome variable $Y(0)$. To stay within the DR class of estimators using IPWRA, $m(\cdot)$ will be the identity, logistic, or exponential function in the vast majority of applications.

For consistent estimation of τ_{ATT} we can get by with the conditional mean version of unconfoundedness of W conditional on X,

$$\mathbb{E}\left[Y(0)|W, X\right] = \mathbb{E}\left[Y\left(0\right)|X\right].$$

Under this assumption, if the mean is correctly specified then α_0 and β_0 are identified by

$$\mathbb{E}\left(Y|W=0,X\right)=\mathbb{E}\left[Y(0)|X\right]=m(\alpha_0+X\beta_0).$$

Letting $q\left(y,m\right)$ be the quasi-log likelihood function, $\hat{\alpha}_0$ and $\hat{\beta}_0$ solve the weighted QMLE problem

$$\max_{a_0,b_0}\sum_{i=1}^{N}\frac{F(X_i\hat{\gamma})}{1-F(X_i\hat{\gamma})}\left(1-W_i\right)q\left(Y_i,m(a_0+X_ib_0)\right),$$

where the estimation is done on the control sample and the weighting ensures the DR property; see Słoczyński and Wooldridge (2018). Given the estimates, the DR estimator of τ_{ATT} is

$$\hat{\tau}_{DRATT}=\bar{Y}_1-N_1^{-1}\sum_{i=1}^{N}W_i\cdot m(\hat{\alpha}_0+X_i\hat{\beta}_0), \tag{33}$$

where $\bar{Y}_1$ is the average outcome over the treated units and N_1 is now the number of treated (not eligible) units. The estimator in (33) has a simple interpretation as an imputation estimator, as the second term is a DR estimator of $\mathbb{E}\left[Y(0)|W=1\right]$ obtained by first imputing $\mathbb{E}\left[Y_i\left(0\right)|W_i=1,X_i\right]$ using the mean function estimated from the $W_i=0$ units. This DR estimator is pre-programmed in popular statistics and econometrics packages.

Given $\hat{\tau}_{DRLATT}$ from Section 3 and $\hat{\tau}_{DRATT}$ in (33), we can test the null hypothesis that treatment is unconfounded given X, provided the instrument Z is such that one-sided noncompliance holds so that $\tau_{LATT}=\tau_{ATT}$. A formal comparison is based on the statistic

$$\frac{\hat{\tau}_{DRLATT}-\hat{\tau}_{DRATT}}{\operatorname{se}\left(\hat{\tau}_{DRLATT}-\hat{\tau}_{DRATT}\right)}. \tag{34}$$

Under the null hypothesis, we cannot say that $\hat{\tau}_{DRATT}$ is the efficient estimator in a suitable class that includes $\hat{\tau}_{DRLATT}$, and so the standard error $\operatorname{se}\left(\hat{\tau}_{DRLATT}-\hat{\tau}_{DRATT}\right)$ does not simplify. Nevertheless, bootstrapping is computationally feasible, or one can extend the calculations in Section 4 to obtain an analytical standard error. Even without one-sided noncompliance, a similar test can also be constructed to assess treatment effect heterogeneity by comparing DR LATE and DR LATT, or IV and DR LATE, or IV and DR LATT estimates.

6 Empirical Applications

In this section we reanalyze the data in Abadie (2003) and Taubman et al. (2014) to illustrate our new doubly robust estimators.

6.1 The Effects of 401(k) Retirement Plans

The 401(k) retirement plans were introduced in the US to increase saving for retirement by allowing tax advantages for the contributions to the retirement account. The policy-relevant empirical question is whether the 401(k) program is effective for increasing savings or only crowds out other personal saving. Since the individuals who participate in 401(k) plans are likely to have different saving preferences than non-participating individuals, a simple comparison of savings of the two groups is likely to provide an upward-biased estimate of the true effect. Different from other saving plans, 401(k) participation requires eligibility which, in turn, is determined by the employer. Abadie (2003) argues, following Poterba, Venti and Wise (1994, 1995), that 401(k) eligibility can be used as a conditionally independent instrument to estimate the effects of 401(k) participation on savings. Several recent papers have revisited this empirical question and estimated the LATE using the same instrument by different methods (e.g., Belloni et al., 2017; Chernozhukov et al., 2018; Heiler, 2022; Sant'Anna, Song and Xu, 2022). Since our new doubly robust estimator relies on the same identifying assumptions, we also choose to reanalyze the effect of 401(k) participation. We use the same dataset as Abadie (2003). The data consists of a sample of 9,275 households from the Survey of Income and Program Participation (SIPP) of 1991. As outcomes, we consider net financial assets (in US dollars) and participation in an individual retirement account (IRA), which is another popular tax-deferred saving plan in the US. The treatment is an indicator for participation in a 401(k) plan. The set of control variables consists of family income, age, marital status, and family size. Age enters the conditional mean functions quadratically.

Table 2 reports the estimates of the parameters of interest together with asymptotic standard errors for both outcome variables. The first two rows display the coefficient estimates for 401(k)

participation from OLS and IV estimation, respectively. OLS estimates of the participation coefficient for both regressions are positive and significant. However, as mentioned above, due to unobserved preferences for saving, it is likely that these overestimate the true effect even after conditioning on various individual characteristics. The usual IV estimates are indeed much smaller than the OLS estimates, although they are still positive and significant. Moreover, the Hausman test for the absence of endogeneity strongly rejects for both outcomes.

Next, we report the average treatment effect (ATE) of the participation in a 401(k) plan, as estimated by IPWRA. The ATE is identified if there are no unmeasured confounders. If this is not the case, similar to the OLS coefficient, we also expect the ATE estimates to be biased. The ATE estimate of 401(k) participation on total net financial assets is smaller than the OLS estimate and slightly larger than the IV estimate. The ATE on having an IRA account is also larger than the corresponding IV estimate but closer to the OLS estimate. Following the estimated ATEs, the LATE estimates are also reported. Specifically, we estimate the LATE using IPW, RA, AIPW, and our proposed IPWRA estimator. The LATE is identified if the assumptions discussed in Section 2 are satisfied. IPW and AIPW estimates of the LATE on net financial assets are small in magnitude and very imprecisely estimated. On the other hand, the RA and IPWRA estimates are closer to the IV estimates and relatively very precise. The IPW estimate of the LATE on IRA participation is insignificant while the estimates based on RA, IPWRA, and AIPW are significant and rather similar in magnitude. Additionally, we estimate the ATT and LATT of participation in a 401(k) plan. Since there is one-sided noncompliance, we can test the equality of ATT and LATT as discussed in Section 5. For net financial assets, we cannot reject the equality of the two treatment effects, ATT and LATT. This suggests that participation in a 401(k) plan might be unconfounded conditional on the set of variables that we control for, following Abadie (2003). However, the ATT and LATT for the probability of having an IRA are statistically different from each other based on our proposed test, which underscores the importance of IV estimation under the maintained assumptions of Section 2.

The results of this application are reassuring given that our proposed DR method provides

estimates in a reasonable range with good precision. At the same time, the AIPW estimate of the LATE on net financial assets – that is, the other doubly robust estimate – is very imprecise, which is in line with previous criticism of AIPW estimation in other contexts (Kang and Schafer, 2007). The RA estimate of the LATE has a smaller standard error but does not enjoy the double robustness property. The differences between the LATE estimates for the binary outcome, IRA participation, are minor, with the exception of the small and insignificant IPW estimate.

6.2 The Effects of Medicaid

In our second empirical application, we revisit Taubman et al. (2014)'s analysis of the data from the Oregon Health Insurance Experiment. In 2008, the state of Oregon decided to offer about 10,000 spots in Medicaid using a lottery that randomly selected eligible households from a larger pool of applicants. Finkelstein et al. (2012), Taubman et al. (2014), Denteh and Liebert (2022), and Johnson, Cao and Kang (2022), among others, used Oregon's lottery assignment as a binary instrument to assess the effects of Medicaid on various outcomes related to health and healthcare utilization. Here, following much of the previous work, we focus on emergency room (ER) visits at the extensive and intensive margins. In other words, we use our proposed method to analyze the effects of Medicaid on a binary outcome indicating any ER visits in the study period as well as on a count outcome indicating the number of visits.

In what follows, we use Taubman et al. (2014)'s administrative data with over 24,000 observational units. The endogenous treatment variable is defined as "ever enrolled in Medicaid" during the study period. Table 3 reports a number of estimates together with their associated standard errors. OLS refers to the coefficient estimates for the endogenous treatment variable. ATE and ATT refer to the IPWRA estimates of these parameters. IV refers to the coefficient estimates for the endogenous treatment variable, which use the binary lottery instrument. Finally, LATE and LATT are estimated using our proposed estimators. Following Taubman et al. (2014), all regressions include indicators for different numbers of household members on the lottery list, which is necessary for instrument validity, and past outcome data, which should improve the precision of

the final estimates.

It turns out that the OLS, ATE, and ATT estimates (and their standard errors) are almost identical for the binary outcome and quite similar for the count outcome, although in the latter case the OLS estimates are smaller than those of the ATE and ATT. In any case, the estimates show a significant positive correlation between Medicaid and ER utilization both on the extensive and intensive margins. Due to treatment endogeneity, however, these estimates cannot be interpreted as causal effects. The last three columns of Table 3 take this endogeneity into account and rely on the instrumental variable for the identification of causal effects. It turns out that these estimates are much smaller than the OLS, ATE, and ATT estimates, especially in the case of the binary outcome, where the estimates are now roughly half as large. Like in Taubman et al. (2014), and unlike in an earlier analysis by Finkelstein et al. (2012) that relied on mail survey data, these estimates are also significantly different from zero. Our analysis, however, also reveals an interesting dimension of treatment effect heterogeneity: LATT, the effect on the treated compliers, appears to be larger than the usual LATE. A formal comparison of the two objects yields a *p*-value of about 0.04 for the binary outcome and 0.47 for the count outcome. The fact that the LATT may be larger than the LATE is likely due to treatment effect heterogeneity across households of different sizes: effects of Medicaid on ER utilization are more pronounced in larger households (cf. Denteh and Liebert, 2022), which are also more likely to be treated given the lottery design.

7 Simulations

In this section, we conduct a Monte Carlo study to assess the bias and precision of our DR LATE estimator in comparison with other existing estimators. In particular, we focus on the effect of certain types of misspecification on the bias and precision. To eliminate to some extent the arbitrariness in choosing the data-generating process, we generate our Monte Carlo samples to mimic some statistical features of the 401(k) dataset we used in Section 6.1. We draw 1,000 samples with $N = 1{,}000$ and the same number of samples with $N = 4{,}000$ observations. Each sample

is constructed in the following steps. First, we draw two random variables from a bivariate normal distribution. The parameters of the bivariate normal distribution are set equal to the empirical means and covariances of age and log income in the 401(k) data. The simulated log income is then exponentiated to generate the income variable. As an additional covariate, we take the square of the simulated age variable. Thus, our full set of covariates, X, includes three variables: income, age, and age squared. The instrumental variable Z is generated according to

$$Z = \mathbb{1}\left(\Lambda\left(\gamma_0 + X\gamma_x\right) > U_z\right), \tag{35}$$

where $\gamma = (\gamma_0, \gamma_x)$ corresponds to the estimated coefficient vector from a logit regression of 401(k) eligibility on a constant, income, age, and age squared using the original data (see column (1) of Table 6 in the Appendix). The random variable U_z is drawn from the standard uniform distribution and $\Lambda(\cdot)$ is the logistic cdf. Then, we construct $D(1)$ as follows:

$$D(1) = \mathbb{1}\left(\Lambda\left(\omega_0 + X\delta_0\right) > U_1\right), \tag{36}$$

where the coefficients are from a logit regression using observations with $Z_i = 1$ (column (2) of Table 6 in the Appendix) in the original data and U_1 is drawn from the standard uniform distribution. Finally, we generate two outcome variables. One mimics the continuous outcome variable, net financial assets, and the other mimics the binary outcome variable, participation in IRA. The continuous outcome $Y(z)$ is generated using the following linear model:

$$Y(z) = \alpha_z + X\beta_z + \varepsilon_z \quad \text{for } z = 0, 1. \tag{37}$$

We use the coefficients from two separate regressions of the outcome variable on the set of covariates for $Z_i = 1$ and $Z_i = 0$ subsamples in the original data (columns (3) and (4) of Table 6 in the Appendix). The error terms ε_z are drawn from a normal distribution with mean zero and variance σ_z^2, where σ_z^2 is the mean squared residual from the regression for $Z_i = z$. The binary outcome

variable is constructed similarly to the potential treatment variable using logit link:

$$Y(z) = \mathbb{1}\left(\Lambda\left(\alpha_z + X\beta_0\right) > U_y\right) \quad \text{for } z = 0, 1, \tag{38}$$

with coefficients from two separate logistic regressions of the binary indicator of IRA participation on the set of covariates (columns (5) and (6) of Table 6 in the Appendix) and U_y drawn from the standard uniform distribution. For the simulated data, the "true" values of the LATE are \$8,816.5 for net financial assets and 0.036 for the probability of IRA participation.

For each simulated sample, in addition to our proposed method, we estimate the LATE using IV, RA, IPW, and AIPW. The RA estimator of the LATE is constructed similarly to our proposed estimator with the crucial difference that the objective functions are not weighted. The IPW and AIPW estimators of the LATE are the ratios of two IPW and AIPW estimators of ATEs of Z on Y and W, respectively.

In general, the LATE estimators based on the identification result in (5) require estimation of four conditional means. However, as mentioned earlier, since in the original 401(k) data and in our simulation design $Z = 0$ implies $D = 0$, the second term in the denominator of (5) is zero and we do not need to estimate (9). Thus, for the RA approach, we estimate the two conditional means in the numerator of (5) by two separate linear (logistic) regressions of continuous (binary) $Y(z)$ using observations from subsamples with $Z_i = 1$ and $Z_i = 0$, respectively. Similarly, the first conditional mean in the denominator is estimated by a logistic regression for the subsample with $Z_i = 1$. On the other hand, the IPW approach requires estimating a binary response model for the instrument propensity score defined in (13). Thus, we estimate the instrument propensity score by a logistic regression. Our proposed IPWRA method requires the same conditional mean specifications as the RA approach and additionally the specification of (13) to construct the weights. For IPWRA, we estimate the conditional means in the numerator of (5) by two separate *weighted* linear (logistic) regressions of continuous (binary) $Y(z)$ with the weight equal to the inverse of the estimated probability of being eligible or not, using the subsample with $Z_i = 1$ or $Z_i = 0$, respectively. Finally, AIPW requires the same set of model specifications as our proposed method, although the

regressions are not weighted.

In our Monte Carlo study, we consider estimators *(i)* when the required models are all correctly specified, *(ii)* when models for (7)–(10) are misspecified, and *(iii)* when the model for (13) is misspecified. Correct specifications for these estimators mean that we use the correct set of covariates for all the regressions, namely simulated income, age, and age squared. Misspecification of a certain model means that the set of regressors does not include age squared.

Tables 4 and 5 present the biases, root mean squared errors of the LATE estimators, and the empirical coverage rates for nominal 95% confidence intervals under the different model specifications for the dependent variables net financial assets and IRA participation, respectively.

Table 4 suggests that, when the relevant models are correctly specified, that is, all the confounding factors are controlled for, the bias is highest for the linear IV estimates. This coincides with the fact that the IV estimand is not equal to the LATE in the case of a conditionally independent instrument (e.g., Słoczyński, 2021; Blandhol et al., 2022). The RA estimator has the smallest bias and RMSE when all the models are correctly specified. For the smaller sample size, our proposed estimator has the second smallest bias and RMSE. For the larger sample size, the bias estimates are very close for IPWRA and AIPW but the precision of our estimator is better than that of IPW and AIPW.

The second block of Table 4 presents the results for the first type of misspecification. The IV estimator becomes heavily biased when we do not control for age squared. The effect of omitting this variable when estimating the conditional mean functions in (7)–(10) is similar for the RA estimator, which is also very biased. The IPW estimator is not affected by this type of misspecification since it only requires the correct specification of the model for the instrument propensity score. The doubly robust estimators, IPWRA and AIPW, are not seriously affected by this type of misspecification either, as predicted by theory. Our proposed method has the smallest bias and RMSE for sample size $N = 1{,}000$, and a slightly larger bias – but still the smallest RMSE – for sample size $N = 4{,}000$.

Finally, we investigate the bias and RMSE for the case where the instrument propensity score

in (13) is estimated without the squared term. As expected, the only estimator that is severely affected by this misspecification is IPW. The doubly robust methods continue to have reasonably small biases. IPWRA has the smaller bias and RMSE for $N = 1{,}000$ and the smaller RMSE for $N = 4{,}000$, as in other cases. The results demonstrate that the double robustness of our proposed method is achieved without significant sacrifices in terms of precision. Coverage rates are close to the nominal coverage rate for both DR estimators in all cases. For IV, RA, and IPW, coverage rates are sometimes substantially lower than 95% when the estimators are otherwise biased.

Table 5 revisits the same measures of estimator performance as Table 4 while focusing on the binary outcome. Unlike in Table 4, the differences between the estimators that we consider are very minor, even in cases when one of the underlying models is misspecified. This is likely due to the fact that, as shown in Table 6 in the Appendix, age squared is insignificant in the logit regressions of the binary outcome, IRA participation. It follows that omitting this variable should have a smaller effect on estimator performance, relative to Table 4.

In any case, even though the differences in estimator performance, as reported in Table 5, are very minor, it is also clear that the performance of RA, IPWRA, and AIPW is better overall than that of IV and IPW. This is again reassuring, given our general preference for IPWRA estimation.

8 Conclusion

In this paper we develop a framework for doubly robust (DR) estimation of local average treatment effects, which uses quasi-likelihood methods weighted by the inverse of the instrument propensity score. These estimators are commonly referred to as inverse probability weighted regression adjustment (IPWRA). We argue that our estimators have appealing small sample properties relative to competing methods, such as augmented inverse probability weighting (AIPW). We discuss inference for IPWRA estimators and propose a DR version of a Hausman test previously suggested by DHL (2014), which compares two estimates of the average treatment effect on the treated (ATT) in settings with one-sided noncompliance.

We discuss two empirical applications. First, we revisit Abadie (2003)'s study of the effects of 401(k) retirement plans, and demonstrate that some of the conclusions are different dependent on whether one uses AIPW or IPWRA, which are the two major classes of DR estimators. While we obviously do not know which estimate is closer to the true effect of interest, we note that our preferred estimate is much more precise. Second, we reanalyze Taubman et al. (2014)'s sample from the Oregon Health Insurance Experiment. Focusing on the effect of Medicaid on emergency room visits, we provide evidence that the local average treatment effect on the treated (LATT) is larger than the usual local average treatment effect (LATE), at least along the extensive margin. We conclude the paper with a Monte Carlo study that demonstrates the very good finite sample properties of our proposed IPWRA estimator.

References

Abadie, Alberto, "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 2003, *113* (2), 231–263.

Angrist, Joshua D. and William N. Evans, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 1998, *88* (3), 450–477.

_ , Guido W. Imbens, and Donald B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, *91* (434), 444–455.

Belloni, Alexandre, Victor Chernozhukov, Iván Fernández-Val, and Christian Hansen, "Program Evaluation and Causal Inference with High-Dimensional Data," *Econometrica*, 2017, *85* (1), 233–298.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When Is TSLS Actually LATE?," 2022.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/Debiased Machine Learning for Treatment and Structural Parameters," *Econometrics Journal*, 2018, *21* (1), C1–C68.

Denteh, Augustine and Helge Liebert, "Who Increases Emergency Department Use? New Insights from the Oregon Health Insurance Experiment," 2022.

Donald, Stephen G., Yu-Chin Hsu, and Robert P. Lieli, "Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT," *Journal of Business & Economic Statistics*, 2014, *32* (3), 395–415.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group, "The Oregon Health Insurance Experiment: Evidence from the First Year," *Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.

Frölich, Markus, "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 2007, *139* (1), 35–75.

_ **and Michael Lechner**, "Exploiting Regional Treatment Intensity for the Evaluation of Labor Market Policies," *Journal of the American Statistical Association*, 2010, *105* (491), 1014–1029.

Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica*, 1978, *46* (6), 1251–1271.

Heiler, Phillip, "Efficient Covariate Balancing for the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 2022, *40* (4), 1569–1582.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, New York: Cambridge University Press, 2015.

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

Johnson, Michael, Jiongyi Cao, and Hyunseung Kang, "Detecting Heterogeneous Treatment Effects with Instrumental Variables and Application to the Oregon Health Insurance Experiment," *Annals of Applied Statistics*, 2022, *16* (2), 1111–1129.

Kang, Joseph D. Y. and Joseph L. Schafer, "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data," *Statistical Science*, 2007, *22* (4), 523–539.

Ogburn, Elizabeth L., Andrea Rotnitzky, and James M. Robins, "Doubly Robust Estimation of the Local Average Treatment Effect Curve," *Journal of the Royal Statistical Society, Series B*, 2015, *77* (2), 373–396.

Poterba, James M., Steven F. Venti, and David A. Wise, "401(k) Plans and Tax-Deferred Saving," in David A. Wise, ed., *Studies in the Economics of Aging*, Chicago–London: University of Chicago Press, 1994, pp. 105–142.

_ **,** _ **, and** _ , "Do 401(k) Contributions Crowd Out Other Personal Saving?," *Journal of Public Economics*, 1995, *58* (1), 1–32.

Sant'Anna, Pedro H. C., Xiaojun Song, and Qi Xu, "Covariate Distribution Balance via Propensity Scores," *Journal of Applied Econometrics*, 2022, *37* (6), 1093–1120.

Singh, Rahul and Liyang Sun, "Double Robustness for Complier Parameters," 2022.

Słoczyński, Tymon, "When Should We (Not) Interpret Linear IV Estimands as LATE?," 2021.

_ **and Jeffrey M. Wooldridge**, "A General Double Robustness Result for Estimating Average Treatment Effects," *Econometric Theory*, 2018, *34* (1), 112–133.

Sun, Baoluo and Zhiqiang Tan, "High-Dimensional Model-Assisted Inference for Local Average Treatment Effects with Instrumental Variables," *Journal of Business & Economic Statistics*, 2022, *40* (4), 1732–1744.

Tan, Zhiqiang, "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 2006, *101* (476), 1607–1618.

Taubman, Sarah L., Heidi L. Allen, Bill J. Wright, Katherine Baicker, and Amy N. Finkelstein, "Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment," *Science*, 2014, *343* (6168), 263–268.

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

Wooldridge, Jeffrey M., "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 2007, *141* (2), 1281–1301.

_ , *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge–London: MIT Press, 2010.

Appendix

Table 1: Combinations of QLLFs and Canonical Link Functions

Support Restrictions	**Mean Function**	**Quasi-LLF**
None	Linear	Gaussian
$Y(w) \in [0, 1]$ (binary, fractional)	Logistic	Bernoulli
$Y(w) \in [0, B]$ (count, corner)	Logistic	Binomial
$Y(w) \geq 0$ (count, continuous, corner)	Exponential	Poisson

Table 2: Estimates of the Effects of 401(k) Participation

	(A) Net financial assets		(B) IRA	
	Estimate	**Std. err.**	**Estimate**	**Std. err.**
OLS	13,527	(1,810)	0.0569	(0.0103)
IV	9,419	(2,152)	0.0274	(0.0132)
	Hausman test			
H_0: OLS = IV	$p = 0.0004$		$p = 0.0004$	
	ATE		**ATE**	
IPWRA	10,767	(1,772)	0.0554	(0.0096)
	LATE		**LATE**	
IPW	3,994	(4,891)	0.0165	(0.0135)
RA	8,467	(1,991)	0.0338	(0.0128)
IPWRA	8,046	(2,587)	0.0361	(0.0128)
AIPW	5,416	(4,176)	0.0404	(0.0131)
	ATT		**ATT**	
IPWRA	12,673	(3,329)	0.0697	(0.0110)
	LATT		**LATT**	
IPWRA	10,918	(3,709)	0.0413	(0.0143)
	Hausman test			
H_0: ATT = LATT	$p = 0.457$		$p = 0.001$	

Notes: The data are Abadie (2003)'s subsample of the Survey of Income and Program Participation (SIPP) of 1991. The sample size is 9,275. The outcomes are net financial assets (Panel A) and a binary indicator for participation in IRAs (Panel B). The treatment is an indicator for 401(k) participation. The instrument is an indicator for 401(k) eligibility. The set of covariates consists of family income, age, age squared, marital status, and family size. "OLS" and "IV" are the estimates of the coefficient on the endogenous treatment with covariates (and instrument) listed above. The remaining estimators are defined in the main text. Standard errors are in parentheses. For OLS and IV, we report robust standard errors. For the remaining estimators, our standard errors follow from the GMM framework in Section 4.

Table 3: Estimates of the Effects of Medicaid

	OLS	ATE	ATT	IV	LATE	LATT
Outcome: ER visits (any)						
Estimate	0.1355	0.1348	0.1360	0.0697	0.0696	0.0812
Std. err.	(0.0068)	(0.0069)	(0.0069)	(0.0239)	(0.0238)	(0.0246)
Outcome: ER visits (#)						
Estimate	0.4611	0.5320	0.5622	0.3880	0.4508	0.4701
Std. err.	(0.0340)	(0.0346)	(0.0385)	(0.1070)	(0.1254)	(0.1141)

Notes: The data are Taubman et al. (2014)'s sample from the Oregon Health Insurance Experiment. The sample sizes are 24,646 (top panel) and 24,615 (bottom panel). The outcomes are an indicator for any ER visits (top panel) and the (censored) number of ER visits (bottom panel) in the study period. The treatment is an indicator for Medicaid coverage. The instrument is an indicator for whether a given household was selected by the Medicaid lottery. The set of covariates consists of indicators for different numbers of household members on the lottery list (both panels), an indicator for any ER visits before the randomization (top panel), and the number of ER visits before the randomization (bottom panel). "OLS" and "IV" are the estimates of the coefficient on the endogenous treatment with covariates (and instrument) listed above. The remaining estimators are defined in the main text. Standard errors, clustered on the household identifier, are in parentheses. For OLS and IV, we report cluster-robust standard errors. For the remaining estimators, our standard errors follow from the GMM framework in Section 4.

Table 4: Simulation Results for the Continuous Outcome Variable

	All Correct			(7)–(10) **misspecified**			(13) **misspecified**		
	Bias	**RMSE**	**Cov.**	**Bias**	**RMSE**	**Cov.**	**Bias**	**RMSE**	**Cov.**
N=1,000									
IV	271.43	6,163.19	95.3	–1,544.41	6,395.32	94.5	271.43	6,163.19	95.3
RA	127.69	6,169.94	95.5	–1,724.20	6,445.16	94.2	127.69	6,169.94	95.5
IPW	162.49	6,958.60	95.8	162.49	6,958.60	95.8	–1,549.48	7,033.47	94.1
IPWRA	159.24	6,300.47	95.4	103.68	6,306.11	95.3	140.70	6,258.49	95.3
AIPW	195.36	6,418.33	95.6	170.13	6,439.62	95.4	170.75	6,304.15	95.4
N=4,000									
IV	114.57	3,097.13	94.8	–1,734.22	3,565.59	89.7	114.57	3,097.13	94.8
RA	–45.16	3,119.43	94.4	–1,907.94	3,662.62	89.2	–45.16	3,119.43	94.4
IPW	–60.69	3,381.29	94.4	–60.69	3,381.29	94.4	–1738.43	3,782.54	91.2
IPWRA	–74.71	3,155.63	94.8	–102.96	3,161.44	94.8	–69.52	3,152.04	94.6
AIPW	–74.41	3,174.61	94.8	–95.49	3,183.18	94.8	–67.71	3,160.11	94.7

Notes: The details of the simulation design are provided in Section 7. Results are based on 1,000 replications. "RMSE" is the root mean squared error of an estimator. "Cov." is the coverage rate for a nominal 95% confidence interval. "IV" is the IV estimate of the coefficient on the endogenous treatment, controlling for X. The remaining estimators are defined in the main text. To calculate the coverage rate, we use robust standard errors (IV) or standard errors that follow from the GMM framework in Section 4 (remaining estimators).

Table 5: Simulation Results for the Binary Outcome Variable

	All Correct			(7)–(10) **misspecified**			(13) **misspecified**		
	Bias	**RMSE**	**Cov.**	**Bias**	**RMSE**	**Cov.**	**Bias**	**RMSE**	**Cov.**
N=1,000									
IV	–0.0005	0.0421	93.5	–0.0048	0.0423	93.6	–0.0005	0.0421	93.5
RA	–0.0031	0.0408	94.0	–0.0030	0.0407	93.3	–0.0031	0.0408	94.0
IPW	–0.0031	0.0441	94.0	–0.0072	0.0438	93.4	–0.0072	0.0438	93.4
IPWRA	–0.0033	0.0410	93.5	–0.0034	0.0411	93.9	–0.0033	0.0409	93.7
AIPW	–0.0033	0.0411	93.6	–0.0033	0.0411	93.8	–0.0032	0.0409	93.8
N=4,000									
IV	0.0028	0.0199	95.0	–0.0017	0.0198	95.6	0.0028	0.0199	95.0
RA	–0.0001	0.0191	95.1	0.0000	0.0192	95.5	–0.0001	0.0191	95.1
IPW	–0.0001	0.0199	95.6	–0.0042	0.0203	95.1	–0.0042	0.0203	95.1
IPWRA	–0.0002	0.0192	95.7	–0.0002	0.0192	95.6	–0.0002	0.0192	95.2
AIPW	–0.0002	0.0192	95.6	–0.0002	0.0191	95.6	–0.0002	0.0192	95.3

Notes: The details of the simulation design are provided in Section 7. Results are based on 1,000 replications. "RMSE" is the root mean squared error of an estimator. "Cov." is the coverage rate for a nominal 95% confidence interval. "IV" is the IV estimate of the coefficient on the endogenous treatment, controlling for X. The remaining estimators are defined in the main text. To calculate the coverage rate, we use robust standard errors (IV) or standard errors that follow from the GMM framework in Section 4 (remaining estimators).

Table 6: Coefficient Values for the Data-Generating Process in Section 7

	401(k) eligibility	401(k) participation	Net total financial assets		IRA participation	
	(1)	(2)	(3)	(4)	(5)	(6)
Household income	0.0000232	0.0000154	1.134	0.762	0.0000318	0.0000342
	(23.00)	(9.18)	(25.67)	(23.92)	(18.87)	(20.87)
Age (minus 25)	0.0581	–0.0285	–106.6	–557.4	0.0420	0.0665
	(7.25)	(–2.03)	(–0.25)	(–2.42)	(2.63)	(4.99)
Age (minus 25) squared	–0.00158	0.000699	41.36	38.28	0.000211	–0.000267
	(–7.45)	(1.88)	(3.68)	(6.32)	(0.52)	(–0.82)
Constant	–1.727	0.387	–36377.2	–19452.5	–3.148	–3.653
	(–24.55)	(3.08)	(–9.68)	(–10.03)	(–20.02)	(–28.04)
Observations	9,275	3,637	3,637	5,638	3,637	5,638
Sample	Full	$Z = 1$	$Z = 1$	$Z = 0$	$Z = 1$	$Z = 0$
Method	Logit	Logit	OLS	OLS	Logit	Logit

Notes: The table presents coefficient estimates obtained using Abadie (2003)'s subsample of the Survey of Income and Program Participation (SIPP) of 1991, as previously analyzed in Table 2. The instrument, Z, is an indicator for 401(k) eligibility. The estimates in column (1) are used as coefficient values for equation (35). The estimates in column (2) are used as coefficient values for equation (36). The estimates in columns (3) and (4) are used as coefficient values for equation (37). The estimates in columns (5) and (6) are used as coefficient values for equation (38). t statistics are in parentheses.