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


Tymon Słoczyński[1] Derya Uysal [2] Jeffrey M. Wooldridge[3]


EEA-ESEM

August 29, 2023


[1]Brandeis University
[2]University of Munich
[2]Michigan State University

# Motivation

Many applied papers in the reduced-form literature use **the IV/2SLS estimator** to estimate:

$$Y = \alpha + \tau W + X\beta + u, \qquad (1)$$

where $W$ is an endogenous binary "treatment" and $X$ is a vector of additional covariates. Also, $Z$ is a binary instrumental variable for $W$.

# Motivation

Many applied papers in the reduced-form literature use **the IV/2SLS estimator** to estimate:

$$Y = \alpha + \tau W + X\beta + u, \tag{1}$$

where $W$ is an endogenous binary "treatment" and $X$ is a vector of additional covariates. Also, $Z$ is a binary instrumental variable for $W$.

Following a series of papers by Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist, Imbens, and Rubin (1996), most papers interpret an IV estimate of $\tau$ as **the local average treatment effect (LATE)**.

# Motivation

Many applied papers in the reduced-form literature use **the IV/2SLS estimator** to estimate:

$$Y = \alpha + \tau W + X\beta + u, \tag{1}$$

where $W$ is an endogenous binary "treatment" and $X$ is a vector of additional covariates. Also, $Z$ is a binary instrumental variable for $W$.

Following a series of papers by Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist, Imbens, and Rubin (1996), most papers interpret an IV estimate of $\tau$ as **the local average treatment effect (LATE)**.

However, $\tau$ is actually equivalent to LATE when there are no covariates in the model (Imbens and Angrist 1994); when $X$ is nonempty, $\tau$ is a weighted average of $X$-specific LATEs (Angrist and Imbens 1995) with rather undesirable weights (Słoczyński 2021).

# Motivation

Many applied papers in the reduced-form literature use **the IV/2SLS estimator** to estimate:

$$Y = \alpha + \tau W + X\beta + u, \tag{1}$$

where $W$ is an endogenous binary "treatment" and $X$ is a vector of additional covariates. Also, $Z$ is a binary instrumental variable for $W$.

Following a series of papers by Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist, Imbens, and Rubin (1996), most papers interpret an IV estimate of $\tau$ as **the local average treatment effect (LATE)**.

However, $\tau$ is actually equivalent to LATE when there are no covariates in the model (Imbens and Angrist 1994); when $X$ is nonempty, $\tau$ is a weighted average of $X$-specific LATEs (Angrist and Imbens 1995) with rather undesirable weights (Słoczyński 2021).

**This paper:** Doubly Robust IV Estimation to estimate the LATE with covariates

# Our contributions

- A new class of doubly robust estimators of LATE and LATT
  - simple to implement
  - avoids the shortcomings of other doubly robust methods
- Provide proofs of double robustness and valid asymptotic inference
- Hausman-type tests to assess the unconfoundedness of the treatment and treastment heterogeneity
- Monte Carlo Experiment for comparison of several LATE-estimators
- Empirical Examples

# Related Literature

## Doubly Estimation Estimation

- Scharfstein et al. (1999) introduced the concept
- Several doubly robust estimators for different parameters are proposed: among others Tan (2006); Wooldridge (2007); Uysal (2011, 2015); Sant'Anna and Zhao (2020);
- For a recent review see Słoczyński and Wooldridge (2018)

## Instrumental Variable estimation of the LATE

- Imbens and Angrist (1994); Angrist et al. (1996) clarify the relationbetween IV and LATE -when the instrument is randomly assigned
- Extensions: Abadie (2003); Frölich (2007); Tan (2006); Uysal (2011); Donald et al. (2014); Ogburn et al. (2015); Belloni et al.(2017); Chernozhukov et al. (2018); Sun and Tan (2022); Heiler (2022); Singh and Sun (2022); Słoczyński et al. (2023)

# Outline

# Recap of LATE

## Recap of LATE

**Potential outcomes:** $Y(1)$ is outcome if $W = 1$, $Y(0)$ is outcome if $W = 0$; what follows, $Y = (1 - W)Y(0) + WY(1)$.

**Treatment effect:** $Y(1i) - Y(0i)$ for individual $i$.

**There are two potential treatments, too:** $W(1)$ is treatment if $Z = 1$, $W(0)$ is treatment if $Z = 0$; what follows, $W = (1 - Z)W(0) + ZW(1)$.

**Standard terminology:** if $W(1) = 1$ and $W(0) = 1$, *always takers*; if $W(1) = 1$ and $W(0) = 0$, *compliers*; if $W(1) = 0$ and $W(0) = 1$, *defiers*; finally, if $W(1) = 0$ and $W(0) = 0$, *never takers*.

# Recap of LATE

**Potential outcomes:** $Y(1)$ is outcome if $W = 1$, $Y(0)$ is outcome if $W = 0$; what follows, $Y = (1 - W)Y(0) + WY(1)$.

**Treatment effect:** $Y(1i) - Y(0i)$ for individual $i$.

**There are two potential treatments, too:** $W(1)$ is treatment if $Z = 1$, $W(0)$ is treatment if $Z = 0$; what follows, $W = (1 - Z)W(0) + ZW(1)$.

**Standard terminology:** if $W(1) = 1$ and $W(0) = 1$, *always takers*; if $W(1) = 1$ and $W(0) = 0$, *compliers*; if $W(1) = 0$ and $W(0) = 1$, *defiers*; finally, if $W(1) = 0$ and $W(0) = 0$, *never takers*.

**LATE**

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

**LATT**

$$\tau_{LATT} = \mathbb{E}[Y(1) - Y(0) \mid W(1) > W(0), W = 1]. \tag{3}$$

# Identifying Assumptions (Abadie, 2003; Frölich, 2007; Tan, 2006)

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

$$\mathbb{P}[Y(w,1) = Y(w,0) \,|X = x] = 1. \ \square$$

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

$$[Y(0), Y(1), W(0), W(1)] \perp Z|X. \ \square$$

**Assumption 3 (Monotonicity):**

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

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

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

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

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

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

$$\mathbb{P}[Y(w,1) = Y(w,0) | X = x] = 1. \square$$

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

$$[Y(0), Y(1), W(0), W(1)] \perp Z|X. \square$$

**Assumption 3 (Monotonicity):**

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

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

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

**Assumption 5' (Overlap for LATT):**
For almost all $x \in \mathcal{X}$

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

# Identification & Estimation under Assumptions 1-5

**Theorem I Frölich (2007), Tan (2006) (Identification of LATE):**

Under Assumptions 1-5,

$$\tau_{LATE} = \frac{\mathbb{E}\left[\mathbb{E}\left(Y|X,Z=1\right) - \mathbb{E}\left(Y|X,Z=0\right)\right]}{\mathbb{E}\left[\mathbb{E}\left(W|X,Z=1\right) - \mathbb{E}\left(W|X,Z=0\right)\right]} = \frac{\mathbb{E}[\mu_1(X)] - \mathbb{E}[\mu_0(X)]}{\mathbb{E}[\rho_1(X)] - \mathbb{E}[\rho_0(X)]}, \tag{4}$$

where

$$\mathbb{E}\left(Y|X,Z=z\right) \equiv \mu_z(X) \tag{5}$$

and

$$\mathbb{E}\left(W|X,Z=z\right) = \mathbb{E}[W\left(z\right)|X] = \rho_z(X), \text{ for } z \in \{0,1\}. \ \square \tag{6}$$

**Theorem II (Proof in the paper) (Identification of LATT):**

Under Assumptions 1-5',

$$\tau_{LATT} = \frac{\mathbb{E}\left(Y|Z=1\right) - \mathbb{E}[\mu_0(X)|Z=1]}{\mathbb{E}\left(W|Z=1\right) - \mathbb{E}[\rho_0(X)|Z=1]}. \ \square \tag{7}$$

# Identification & Estimation under Assumptions 1-5

**Theorem I Frölich (2007), Tan (2006) (Identification of LATE):**

Under Assumptions 1-5,

$$\tau_{LATE} = \frac{\mathbb{E}\left[\mathbb{E}\left(Y|X, Z=1\right) - \mathbb{E}\left(Y|X, Z=0\right)\right]}{\mathbb{E}\left[\mathbb{E}\left(W|X, Z=1\right) - \mathbb{E}\left(W|X, Z=0\right)\right]} = \frac{\mathbb{E}[\mu_1(X)] - \mathbb{E}[\mu_0(X)]}{\mathbb{E}[\rho_1(X)] - \mathbb{E}[\rho_0(X)]}, \quad (4)$$

where

$$\mathbb{E}\left(Y|X, Z=z\right) \equiv \mu_z(X) \quad (5)$$

and

$$\mathbb{E}\left(W|X, Z=z\right) = \mathbb{E}[W(z)|X] = \rho_z(X), \text{ for } z \in \{0,1\}. \ \square \quad (6)$$

**Theorem II (Proof in the paper) (Identification of LATT):**

Under Assumptions 1-5',

$$\tau_{LATT} = \frac{\mathbb{E}\left(Y|Z=1\right) - \mathbb{E}[\mu_0(X)|Z=1]}{\mathbb{E}\left(W|Z=1\right) - \mathbb{E}[\rho_0(X)|Z=1]}. \ \square \quad (7)$$

# A Doubly Robust Approach to Estimating LATE & LATT

# General Idea

- We propose an IPWRA (inverse probability weighted regression adjustment) type doubly robust estimator for the LATE and LATT parameter.
- It combines inverse probability weighting (IPW) and regression adjustment (RA) using a particular quasi-maximum likelihood estimator.
- The resulting estimator for the LATE/LATT is consistent
    - if model specification related to IPW part is correct, **or**
    - if the model specifications related to RA part are correct,

  i.e. it is doubly robust.

# Estimation Procedure DR LATE

Specify Parametric Models:

- Instrument propensity score:
  $\mathbb{P}(Z_i = 1|X_i) = G(X_i, \gamma)$

- Conditional means:
  $\mathbb{P}(W_i = 1|X_i, Z_i = z) \equiv \rho_z(X_i) \Rightarrow \Lambda(\omega_z + X_i\delta_z)$
  $\mathbb{E}(Y_i|X_i, Z_i = z) \equiv \mu_z(X_i) \Rightarrow m(\alpha_z + X_i\beta_z)$, for $z = 0, 1$

! Choose $m(\cdot)$ considering the nature of $Y$ from LEF with canonical link function.

Estimation:

1. Estimate $\gamma \Rightarrow G(X_i, \hat{\gamma})$.
2. Estimate $(\omega_z, \delta_z)$ and $(\alpha_z, \beta_z)$ by a weighted QMLE with weights $1/[1 - G(X_i, \hat{\gamma})]$ for $z = 0$ and and $1/G(X_i, \hat{\gamma})$ for $z = 1$.

## DR Estimators of $\tau_{LATE}$

- Hence, we obtain the DR estimator of $\tau_{LATE}$ as

**$\hat{\tau}_{DRLATE}$**

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

- Under standard regularity conditions, $\hat{\tau}_{DRLATE}$ is consistent for $\tau_{LATE}$
  - if the model for $\mathbb{P}(Z = 1|X)$ is correct, **or**
  - if the models for $\mu_z(X)$ and $\rho_z(X)$ are correct.

# Estimation Procedure DR LATT

Specify Parametric Models:

- Instrument propensity score:
  $\mathbb{P}(Z_i = 1|X_i) = G(X_i, \gamma)$
- Conditional means (only for $z = 0$):
  $\mathbb{P}(W_i = 1|X_i, Z_i = 0) \equiv \rho_0(X_i) \Rightarrow \Lambda(\omega_0 + X_i\delta_0)$
  $\mathbb{E}(Y_i|X_i, Z_i = 0) \equiv \mu_0(X_i) \Rightarrow m(\alpha_0 + X_i\beta_0)$

Estimation:

1. Estimate $\gamma \Rightarrow G(X_i, \hat{\gamma})$.
2. Estimate $(\omega_0, \delta_0)$ and $(\alpha_0, \beta_0)$ by a weighted QMLE with weights $G(X_i, \hat{\gamma}) / [1 - G(X_i, \hat{\gamma})]$
3. Obtain the DR estimator of $\tau_{LATT}$ as

$$\hat{\tau}_{DRLATT} = \frac{\bar{Y}_1 - N_1^{-1}\sum_{i=1}^N Z_i m(\hat{\alpha}_0 + X_i\hat{\beta}_0)}{\bar{W}_1 - N_1^{-1}\sum_{i=1}^N Z_i \Lambda(\hat{\omega}_0 + X_i\hat{\delta}_0)}, \tag{8}$$

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

# Doubly Robustness

- Doubly robustness of the denominator follows the doubly robustness of the ATE by weighted regression adjustment (see Wooldridge; 2007)
- For the numerator, it is a different story 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.
- The trick is to write $Y$ with a zero conditional mean error term:

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

# A Test Comparing LATT and ATT Estimators

## A Test Comparing LATT and ATT Estimators

- Under one-sided noncompliance, LATT is the same as ATT (Donald et al., 2014)
- We can test the null hypothesis that treatment is unconfounded given $X$!
- We propose 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

$$\tau_{ATT} = \mathbb{E}\left[Y(1) - Y(0) \mid W = 1\right]. \tag{9}$$

- Let $\hat{\tau}_{DRATT}$ be the DR estimator of the ATT.

# A Test Comparing LATT and ATT Estimators

- A formal comparison is based on the statistic

$$\frac{\hat{\tau}_{DRLATT} - \hat{\tau}_{DRATT}}{\mathrm{se}\left(\hat{\tau}_{DRLATT} - \hat{\tau}_{DRATT}\right)}. \tag{10}$$

- The standard error $\mathrm{se}\left(\hat{\tau}_{DRLATT} - \hat{\tau}_{DRATT}\right)$ does not simplify, but, bootstrapping is computationally feasible, or one can extend the calculations for the asmyptotics 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.

- We are preparing a Stata package which implements all these tests with analytical standard errors.

# Simulations

# DGP

- We simulated data to mimic certain statistical features of the data used by Abadie (2003)
- 1,000 samples with $N = 1{,}000$ and the same number of samples with $N = 4{,}000$ observations.
- Draw two random variables from a bivariate normal distribution matching the empirical means and covariances of age and log income in the 401(k) data.
- Our full set of covariates, $X$, includes three variables: income, age, and age squared.

$$Z = \mathbb{1}\left(\Lambda\left(\gamma_0 + X\gamma_x\right) > U_z\right), \tag{11}$$

$$W(1) = \mathbb{1}\left(\Lambda\left(\omega_0 + X\delta_0\right) > U_1\right), \tag{12}$$

$$Y(z) = \alpha_z + X\beta_z + \varepsilon_z \quad \text{for } z = 0, 1. \tag{13}$$

# Estimation

- Estimate Instrument Propensity Score by a logit
- Estimate the parametric models for $\mathbb{P}(W_i = 1|X_i, Z_i = z)$
  - unweighted and weighted QMLE with a logit link
- Estimate the parametric models for $\mathbb{E}(Y_i|X_i, Z_i = z)$
  - unweighted and weighted QMLE with the identity link

Estimation of the LATE by:

- IPW (uses only fitted IPS)
- RA (uses fitted $\mathbb{E}(W_i|X_i, Z_i = z)$ and $\mathbb{E}(Y_i|X_i, Z_i = z)$ using unweighted regression )
- IPWRA (uses fitted $\mathbb{E}(W_i|X_i, Z_i = z)$ and $\mathbb{E}(Y_i|X_i, Z_i = z)$ using weighted regression)
- AIPW (uses fitted $\mathbb{E}(W_i|X_i, Z_i = z)$ and $\mathbb{E}(Y_i|X_i, Z_i = z)$ using unweighted regression and fitted IPS for the IPW part)

# Estimation

In our Monte Carlo study, we consider estimators

*(i)* when the required models are all correctly specified,

*(ii)* when models for $\mathbb{P}(W_i = 1|X_i, Z_i = z)$ and $\mathbb{E}(Y_i|X_i, Z_i = z)$ for $z = 0, 1$ are misspecified, and

*(iii)* when the model for the IPS $(\mathbb{P}(Z_i = 1|X_i))$ is misspecified.

Correct specifications for these estimators mean that we use the correct set of covariates for all the regressions, and misspecification of a certain model means that one of the regressors is omitted.

# Summary of Simulation Results

- Our proposed method is never substantially more biased than the competing estimators
- Its RMSE is better than that of AIPW, which is the only alternative that shares the double robustness property of IPWRA.

# Simulation Results

**Table 1:** Simulation Results for the Continuous Outcome Variable

| | All Correct | | | $\rho_z(X_i)$ and $\mu_z(X_i)$ misspecified | | | $G(x,\gamma)$ 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:* 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 framewor. (remaining estimators).

# Outlook

## Outlook

- We propose a new class of estimators for the LATE parameter:
    - two ways protection against certain types of misspecification, i.e. doubly robust
    - fitted values are ensured to be in the logical range determined by the response variable
    - desirable finite sample properties
- We also propose a DR version of the Hausman test that can be used to assess the unconfoundedness assumption and treatment heterogeneity