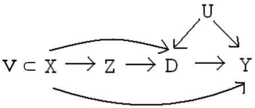
Doubly robust estimation of the local average treatment effect curve

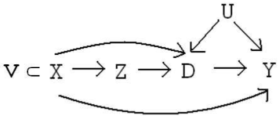
Elizabeth L. Ogburn, Andrea Rotnitzky, and James M. Robins

background
hypothetical example



- We want to know the causal effect of colonoscopy (D) on colorectal cancer (Y).
- High rates of noncompliance for colonoscopy.
- Unmeasured confounders (U) of receiving colonoscopy and outcome include mental health, underlying attitudes and behaviors related to health.
- Treatment assignment (Z) is an instrument for treatment.
- Covariates (X) include age, family history, BMI, smoking history, fecal occult bleeding test (ng/mL).
- Counterfactuals $Y_{z,d}$, D_z

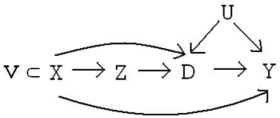
background
setting and definitions



Compliance types:

- **always takers** take treatment regardless of the instrument: $D_1 = D_0 = 1$
- **never takers** do not take treatment regardless of the instrument: $D_1 = D_0 = 0$
- **compliers** follow their assignment: $D_Z = Z$
- **defiers** do the opposite of their assignment: $D_Z = 1 - Z$

background
setting and definitions



- The **local average treatment effect (LATE)** is the treatment effect among compliers

$$LATE(v) = E[Y_1 - Y_0 | Complier, V = v]$$

- Our goal is to model $LATE(v)$ as robustly as possible.
- V may be a strict subset of X, it may be equal to X, or it may be the empty set.

outline

1. Background.
2. Warm up with the DR model for $LATE(x)$.
3. Present the more general model for $LATE(v)$.
4. Data analysis.
5. A surprising link between our model for $LATE(v)$ and a different class of models.

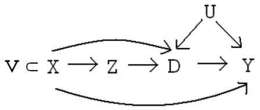
context

- Estimation of LATE first proposed by Angrist, Imbens, and Rubin (1993, 1996) and Baker and Lindeman (1994).
- Froelich (2007), Tan (2006), and Uysal (2011) proposed DR estimators for $LATE$ marginalized over covariates (see also Okui et al, 2012).
- There have been many proposals for estimation of $LATE(x)$ (Abadie, 2003; Hirano et al, 2000; Little and Yau, 1998; Tan, 2006).
    - They are not DR.
    - They require more modeling restrictions than ours when X is high dimensional and result in parametric specifications for $LATE(x)$ that are difficult to interpret.
- We are not aware of any previous proposals for estimating $LATE(v)$, though it would be possible using the methods in Tan (2010). However, the methods in Tan (2010) could suffer from model incompatibility.

context

- We propose doubly robust estimators for $LATE(v)$ that parameterize $LATE(v)$ directly, ensuring interpretability of the model of interest, along with two nuisance models.
- But first, $LATE(x)$...

background
I.V. assumptions for identifiability



(i) **exclusion**: there is no direct effect of Z on Y, $Y_{z,d} = Y_d$

(ii) **instrumentation**: Z has a causal effect on D for all X, i.e.
$P(D_1 = 1|X) - P(D_0 = 1|X) \neq 0$ w.p. 1

(iii) **randomization**: Z is independent of the counterfactuals for D and Y conditional on X, i.e.$\{Y_d, D_z\} \perp Z \mid X$

(iv) **monotonicity**: there are no defiers in the population, i.e. $D_1 \geq D_0$

(v) **positivity**: the support of X is the same among those with $Z = 1$ and $Z = 0$, i.e. $0 < P(Z = 1|X) < 1$

(vi) **consistency**: The observed outcome (treatment) is the counterfactual corresponding to the observed treatment (instrument), i.e.
$Y = DY_1 + (1 - D)Y_0$ and $D = ZD_1 + (1 - Z)D_0$

background
modeling assumptions

Under these assumptions, $LATE(x)$ is identified by the I.V. estimand

$$IVE(x) \equiv \frac{E\left[Y|Z=1, X=x\right] - E\left[Y|Z=0, X=x\right]}{E\left[D|Z=1, X=x\right] - E\left[D|Z=0, X=x\right]}.$$

(Angrist, Imbens, and Rubin, 1993, 1996; Imbens and Angrist, 1994)

background
modeling assumptions

- We would like to be able to estimate $LATE(x)$ under the semiparametric model that posits only the I.V. assumptions and a parametric model for $LATE(x)$.
- But the curse of dimensionality is such that, for X high dimensional, additional modeling assumptions are required.

observed data restrictions

$$\text{I.V. assumptions} \Longrightarrow \begin{cases} P(y < Y \leq y', D = 1 | Z = 1, X) - P(y < Y \leq y', D = 1 | Z = 0, X) \geq 0 \\ P(y < Y \leq y', D = 0 | Z = 0, X) - P(y < Y \leq y', D = 0 | Z = 1, X) \geq 0 \\ E(D | Z = 1, X) - E(D | Z = 0, X) > 0 \\ 0 < P(Z = 1 | X) < 1 \end{cases}$$

$$m(x; \beta^*) = LATE(x) \implies m(x; \beta^*) = \frac{E[Y | Z = 1, X = x] - E[Y | Z = 0, X = x]}{E[D | Z = 1, X = x] - E[D | Z = 0, X = x]}$$

observed data restrictions

$$\text{I.V. assumptions} \Longrightarrow \begin{cases} P\left(y < Y \leq y', D = 1 | Z = 1, X\right) - P\left(y < Y \leq y', D = 1 | Z = 0, X\right) \geq 0 \\ P\left(y < Y \leq y', D = 0 | Z = 0, X\right) - P\left(y < Y \leq y', D = 0 | Z = 1, X\right) \geq 0 \\ E\left(D | Z = 1, X\right) - E\left(D | Z = 0, X\right) > 0 \\ 0 < P(Z = 1 | X) < 1 \end{cases}$$

$$m\left(x; \beta^*\right) = LATE(x) \implies m\left(x; \beta^*\right) = \frac{E\left[Y | Z = 1, X = x\right] - E\left[Y | Z = 0, X = x\right]}{E\left[D | Z = 1, X = x\right] - E\left[D | Z = 0, X = x\right]}$$

the model

When Z is binary, as we assume throughout,

$$m(x;\beta^*) = \frac{E[Y|Z=1,X=x] - E[Y|Z=0,X=x]}{E[D|Z=1,X=x] - E[D|Z=0,X=x]}$$

is equivalent to

$$Cov\left(\underbrace{Y - m(X;\beta^*)D}_{H(\beta^*)}, Z \middle| X\right) = 0.$$

the model

- Inference is based on the conditional moment restriction

$$Cov\left(\underbrace{Y - m(X;\beta^*)D}_{H(\beta^*)}, Z \middle| X\right) = 0.$$

- The set gradients for β^* is

$$\left\{ q(X)\left(\underbrace{Y - m(X;\beta^*)D}_{H(\beta^*)} - \underbrace{E\left[Y - m(X;\beta^*)D|X\right]}_{E[H(\beta^*)|X]}\right)(Z - E[Z|X])\right\}.$$

the model

$$\left\{ q(X) \left(\underbrace{Y - m(X;\beta^*)D}_{H(\beta^*)} - \underbrace{E\left[Y - m(X;\beta^*)D \middle| X \right]}_{E[H(\beta^*)|X]} \right) (Z - E\left[Z|X\right]) \right\}$$

- For high dimensional X, we cannot hope to find an estimator with influence function in this set.

the model

$$\left\{ q(X) \left(\underbrace{Y - m(X;\beta^*)D}_{H(\beta^*)} - \underbrace{E\left[Y - m(X;\beta^*)D | X \right]}_{E[H(\beta^*)|X]} \right) (Z - E[Z|X]) \right\}$$

- For high dimensional X, we cannot hope to find an estimator with influence function in this set.
- Postulate two additional models:
  - (1) $E[Z|X] = \pi(X;\alpha)$
    - $\widehat{\alpha}$ solves the score equations $E_n\left[\frac{\partial}{\partial\alpha}\text{logit}\pi(X;\alpha)\{Z - \pi(X;\alpha)\}\right] = 0$
  - (2) $E[H(\beta)|X] = h(X;\eta(\beta))$
    - For each β, $\widehat{\eta}(\beta)$ solves the estimating equation $E_n\left[\frac{\partial}{\partial\eta}h(X;\eta)\{H(\beta) - h(X;\eta)\}\right] = 0$

the model

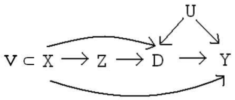
- DR estimating equations:

$$E_n\left[q(X)\left(H(\beta)-\underbrace{h(X;\widehat{\eta}(\beta))}_{\substack{\text{model for}\\ E[H(\beta)|X]}}\right)\left(Z-\underbrace{\pi(X;\widehat{\alpha})}_{\substack{\text{model for}\\ E[Z|X]}}\right)\right]=0$$

- CAN for β^* if either $h(X;\eta)$ or $\pi(X;\alpha)$ is correctly specified.
- If both are correctly specified, and if $q(X)=q_{opt}(X)$, then our estimator attains the asymptotic semiparametric efficiency bound.

$$LATE(v)$$

identifying assumptions



(i) **exclusion**: there is no direct effect of Z on Y, $Y_{z,d} = Y_d$

(ii) **instrumentation**: Z has a causal effect on D for all V, i.e. $P[D_1 = 1|V] - P[D_0 = 1|V] \neq 0$ w.p 1

(iii) **randomization**: Z is independent of the counterfactuals for D and Y conditional on X, i.e.$\{Y_d, D_z\} \perp Z \mid X$

(iv) **monotonicity**: there are no defiers in the population, i.e. $D_1 \geq D_0$

(v) **positivity**: the support of X is the same among those with $Z = 1$ and $Z = 0$, i.e. $0 < P(Z = 1|X) < 1$

(vi) **consistency**: The observed outcome (treatment) is the counterfactual corresponding to the observed treatment (instrument), i.e. $Y = DY_1 + (1 - D)Y_0$ and $D = ZD_1 + (1 - Z)D_0$

estimation of $LATE(v)$

Under the I.V. assumptions, $LATE(v)$ is identified by the conditional I.V. estimand

$$IVE(v) = \frac{E\left[E\left(Y|Z=1,X\right)-E\left(Y|Z=0,X\right)|V=v\right]}{E\left[E\left(D|Z=1,X\right)-E\left(D|Z=0,X\right)|V=v\right]}.$$

observed data restrictions

$$\text{I.V. assumptions} \Longrightarrow \begin{cases} P(y < Y \leq y', D = 1|Z = 1, X) - P(y < Y \leq y', D = 1|Z = 0, X) \geq 0 \\ P(y < Y \leq y', D = 0|Z = 0, X) - P(y < Y \leq y', D = 0|Z = 1, X) \geq 0 \\ E(D|Z = 1, X) - E(D|Z = 0, X) > 0 \\ 0 < P(Z = 1|X) < 1 \end{cases}$$

$$m(v;\beta^*) = LATE(v) \Longrightarrow m(v;\beta^*) = \frac{E[E(Y|Z=1,X) - E(Y|Z=0,X)|V=v]}{E[E(D|Z=1,X) - E(D|Z=0,X)|V=v]}$$

the model

$$m(v;\beta^*) = \frac{E\left[E\left(Y|Z=1,X\right) - E\left(Y|Z=0,X\right)|V=v\right]}{E\left[E\left(D|Z=1,X\right) - E\left(D|Z=0,X\right)|V=v\right]}$$

is equivalent to

$$E\left\{E\left[\underbrace{Y - m(V;\beta^*)D}_{H(\beta^*)} \mid Z=1,X\right] \mid V\right\} - E\left\{E\left[\underbrace{Y - m(V;\beta^*)D}_{H(\beta^*)} \mid Z=0,X\right] \mid V\right\} = 0$$

and to

$$E\left\{\left(\frac{1}{P[Z=1|X]}\right)^{Z}\left(-\frac{1}{P[Z=0|X]}\right)^{1-Z} H(\beta^*) \mid V\right\} = 0.$$

estimation of $LATE(v)$

- The set of gradients for β^* is

$$\left\{ q(V) \left[\left(\frac{1}{P[Z=1|X]} \right)^{Z} \left(-\frac{1}{P[Z=0|X]} \right)^{1-Z} H(\beta^*) - (Z - P[Z=1|X]) \left(\frac{E[H(\beta^*)|Z=1,X]}{P[Z=1|X]} + \frac{E[H(\beta^*)|Z=0,X]}{P[Z=0|X]} \right) \right] \right\}$$

- For high dimensional X, we cannot hope to find an estimator with influence function in this set.

estimation of $LATE(v)$

- The set of gradients for β^* is

$$\left\{ q(V)\left[\left(\frac{1}{P[Z=1|X]}\right)^{Z}\left(-\frac{1}{P[Z=0|X]}\right)^{1-Z} H(\beta^*) \right.\right.$$
$$\left.\left. -(Z-P[Z=1|X])\left(\frac{E[H(\beta^*)|Z=1,X]}{P[Z=1|X]}+\frac{E[H(\beta^*)|Z=0,X]}{P[Z=0|X]}\right)\right]\right\}$$

- For high dimensional X, we cannot hope to find an estimator with influence function in this set.

estimation of $LATE(v)$

$$\Bigg\{ q(V) \Bigg[\left(\frac{Z}{P[Z=1|X]} \right)^{Z} \left(-\frac{1-Z}{P[Z=0|X]} \right)^{1-Z} H(\beta^*) - (Z - P[Z=1|X]) \left(\frac{E[H(\beta^*)|Z=1,X]}{P[Z=1|X]} + \frac{E[H(\beta^*)|Z=0,X]}{P[Z=0|X]} \right) \Bigg] \Bigg\}$$

We postulate two additional models

$$(1) \quad E[Z|X] = \pi(X;\alpha)$$

$$(2) \quad E[H(\beta)|Z,X] = h(Z,X;\eta(\beta)).$$

estimation of $LATE(v)$

- DR estimating equations

$$E_n\left\{q(V)\left[\left(\frac{Z}{\pi(X;\hat{\alpha})}\right)^{Z}\left(-\frac{1-Z}{1-\pi(X;\hat{\alpha})}\right)^{1-Z}H(\beta)\right.\right.$$
$$\left.\left.-(Z-\pi(X;\hat{\alpha}))\left(\frac{h(1,X;\hat{\eta}(\beta))}{\pi(X;\hat{\alpha})}+\frac{h(0,X;\hat{\eta}(\beta))}{1-\pi(X;\hat{\alpha})}\right)\right]\right\}=0$$

- The solution $\hat{\beta}$ is CAN for β^* if either $\pi(X;\alpha)$ or $h(Z,X;\eta(\beta))$ is correctly specified.
- If both are correctly specified and if $q(V) = q_{opt}(V)$, our estimator attains the asymptotic semiparametric efficiency bound (see page 383).

estimation of $LATE(v)$

- The model for $E[H(\beta)|Z,X]$ has to respect the constraint that $E\{E[H(\beta^*) \mid Z=1,X] \mid -E[H(\beta^*) \mid Z=0,X] \mid V\}=0.$
  - This is immediate when $V=X$ but not straightforward when V is a strict, non-empty subset of X.
  - We give one possible modeling strategy on p. 380.
- Models $\pi(X;\alpha)$ and $h(Z,X;\eta)$ are variation independent of $m(V;\beta)$. That is, no modeling assumptions incorporated into $\pi(X;\alpha)$ and $h(Z,X;\eta)$ can conflict with any parametric specification $m(V;\beta)$ of $LATE(v)$.
- The asymptotic variance of $\hat{\beta}$ can be estimated by the sandwich variance estimator or by the bootstrap.

data analysis

- What is the effect of 401(k) tax-deferred retirement plans on household saving in the U.S.? Do 401(k) plans represent increased saving or do they replace other modes of saving?
- Survey of Income and Program Participation (SIPP) data, previously analyzed by Abadie (2003), included 9725 household reference subjects.
- $Y =$ net financial assets; $Z =$ 401(k) eligibility; $D =$ 401(k) participation;
  $X =$ (age, married, family size, household income)
- Eligibility is determined by employers; $D = 0$ whenever $Z = 0$.
  - Therefore there are no defiers or always takers.
- We estimated $LATE(income)$.

data analysis

Estimators of (β_0, β_1) and their bootstrap standard errors under model $LATE(income) = \beta_0 + \beta_1 \cdot income$.					
		Power k of income			
		1	2	4	8
intercept	$\widehat{\beta}_{dr}^{opt}$	-4640 (2940)	-1845 (3220)	-1490 (2900)	-1566 (2896)
	$\widehat{\beta}_{dr}^{ineff,stable}$	-418 (4827)	-4958 (5547)	-1814 (4527)	-1590 (4543)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	12331 (6076)	-3489 (5632)	-1478 (4019)	-1179 (4409)
	$\widehat{\beta}_{reg}$	-6992 (7019)	1929 (7665)	-1266 (6796)	-1494 (7004)
income	$\widehat{\beta}_{dr}^{opt}$	382 (88)	337 (92)	328 (82)	328 (83)
	$\widehat{\beta}_{dr}^{ineff,stable}$	272 (128)	425 (149)	340 (123)	331 (120)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	14 (161)	385 (154)	339 (117)	329 (123)
	$\widehat{\beta}_{reg}$	510 (187)	272 (210)	345 (183)	353 (194)

data analysis

Estimators of (β_0, β_1) and their bootstrap standard errors under model $LATE(income) = \beta_0 + \beta_1 income$.

		Power k of income			
		1	2	4	8
intercept	$\widehat{\beta}_{dr}^{opt}$	-4640 (2940)	-1845 (3220)	-1490 (2900)	-1566 (2896)
	$\widehat{\beta}_{dr}^{ineff,stable}$	-418 (4827)	-4958 (5547)	-1814 (4527)	-1590 (4543)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	12331 (6076)	-3489 (5632)	-1478 (4019)	-1179 (4409)
	$\widehat{\beta}_{reg}$	-6992 (7019)	1929 (7665)	-1266 (6796)	-1494 (7004)
income	$\widehat{\beta}_{dr}^{opt}$	382 (88)	337 (92)	328 (82)	328 (83)
	$\widehat{\beta}_{dr}^{ineff,stable}$	272 (128)	425 (149)	340 (123)	331 (120)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	14 (161)	385 (154)	339 (117)	329 (123)
	$\widehat{\beta}_{reg}$	510 (187)	272 (210)	345 (183)	353 (194)

Different DR estimators give similar estimates; this is consistent with approximately correct specification of the LATE(income) model.

data analysis

Estimators of (β_0, β_1) and their bootstrap standard errors under model $LATE(income) = \beta_0 + \beta_1 income$.

		Power k of income			
		1	2	4	8
intercept	$\widehat{\beta}_{dr}^{opt}$	-4640 (2940)	-1845 (3220)	-1490 (2900)	-1566 (2896)
	$\widehat{\beta}_{dr}^{ineff,stable}$	-418 (4827)	-4958 (5547)	-1814 (4527)	-1590 (4543)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	12331 (6076)	-3489 (5632)	-1478 (4019)	-1179 (4409)
	$\widehat{\beta}_{reg}$	-6992 (7019)	1929 (7665)	-1266 (6796)	-1494 (7004)
income	$\widehat{\beta}_{dr}^{opt}$	382 (88)	337 (92)	328 (82)	328 (83)
	$\widehat{\beta}_{dr}^{ineff,stable}$	272 (128)	425 (149)	340 (123)	331 (120)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	14 (161)	385 (154)	339 (117)	329 (123)
	$\widehat{\beta}_{reg}$	510 (187)	272 (210)	345 (183)	353 (194)

The income coefficient is approximately 330, suggesting that 401(k) plans have more effect on the savings of families with higher incomes.

data analysis

Estimators of (β_0, β_1) and their bootstrap standard errors under model $LATE(income) = \beta_0 + \beta_1 income$.					
		Power k of income			
		1	2	4	8
intercept	$\widehat{\beta}_{dr}^{opt}$	-4640 (2940)	-1845 (3220)	-1490 (2900)	-1566 (2896)
	$\widehat{\beta}_{dr}^{ineff,stable}$	-418 (4827)	-4958 (5547)	-1814 (4527)	-1590 (4543)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	12331 (6076)	-3489 (5632)	-1478 (4019)	-1179 (4409)
	$\widehat{\beta}_{reg}$	-6992 (7019)	1929 (7665)	-1266 (6796)	-1494 (7004)
income	$\widehat{\beta}_{dr}^{opt}$	382 (88)	337 (92)	328 (82)	328 (83)
	$\widehat{\beta}_{dr}^{ineff,stable}$	272 (128)	425 (149)	340 (123)	331 (120)
	$\widehat{\beta}_{ipw}^{ineff,stable}$	14 (161)	385 (154)	339 (117)	329 (123)
	$\widehat{\beta}_{reg}$	510 (187)	272 (210)	345 (183)	353 (194)

As expected, the DR estimator with the optimal $q(\cdot)$ function has the smallest standard errors.

connection to ETT

- Robins (1994) and Tan (2010) estimated the average treatment effect on the treated

$$ETT(V) = E\left[Y_1 - Y_0 | D = 1, V\right].$$

- Though the Robins-Tan model is quite different from ours, estimating procedures are the same under the two models!

connection to ETT

- The Robins-Tan model for the ETT assumes
  - exclusion, randomization, instrumentation, positivity, and consistency – but **not monotonicity**.
  - no treatment-instrument interaction: $ETT(z, v) = ETT(v)$.
  - a parametric model $m(v; \beta^*) = ETT(v)$.
- Under these assumptions, $ETT(v)$ is identified by the I.V. estimand

$$IVE(v) = \frac{E\left[E\left(Y|Z=1,X\right) - E\left(Y|Z=0,X\right)|V=v\right]}{E\left[E\left(D|Z=1,X\right) - E\left(D|Z=0,X\right)|V=v\right]}$$

connection to ETT

The Robins-Tan model imposes these constraints on the observed data:

$$\text{Assumptions} \Longrightarrow \begin{cases} E\left[P\left(D=1|Z=1,X\right)|V\right] \neq E\left[P\left(D=1|Z=0,X\right)|V\right] \text{ w.p. } 1 \\ 0 < P(Z=1|X) < 1 \end{cases}$$

$$m(v;\beta^*) = ETT(v) \implies m(v;\beta^*) = \frac{E\left[E\left(Y|Z=1,X\right) - E\left(Y|Z=0,X\right)|V=v\right]}{E\left[E\left(D|Z=1,X\right) - E\left(D|Z=0,X\right)|V=v\right]}$$

connection to ETT

- Every observed data distribution is compatible with a counterfactual world in which the models are the same and the ETT is equal to the $LATE$.
- This counterfactual world is characterized by
  - no defiers
  - the effect of treatment is the same among compliers and always takers.
- The treated population is comprised of compliers and always takers, so if we assume effect homogeneity then $ETT(v) = LATE(v)$.
- This condition is unlikely to hold in reality, but it is untestable. Because it is compatible with any observed data distribution inference must be the same whether it holds or not.

summary

DR estimation of $LATE(v)$ where $V \subseteq X$

- requires nuisance models for $E[H(\beta)|Z, X]$ and $P(Z = 1|X)$.
- can be important effect for clinical decisions.
- is the same as DR estimation of $ETT(v)$ under a different set of identifying assumptions.

Please see the paper for

- extra efficiency protection.
- DR estimation of $MLATE(v) = \frac{E[Y_1|Complier, V=v]}{E[Y_0|Complier, V=v]}$.
- DR estimation of the least squares approximation to $LATE(v)$.
- further exploration of the SIPP data.

references

Baker SG and Lindeman KS (1994) The Paired Availability Design: A Proposal for Evaluating Epidural Analgesia During Labor. *Statistics in Medicine*, **13**, 2269-2278.

Okui R, Small DS, Tan Z, and Robins JM (2012) Doubly Robust Instrumental Variable Regression. *Statistica Sinica*, **22**, 173-205.

All other references can be found in the paper.

Thank you!