Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect

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

2023 Stata Virtual Symposium
November 9, 2023

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

Motivation

- What happens when we use the IV/2SLS estimator to estimate:

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

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

Motivation

- What happens when we use the IV/2SLS estimator to estimate:

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

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

- 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 τ as **the local average treatment effect (LATE)**

Motivation

- What happens when we use the IV/2SLS estimator to estimate:

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

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

- 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 τ as **the local average treatment effect (LATE)**

- In fact: when X is nonempty, τ is a weighted average of X-specific LATEs (Angrist and Imbens 1995) with rather undesirable weights (Słoczyński 2022).

- This paper: weighting type estimators of the LATE with covariates (mostly) based on Abadie (2003)

Introduction

- Abadie (2003) demonstrates how to identify any parameter that is defined in terms of moments of the joint distribution of the data for compliers by "kappa weighting".

- Abadie (2003)'s result has been used in applied work to:
  - estimate mean covariate values for compliers (e.g., Angrist et al., 2013; Dahl et al., 2014; Bisbee et al., 2017)
  - approximate the conditional mean of an outcome of interest in this subpopulation (e.g., Cruces and Galiani, 2007; Angrist et al., 2013; Goda et al., 2017)

- Yet, kappa weighting has been **rarely discussed for the estimation of the local average treatment effects**

- Alternative way to construct weighting estimators of the LATE: the ratio of the IPW (Inverse Probability Weighting) estimator of the ATE of the instrument on the outcome and the IPW estimator of the ATE of the instrument on the treatment (Frölich, 2007)

This paper

A comprehensive treatment of different approaches to constructing weighting estimators of the LATE with covariates (mostly based on Abadie; 2003):

- We stress the importance of normalization for the weighting type estimators and provide an objective and intuitively appealing criterion.
- We show that in case of one-sided noncompliance certain unnormalized estimators have also an appealing property
- We show with a certain choice of IPS (Instrument Propenstiy Score) estimation we can have both properties together.

This paper

We study estimation of each of these estimators in a unified framework of M-estimation.

We illustrate our findings with a simulation study and three empirical applications.

Empirical applications:

- In a replication of Angrist (1990), unnormalized estimators are highly variable across specifications.
- In a replication of Card (1995), the unnormalized estimates are either "too large" in magnitude or otherwise negative, which makes little sense for causal effects of college education.
- In a replication of Angrist and Evans (1998), some of the unnormalized estimates of the effects of childbearing on female labor market outcomes are positive, which is again not believable.

Running example: Card (1995)

Data: 3,010 workers with valid information on wage and education from the 1976 subset of the National Longitudinal Survey of Young Men (NLSYM).

Outcome: log wage in 1976
Treatment: whether completed at least one year, or at least four years of college
Extra covariates: one specification based on Card (1995), X_C; another based on Kitagawa (2015), X_K
IV for treatment: whether lived close to a college as a teenager

Data: 3,010 workers with valid information on wage and education from the 1976 subset of the National Longitudinal Survey of Young Men (NLSYM).

Outcome: log wage in 1976
Treatment: whether completed at least one year, or at least four years of college
Extra covariates: one specification based on Card (1995), X_C; another based on Kitagawa (2015), X_K
IV for treatment: whether lived close to a college as a teenager

IV estimates:

	Some college		Four-year degree	
	X_C	X_K	X_C	X_K
IV	0.661	0.575	1.392	0.991
	(0.294)	(0.308)	(0.798)	(0.610)

Outline

Recap of Abadie (2003)

Estimation of LATE

Empirical Illustrations

- Stata Module KAPPALATE

Monte Carlo Simulations

Recap of Abadie (2003)

Framework & Notation

Potential outcomes: Y_1 is outcome if $D = 1$, Y_0 is outcome if $D = 0$; what follows, $Y = (1 - D)Y_0 + DY_1$.

Treatment effect: $Y_{1i} - Y_{0i}$ for individual i.

There are two potential treatments, too: D_1 is treatment if $Z = 1$, D_0 is treatment if $Z = 0$; what follows, $D = (1 - Z)D_0 + ZD_1$.

Standard terminology: if $D_1 = 1$ and $D_0 = 1$, *always takers*; if $D_1 = 1$ and $D_0 = 0$, *compliers*; if $D_1 = 0$ and $D_0 = 1$, *defiers*; finally, if $D_1 = 0$ and $D_0 = 0$, *never takers*.

Potential outcomes: Y_1 is outcome if $D = 1$, Y_0 is outcome if $D = 0$; what follows, $Y = (1 - D)Y_0 + DY_1$.

Treatment effect: $Y_{1i} - Y_{0i}$ for individual i.

There are two potential treatments, too: D_1 is treatment if $Z = 1$, D_0 is treatment if $Z = 0$; what follows, $D = (1 - Z)D_0 + ZD_1$.

Standard terminology: if $D_1 = 1$ and $D_0 = 1$, *always takers*; if $D_1 = 1$ and $D_0 = 0$, *compliers*; if $D_1 = 0$ and $D_0 = 1$, *defiers*; finally, if $D_1 = 0$ and $D_0 = 0$, *never takers*.

LATE

$$\tau_{LATE} = \mathbb{E}[Y_1 - Y_0 | D_1 > D_0]. \tag{2}$$

Identifying Assumptions (Abadie, 2003)

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

$$\mathbb{P}[Y_{d,1} = Y_{d,0} | X = x] = 1. \square$$

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

$$[Y_0, Y_1, D_1, D_0] \perp Z | X. \square$$

Assumption 3 (Monotonicity):

$$\mathbb{P}[D_1 \geq D_0] = 1. \square$$

Assumption 4 (Existence of Compliers):

$$\mathbb{P}[D_1 > D_0] > 0. \square$$

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

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

Main result in Abadie (2003)

Theorem (Abadie 2003, pp. 236–237)
Let $g(\cdot)$ be any measurable real function of (Y, D, X) such that $E|g(Y, D, X)| < \infty$. Let $p(X) = P(Z = 1|X)$. Define

$$\begin{aligned} \kappa_0 &= (1 - D)\,\frac{(1 - Z) - (1 - p(X))}{p(X)\,(1 - p(X))}, \\ \kappa_1 &= D\,\frac{Z - p(X)}{p(X)\,(1 - p(X))}, \\ \kappa = \kappa_0\,(1 - p(X)) + \kappa_1 p(X) &= 1 - \frac{D\,(1 - Z)}{1 - p(X)} - \frac{(1 - D)\,Z}{p(X)}. \end{aligned}$$

Under Assumption 2.1 (Abadie 2003, pp. 234–235),

(a) $E\left[g(Y, D, X)|D_1 > D_0\right] = \frac{1}{P(D_1 > D_0)} E\left[\kappa\, g(Y, D, X)\right]$. *Also,*

(b) $E\left[g(Y_0, X)|D_1 > D_0\right] = \frac{1}{P(D_1 > D_0)} E\left[\kappa_0\, g(Y, X)\right]$, *and*

(c) $E\left[g(Y_1, X)|D_1 > D_0\right] = \frac{1}{P(D_1 > D_0)} E\left[\kappa_1\, g(Y, X)\right]$.

Identification of LATE

- To see how Abadie's theorem identifies the LATE, take $g(Y_0, X) = Y_0$ and $g(Y_1, X) = Y_1$, and write:

$$\tau_{LATE} = \frac{1}{\Pr[D_1 > D_0]} \operatorname{E}[\kappa_1 Y] - \frac{1}{\Pr[D_1 > D_0]} \operatorname{E}[\kappa_0 Y]. \tag{3}$$

- Equivalently:

$$\begin{aligned} \tau_{LATE} &= \frac{1}{\Pr[D_1 > D_0]} \operatorname{E}[(\kappa_1 - \kappa_0) Y] \\ &= \frac{1}{\Pr[D_1 > D_0]} \operatorname{E}\left[Y \frac{Z - p(X)}{p(X)(1 - p(X))}\right]. \end{aligned} \tag{4}$$

- As noted by Abadie (2003), $\Pr[D_1 > D_0] = \operatorname{E}[\kappa]$.
- Similarly, $\Pr[D_1 > D_0] = \operatorname{E}[\kappa_1]$ and $\Pr[D_1 > D_0] = \operatorname{E}[\kappa_0]$.

Estimation of LATE

Estimation of LATE

Given a random sample $\{(D_i, Z_i, X_i, Y_i) : i = 1, \ldots, N\}$, equation (4) suggests that we can consistently estimate τ_{LATE} as follows:

$$\widehat{\tau}_{LATE} = \frac{1}{\widehat{\Pr}\left[D_1 > D_0\right]} \left[N^{-1} \sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)} \right], \qquad (5)$$

where $\widehat{\Pr}\left[D_1 > D_0\right] \xrightarrow{p} \Pr\left[D_1 > D_0\right] > 0$.

Given a random sample $\{(D_i, Z_i, X_i, Y_i) : i = 1, \ldots, N\}$, equation (4) suggests that we can consistently estimate τ_{LATE} as follows:

$$\widehat{\tau}_{LATE} = \frac{1}{\widehat{\Pr}\left[D_1 > D_0\right]} \left[N^{-1} \sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)} \right], \tag{5}$$

where $\widehat{\Pr}\left[D_1 > D_0\right] \xrightarrow{p} \Pr\left[D_1 > D_0\right] > 0$.

Our discussion so far also implies that there are three candidate estimators for $\Pr\left[D_1 > D_0\right]$, namely $N^{-1} \sum_{i=1}^{N} \kappa_i$, $N^{-1} \sum_{i=1}^{N} \kappa_{i1}$, and $N^{-1} \sum_{i=1}^{N} \kappa_{i0}$.

Three estimators of LATE

It follows from equation (5) that we have the following consistent estimators of τ_{LATE}:

$$\widehat{\tau}_a = \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{6}$$

Three estimators of LATE

It follows from equation (5) that we have the following consistent estimators of τ_{LATE}:

$$\widehat{\tau}_a = \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{6}$$

$$\widehat{\tau}_{a,1} = \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{7}$$

Three estimators of LATE

It follows from equation (5) that we have the following consistent estimators of τ_{LATE}:

$$\widehat{\tau}_a = \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{6}$$

$$\widehat{\tau}_{a,1} = \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{7}$$

$$\widehat{\tau}_{a,0} = \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right]. \tag{8}$$

Three estimators of LATE

It follows from equation (5) that we have the following consistent estimators of τ_{LATE}:

$$\widehat{\tau}_a = \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{6}$$

$$\widehat{\tau}_{a,1} = \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right], \tag{7}$$

$$\widehat{\tau}_{a,0} = \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} Y_i \frac{Z_i - p(X_i)}{p(X_i)\left(1 - p(X_i)\right)}\right]. \tag{8}$$

Even though $\mathrm{E}\left[\kappa\right] = \mathrm{E}\left[\kappa_1\right] = \mathrm{E}\left[\kappa_0\right]$, $N^{-1}\sum_{i=1}^{N}\kappa_i$, $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$, and $N^{-1}\sum_{i=1}^{N}\kappa_{i0}$ will generally be different.

Tan (2006)'s estimator

Given our interest in weighting estimators, a natural candidate estimator is

$$\widehat{\tau}_t = \left[\sum_{i=1}^{N} \frac{D_i Z_i}{p(X_i)} - \sum_{i=1}^{N} \frac{D_i \left(1 - Z_i\right)}{1 - p(X_i)} \right]^{-1} \left[\sum_{i=1}^{N} \frac{Y_i Z_i}{p(X_i)} - \sum_{i=1}^{N} \frac{Y_i \left(1 - Z_i\right)}{1 - p(X_i)} \right],$$

which was first suggested by Tan (2006).

Tan (2006)'s estimator

Given our interest in weighting estimators, a natural candidate estimator is

$$\widehat{\tau}_t = \left[\sum_{i=1}^{N} \frac{D_i Z_i}{p(X_i)} - \sum_{i=1}^{N} \frac{D_i \left(1 - Z_i\right)}{1 - p(X_i)} \right]^{-1} \left[\sum_{i=1}^{N} \frac{Y_i Z_i}{p(X_i)} - \sum_{i=1}^{N} \frac{Y_i \left(1 - Z_i\right)}{1 - p(X_i)} \right],$$

which was first suggested by Tan (2006).

This is by far the most popular weighting estimator of the LATE. See, for example, Frölich (2007), MaCurdy et al. (2011), Donald et al. (2014a,b), and Abdulkadiroğlu et al. (2017).

Tan (2006)'s estimator

Given our interest in weighting estimators, a natural candidate estimator is

$$\widehat{\tau}_t = \left[\sum_{i=1}^{N} \frac{D_i Z_i}{p(X_i)} - \sum_{i=1}^{N} \frac{D_i (1 - Z_i)}{1 - p(X_i)}\right]^{-1} \left[\sum_{i=1}^{N} \frac{Y_i Z_i}{p(X_i)} - \sum_{i=1}^{N} \frac{Y_i (1 - Z_i)}{1 - p(X_i)}\right],$$

which was first suggested by Tan (2006).

This is by far the most popular weighting estimator of the LATE. See, for example, Frölich (2007), MaCurdy et al. (2011), Donald et al. (2014a,b), and Abdulkadiroğlu et al. (2017).

Remark

$\widehat{\tau}_t = \widehat{\tau}_{a,1}$.

Running example: Card (1995)

Something is really off here...

	Some college		Four-year degree	
	X_C	X_K	X_C	X_K
$\widehat{\tau}_a$	–0.319	2.248	–0.594	4.317
	(1.182)	(0.971)	(2.184)	(2.485)
$\widehat{\tau}_t = \widehat{\tau}_{a,1}$	–0.321	2.053	–0.601	3.651
	(1.201)	(0.813)	(2.251)	(1.780)
$\widehat{\tau}_{a,0}$	–0.290	2.846	–0.501	7.241
	(1.036)	(1.592)	(1.728)	(7.245)

Unnormalized and normalized weights

Unnormalized and normalized weights

An important point, due to Imbens (2004), Millimet and Tchernis (2009), and Busso et al. (2014) in the context of treatment effects under unconfoundedness, is that **weighting estimators perform badly when the weights are not normalized** or, in other words, **when they do not sum to one** in finite samples.

Unnormalized and normalized weights

An important point, due to Imbens (2004), Millimet and Tchernis (2009), and Busso et al. (2014) in the context of treatment effects under unconfoundedness, is that **weighting estimators perform badly when the weights are not normalized** or, in other words, **when they do not sum to one** in finite samples.

It follows immediately that $\widehat{\tau}_t$ is likely inferior to the ratio of two normalized estimators of the ATE under unconfoundedness:

$$\widehat{\tau}_{t,norm} = \frac{\left[\sum_{i=1}^{N} \frac{Z_i}{p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i Z_i}{p(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{Y_i(1-Z_i)}{1-p(X_i)}}{\left[\sum_{i=1}^{N} \frac{Z_i}{p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i Z_i}{p(X_i)} - \left[\sum_{i=1}^{N} \frac{1-Z_i}{1-p(X_i)}\right]^{-1} \sum_{i=1}^{N} \frac{D_i(1-Z_i)}{1-p(X_i)}},$$

which was first suggested by Uysal (2011) and subsequently discussed by Bodory and Huber (2018) and Heiler (2021).

$\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized

To see that $\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized, note that each of these estimators can equivalently be represented as an analogue of equation (3):

To see that $\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized, note that each of these estimators can equivalently be represented as an analogue of equation (3):

$$\widehat{\tau}_a = \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right],$$

$\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized

To see that $\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized, note that each of these estimators can equivalently be represented as an analogue of equation (3):

$$
\begin{aligned}
\widehat{\tau}_a &= \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right], \\
\widehat{\tau}_{a,1} &= \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right],
\end{aligned}
$$

$\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized

To see that $\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized, note that each of these estimators can equivalently be represented as an analogue of equation (3):

$$
\begin{aligned}
\widehat{\tau}_a &= \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_i\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right], \\
\widehat{\tau}_{a,1} &= \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_{i1}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right], \\
\widehat{\tau}_{a,0} &= \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i1} Y_i\right] - \left[\sum_{i=1}^{N} \kappa_{i0}\right]^{-1} \left[\sum_{i=1}^{N} \kappa_{i0} Y_i\right].
\end{aligned}
$$

$\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized

To see that $\widehat{\tau}_a$, $\widehat{\tau}_{a,1}$, and $\widehat{\tau}_{a,0}$ are unnormalized, note that each of these estimators can equivalently be represented as an analogue of equation (3):

$$
\begin{aligned}
\widehat{\tau}_a &= \left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}Y_i\right]-\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}Y_i\right],\\
\widehat{\tau}_{a,1} &= \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}Y_i\right]-\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}Y_i\right],\\
\widehat{\tau}_{a,0} &= \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}Y_i\right]-\left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}Y_i\right].
\end{aligned}
$$

It turns out that neither of these estimators is properly normalized. For example, $\widehat{\tau}_a$ uses weights of $\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\kappa_{i1}$ and $\left[\sum_{i=1}^{N}\kappa_i\right]^{-1}\kappa_{i0}$, which do not sum to unity across i.

How to normalize?

It is straightforward to construct a normalized Abadie estimator of the LATE. It turns out that the two denominators in equation (3) need to be estimated separately, using different estimators of $\Pr\left[D_1 > D_0\right]$, $N^{-1}\sum_{i=1}^{N}\kappa_{i1}$ and $N^{-1}\sum_{i=1}^{N}\kappa_{i0}$. The resulting estimator becomes

$$\widehat{\tau}_{a,10} = \left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i1}Y_i\right] - \left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\left[\sum_{i=1}^{N}\kappa_{i0}Y_i\right], \qquad (9)$$

where both sets of weights, $\left[\sum_{i=1}^{N}\kappa_{i1}\right]^{-1}\kappa_{i1}$ and $\left[\sum_{i=1}^{N}\kappa_{i0}\right]^{-1}\kappa_{i0}$, necessarily sum to unity across i.

Why is it so important that weights sum to unity?

Many of the recommendations to date are based on simulation results (e.g., Millimet and Tchernis, 2009; Busso et al., 2014), and it is not clear to what extent such evidence should guide estimator choice (cf. Advani et al., 2019).

Why is it so important that weights sum to unity?

Many of the recommendations to date are based on simulation results (e.g., Millimet and Tchernis, 2009; Busso et al., 2014), and it is not clear to what extent such evidence should guide estimator choice (cf. Advani et al., 2019).

We provide an objective and intuitively appealing criterion that differentiates the normalized from the unnormalized estimators.

Why is it so important that weights sum to unity?

Many of the recommendations to date are based on simulation results (e.g., Millimet and Tchernis, 2009; Busso et al., 2014), and it is not clear to what extent such evidence should guide estimator choice (cf. Advani et al., 2019).

We provide an objective and intuitively appealing criterion that differentiates the normalized from the unnormalized estimators.

Let $\mathbf{Y}$ be a vector of observed data on outcomes and $\mathbf{W} = (\mathbf{D}\ \mathbf{Z}\ \mathbf{X})$ be a matrix of observed data on the remaining variables.

Why is it so important that weights sum to unity?

TI (Translation Invariance)

We say that an estimator $\widehat{\tau} = \widehat{\tau}(\mathbf{Y}, \mathbf{W})$ is translation invariant if $\widehat{\tau}(\mathbf{Y}, \mathbf{W}) = \widehat{\tau}(\mathbf{Y} + k, \mathbf{W})$ for all $\mathbf{Y}$, $\mathbf{W}$, and k.

Why is it so important that weights sum to unity?

TI (Translation Invariance)

We say that an estimator $\widehat{\tau} = \widehat{\tau}(\mathbf{Y}, \mathbf{W})$ is translation invariant if $\widehat{\tau}(\mathbf{Y}, \mathbf{W}) = \widehat{\tau}(\mathbf{Y} + k, \mathbf{W})$ for all $\mathbf{Y}$, $\mathbf{W}$, and k.

SI (Scale Invariance)

We say that an estimator $\widehat{\tau} = \widehat{\tau}(\mathbf{Y}, \mathbf{W})$ is scale invariant with respect to g if $\widehat{\tau}(f(\mathbf{Y}), \mathbf{W}) = \widehat{\tau}(f(a\mathbf{Y}), \mathbf{W})$, $f(\mathbf{Y}) = (g(Y_1), \ldots, g(Y_N))$, for all $\mathbf{Y} > 0$, $\mathbf{W}$, and $a > 0$.

Why is it so important that weights sum to unity?

TI (Translation Invariance)

We say that an estimator $\widehat{\tau} = \widehat{\tau}(\mathbf{Y}, \mathbf{W})$ is translation invariant if $\widehat{\tau}(\mathbf{Y}, \mathbf{W}) = \widehat{\tau}(\mathbf{Y} + k, \mathbf{W})$ for all $\mathbf{Y}$, $\mathbf{W}$, and k.

SI (Scale Invariance)

We say that an estimator $\widehat{\tau} = \widehat{\tau}(\mathbf{Y}, \mathbf{W})$ is scale invariant with respect to g if $\widehat{\tau}(f(\mathbf{Y}), \mathbf{W}) = \widehat{\tau}(f(a\mathbf{Y}), \mathbf{W})$, $f(\mathbf{Y}) = (g(Y_1), \ldots, g(Y_N))$, for all $\mathbf{Y} > 0$, $\mathbf{W}$, and $a > 0$.

Proposition

$\widehat{\tau}_{t,norm}$ and $\widehat{\tau}_{a,10}$ are translation invariant and scale invariant with respect to the natural logarithm. $\widehat{\tau}_a$, $\widehat{\tau}_t$ $(= \widehat{\tau}_{a,1})$, and $\widehat{\tau}_{a,0}$ are not translation invariant and not scale invariant with respect to the natural logarithm.

Near-zero denominators

Near-zero denominators

Weighting estimators of the LATE, like two-stage least squares and many other IV methods, are an example of ratio estimators. A common problem with such estimators is that they **behave badly if their denominator is close to zero**.

Weighting estimators of the LATE, like two-stage least squares and many other IV methods, are an example of ratio estimators. A common problem with such estimators is that they **behave badly if their denominator is close to zero**.

In this paper, we identify two situations under which certain *unnormalized* estimators have the advantage of being based on a denominator that is nonnegative by construction and bounded away from zero in all practically relevant situations.

Near-zero denominators

Weighting estimators of the LATE, like two-stage least squares and many other IV methods, are an example of ratio estimators. A common problem with such estimators is that they **behave badly if their denominator is close to zero**.

In this paper, we identify two situations under which certain *unnormalized* estimators have the advantage of being based on a denominator that is nonnegative by construction and bounded away from zero in all practically relevant situations.

To be specific, $\widehat{\tau}_t = \widehat{\tau}_{a,1}$ has this property **when there are no always-takers** while $\widehat{\tau}_{a,0}$ has this property **when there are no never-takers**.

Simplified formulas for κ, κ_1, and κ_0

	κ	$\text{sgn}(\kappa)$	κ_1	$\text{sgn}(\kappa_1)$	κ_0	$\text{sgn}(\kappa_0)$
$Z = 1, D = 1$	1	$+$	$\frac{1}{p(X)}$	$+$	0	0
$Z = 1, D = 0$	$-\frac{1-p(X)}{p(X)}$	$-$	0	0	$-\frac{1}{p(X)}$	$-$
$Z = 0, D = 1$	$-\frac{p(X)}{1-p(X)}$	$-$	$-\frac{1}{1-p(X)}$	$-$	0	0
$Z = 0, D = 0$	1	$+$	0	0	$\frac{1}{1-p(X)}$	$+$

Simplified formulas for κ, κ_1, and κ_0

	κ	$\text{sgn}(\kappa)$	κ_1	$\text{sgn}(\kappa_1)$	κ_0	$\text{sgn}(\kappa_0)$
$Z=1, D=1$	1	$+$	$\frac{1}{p(X)}$	$+$	0	0
$Z=1, D=0$	$-\frac{1-p(X)}{p(X)}$	$-$	0	0	$-\frac{1}{p(X)}$	$-$
$Z=0, D=1$	$-\frac{p(X)}{1-p(X)}$	$-$	$-\frac{1}{1-p(X)}$	$-$	0	0
$Z=0, D=0$	1	$+$	0	0	$\frac{1}{1-p(X)}$	$+$

Remark

If there are no always-takers, $N^{-1}\sum_{i=1}^{N}\kappa_{i1} > \widehat{\Pr}\left[D=1\right] > 0$.

Simplified formulas for κ, κ_1, and κ_0

	κ	$\text{sgn}(\kappa)$	κ_1	$\text{sgn}(\kappa_1)$	κ_0	$\text{sgn}(\kappa_0)$
$Z=1, D=1$	1	$+$	$\frac{1}{p(X)}$	$+$	0	0
$Z=1, D=0$	$-\frac{1-p(X)}{p(X)}$	$-$	0	0	$-\frac{1}{p(X)}$	$-$
$Z=0, D=1$	$-\frac{p(X)}{1-p(X)}$	$-$	$-\frac{1}{1-p(X)}$	$-$	0	0
$Z=0, D=0$	1	$+$	0	0	$\frac{1}{1-p(X)}$	$+$

Remark
If there are no always-takers, $N^{-1}\sum_{i=1}^{N}\kappa_{i1} > \widehat{\Pr}[D=1] > 0$.

Remark
If there are no never-takers, $N^{-1}\sum_{i=1}^{N}\kappa_{i0} > \widehat{\Pr}[D=0] > 0$.

Estimation of the Instrument Propensity Score (IPS)

Estimation of the Instrument Propensity Score (IPS) by Maximum Likelihood

- In practice the IPS ($p(X)$) is usually unknown
- Typically researchers adopt a parametric model for $p(X)$, say $F(X, \alpha)$, and estimate the unknown parameters by maximum likelihood.
- The corresponding score function

$$\mathrm{E}\left[\frac{(Z_i - F(X_i, \alpha))}{F(X_i, \alpha)(1 - F(X_i, \alpha))} \nabla_\alpha F(X_i, \alpha)\right] \tag{10}$$

- Solve for $\widehat{\alpha}_{ML}$ and plug in $F(X_i, \widehat{\alpha}_{ML})$

Estimation of the Instrument Propensity Score (IPS) by covariate balancing approach

- Alternatively, one could estimate the IPS using the covariate balancing approach proposed by Imai and Ratkovis (2014).
- The moment function that balances the first moments of the covariates:

$$E\left[\frac{(Z_i - F(X_i, \alpha))}{F(X_i, \alpha)(1 - F(X_i, \alpha))} X_i\right] = 0 \tag{11}$$

- Solve the sample counterpart of the moment equation for $\widehat{\alpha}_{cb}$ and get the estimated IPS plug in $F(X_i, \widehat{\alpha}_{cb})$

Estimation of the Instrument Propensity Score (IPS) by covariate balancing approach

Proposition

If X includes a constant, $\widehat{\tau}_t$ $(= \widehat{\tau}_{a,1})$, $\widehat{\tau}_{a,0}$, $\widehat{\tau}_{a,10}$, and $\widehat{\tau}_{t,norm}$ are numerically identical when estimated with the covariate balancing IPS.

Note: We denote any of those by $\widehat{\tau}_{cb}$.

- Any of these estimators with covariate balancing IPS is translation invariant and scale invariant with respect to $g(Y) = \log(Y)$.
- At the same time, because they share the structure of $\widehat{\tau}_t$ $(= \widehat{\tau}_{a,1})$ and $\widehat{\tau}_{a,0}$, it avoids near-zero denominators when there are no always-takers *and also* when there are instead no never-takers.

Empirical Illustrations

Stata Module KAPPALATE

- You can install the module
  `ssc install kappalate`
- The syntax of the kappalate command is:
  `kappalate depvar [indepvars] (treatment = instrument) [if] [in] [, options]`

Options	*Description*
zmodel(*string*)	select the approach to estimating the instrument propensity score; options include **logit**, **probit**, and **cbps**; default is **cbps**
vce(*vcetype*)	vcetype is passed on to Stata's **gmm** command and specifies the type of standard error reported; default is **robust**
std(*string*)	string may be **on** or **off**, which determines whether nonbinary covariates are standardized prior to estimation; default is **on**
which(*string*)	string may be **all** or **norm**, which determines whether all estimates or only normalized estimates are displayed; default is **norm**

Example 1: Angrist (1990)

- Study of causal effects of military service using the draft eligibility instrument
- Data: a sample of 3,027 individuals from the 1984 Survey of Income andProgram Participation (SIPP)
- Outcome: log(wage(dollars)), log(wage(cents))
- Treatment: an indicator for being veteran
- Instrument: an indicator for whether an individual had a lottery number below the draft eligibility ceiling
- Controls: because the ceilings were cohort specific, it is essential to control for age in subsequent analysis.

Example 1: Angrist (1990) cont'd

```
. kappalate lwage (nvstat=rsncode) age, which(all) zmodel(logit)

Weighting estimation of the LATE

Outcome     :   lwage
Treatment   :   nvstat
Instrument  :   rsncode
IPS         :   logit

Number of obs   =   3027
```

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tau_a	.0146861	.206792	0.07	0.943	-.3906189	.4199911
tau_a,1	.0155666	.2191482	0.07	0.943	-.4139559	.4450892
tau_a,0	.0141231	.1988904	0.07	0.943	-.3756948	.4039411
tau_a,10	.2267641	.2037929	1.11	0.266	-.1726627	.6261909
tau_t,norm	.2341665	.2110309	1.11	0.267	-.1794465	.6477794

```
. kappalate lwage (nvstat=rsncode) age, zmodel(cbps)

Weighting estimation of the LATE

Outcome     :   lwage
Treatment   :   nvstat
Instrument  :   rsncode
IPS         :   CBPS

Number of obs   =   3027
```

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tau_t,norm	.2294623	.2130195	1.08	0.281	-.1880483	.6469729

Example 1: Angrist (1990) cont'd

```
. kappalate lwage_cnt (nvstat=rsncode) age, which(all) zmodel(logit)

Weighting estimation of the LATE

Outcome     :   lwage_cnt
Treatment   :   nvstat
Instrument  :   rsncode
IPS         :   logit

Number of obs    =   3027
```

lwage_cnt	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tau_a	-.4293729	.2582931	-1.66	0.096	-.9356181	.0768723
tau_a,1	-.4551167	.2785057	-1.63	0.102	-1.000978	.0907445
tau_a,0	-.4129138	.2456693	-1.68	0.093	-.8944167	.0685892
tau_a,10	.2267641	.2037929	1.11	0.266	-.1726627	.6261909
tau_t,norm	.2341665	.2110309	1.11	0.267	-.1794465	.6477794

```
. kappalate lwage_cnt (nvstat=rsncode) age, zmodel(cbps)

Weighting estimation of the LATE

Outcome     :   lwage_cnt
Treatment   :   nvstat
Instrument  :   rsncode
IPS         :   CBPS

Number of obs    =   3027
```

lwage_cnt	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tau_t,norm	.2294623	.2130195	1.08	0.281	-.1880483	.6469729

Example 1: Angrist (1990) cont'd

Table 1: Causal Effects of Military Service on Log Wages

Controls	age	age	cubic in age	cubic in age	i.age	i.age
Outcome	ln(dollars)	ln(cents)	ln(dollars)	ln(cents)	ln(dollars)	ln(cents)
τ_a	0.015	-0.429*	0.314	0.537*	0.241	0.241
	(0.207)	(0.258)	(0.252)	(0.322)	(0.229)	(0.229)
$\tau_{a,1}$	0.016	-0.455	0.302	0.515*	0.241	0.241
	(0.219)	(0.279)	(0.240)	(0.301)	(0.229)	(0.229)
$\tau_{a,0}$	0.014	-0.413*	0.317	0.540*	0.241	0.241
	(0.199)	(0.246)	(0.255)	(0.326)	(0.229)	(0.229)
$\tau_{a,10}$	0.227	0.227	0.204	0.204	0.241	0.241
	(0.204)	(0.204)	(0.239)	(0.239)	(0.229)	(0.229)
$\tau_{t,norm}$	0.234	0.234	0.202	0.202	0.241	0.241
	(0.211)	(0.211)	(0.235)	(0.235)	(0.229)	(0.229)
τ_{cb}	0.229	0.229	0.208	0.208	0.241	0.241
	(0.213)	(0.213)	(0.232)	(0.232)	(0.229)	(0.229)

Example 2: Card (1995)

- **Data:** 3,010 workers with valid information on wage and education from the 1976 subset of the National Longitudinal Survey of Young Men (NLSYM).
- **Outcome:** log of wage measured in dollars or in cents (1976)
- **Treatment:** whether completed at least one year, or at least four years of college
- **Extra covariates:** one specification based on Card (1995), X_C; another based on Kitagawa (2015), X_K
- **IV for treatment:** whether lived close to a college as a teenager

Example 2: Card (1995)

Table 2: Causal Effects of College Education on Log Wages

treatment	Some College				College Graduate			
controls	x_C		x_K		x_C		x_K	
outcome	ln(dollars)	ln(cents)	ln(dollars)	ln(cents)	ln(dollars)	ln(cents)	ln(dollars)	ln(cents)
τ_a	0.170	-0.319	0.842**	2.248**	0.315	-0.594	1.617*	4.317*
	(0.370)	(1.182)	(0.362)	(0.971)	(0.696)	(2.184)	(0.891)	(2.485)
$\tau_{a,1}$	0.171	-0.321	0.769**	2.053**	0.319	-0.601	1.367**	3.651**
	(0.367)	(1.201)	(0.308)	(0.813)	(0.687)	(2.251)	(0.648)	(1.780)
$\tau_{a,0}$	0.154	-0.290	1.066*	2.846*	0.266	-0.501	2.712	7.241
	(0.354)	(1.036)	(0.574)	(1.592)	(0.639)	(1.728)	(2.577)	(7.246)
$\tau_{a,10}$	0.346*	0.346*	0.293	0.293	0.586*	0.586*	0.836	0.836
	(0.200)	(0.200)	(0.252)	(0.252)	(0.356)	(0.356)	(0.821)	(0.821)
$\tau_{t,norm}$	0.331	0.331	0.356	0.356	0.619	0.619	0.628	0.628
	(0.202)	(0.202)	(0.244)	(0.244)	(0.387)	(0.387)	(0.448)	(0.448)
τ_{cb}	0.376*	0.376*	0.331	0.331	0.853	0.853	0.588	0.588
	(0.223)	(0.223)	(0.236)	(0.236)	(0.549)	(0.549)	(0.433)	(0.433)

Example 3: Angrist and Evans (1998)

- **Data:** Farbmacher et al.'s (2018) subsample of the 1980 US Census that consists of all women aged 21–35 with at least two children (sample size = 394,840)
- **Outcome:** log income and an indicator for labor force participation
- **Treatment:** having more than two children
- **Covariates:** age, age at first birth, sex of the first and second children, and indicators for whether Black, whether Hispanic, and whether another race
- **IV for treatment:** whether the first two children areof the same sex

Example 3: Angrist and Evans (1998)

Table 3: Causal Effects of Childbearing on Labor Force Participation and Log Income

outcome	**labor force participation**			**log income**			
unit	0,1	1,2	1,0	dollars	cents	1,000s $	100,000s $
τ_a	-0.100***	-0.070***	-0.131***	0.143	0.286**	-0.073	-0.216**
	(0.025)	(0.026)	(0.025)	(0.102)	(0.113)	(0.093)	(0.093)
$\tau_{a,1}$	-0.099***	-0.069***	-0.129***	0.140	0.282**	-0.072	-0.213**
	(0.025)	(0.025)	(0.025)	(0.100)	(0.111)	(0.092)	(0.091)
$\tau_{a,0}$	-0.102***	-0.071***	-0.133***	0.145	0.291**	-0.074	-0.220**
	(0.025)	(0.026)	(0.026)	(0.104)	(0.115)	(0.094)	(0.094)
$\tau_{a,10}$	-0.117***	-0.117***	-0.117***	-0.132	-0.132	-0.132	-0.132
	(0.025)	(0.025)	(0.025)	(0.093)	(0.093)	(0.093)	(0.093)
$\tau_{t,norm}$	-0.117***	-0.117***	-0.117***	-0.135	-0.135	-0.135	-0.135
	(0.025)	(0.025)	(0.025)	(0.092)	(0.092)	(0.092)	(0.092)
τ_{cb}	-0.117***	-0.117***	-0.117***	-0.135	-0.135	-0.135	-0.135
	(0.025)	(0.025)	(0.025)	(0.092)	(0.092)	(0.092)	(0.092)

Monte Carlo Simulations

DGP

We focus on data-generating processes from Heiler (2022):

$$\begin{aligned}
Z &= 1[u < \pi(X)], \\
\pi(X) &= 1/\left(1 + \exp\left(-\mu_z(X) \cdot \theta_0\right)\right), \\
D_z &= 1[\mu_d(X, z) > v], \\
Y_1 &= \mu_{y_1}(X) + \varepsilon_1, \\
Y_0 &= \varepsilon_0,
\end{aligned}$$

where u and X are i.i.d. standard uniform,

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_0 \\ v \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix} \right),$$ $\theta_0 = \ln((1-\delta)/\delta)$, and $\delta \in \{0.01, 0.02, 0.05\}$.

Table 4: Simulation Designs

	Design A.1	Design A.2	Design B	Design C	Design D
$\mu_d(x,z)$	$4z$	$4(z-1)$	$-1+2x+2.122z$	$-1+2x+2.122z$	$-1+2x+2.122z$
$\mu_{y_1}(x)$	0.3989	0.3989	0.3989	$9(x+3)^2$	$9(x+3)^2$
$\mu_z(x)$	$2x-1$	$2x-1$	$2x-1$	$2x-1$	$x+x^2-1$

Designs

- To illustrate our findings from on near-zero denominators
  - Design A.1: with no never-takers.
  - Design A.2: with no always takers
- Designs A.1 and A.2 correspond to the case of a fully independent instrument while in the remaining designs the instrument is conditionally independent.
- In Designs A.1, A.2, and B, treatment effect heterogeneity is only due to the correlation between ε_1 and v
- In Designs C and D, on the other hand, the dependence of $\mu_{y_1}(X)$ on X constitutes another source of heterogeneity.
- The linear IV estimator that controls for X is expected to perform very well in Designs A.1, A.2, and B but not necessarily elsewhere.

Estimation

- The linear IV estimator as a benchmark
- Normalized estimators $\widehat{\tau}_{t,norm}$, $\widehat{\tau}_{a,10}$ (with the IPS estimated by MLE) and $\widehat{\tau}_{cb}$, controlling for X.
- Unnormalized estimators $\widehat{\tau}_{a}$, $\widehat{\tau}_{a,1}$ $(= \widehat{\tau}_{t})$, and $\widehat{\tau}_{a,0}$ with the IPS estimated by MLE, also controlling for X.
- We consider three sample sizes, $N = 500$, $N = 1{,}000$, and $N = 5{,}000$, and 10,000 replications for each combination of a design, a value of δ, and a sample size.

Summary of Simulation Results

- $\widehat{\tau}_t$, $\widehat{\tau}_a$, and $\widehat{\tau}_{a,10}$, are mostly very unstable when overlap is sufficiently poor and samples are small
- $\widehat{\tau}_{a,0}$ does not suffer from instability in no never takers case
- $\widehat{\tau}_{a,1}$ does not suffer from instability in no always takers case
- $\widehat{\tau}_{t,norm}$ and $\widehat{\tau}_{cb}$ perform better than other weighting estimators in term of MSE (in many cases also in terms of bias)
- Design D: $\widehat{\tau}_{cb}$ dominates all other estimators in terms of MSE, bias, and coverage.

Conclusion

This paper studies the properties of various **weighting estimators of the local average treatment effect (LATE)**, several of which are based on the identification results of Abadie (2003) and Frölich (2007).

We make several novel observations:

- on the scale invariance with respect to the natural logarithm and translation invariance of normalized estimators
- on the advantages of some estimators in case of one-sided noncompliance
- on the desirable properties of some estimator when IPS is estimated by the covariate balancing method

We illustrate our findings with a simulation study and three empirical applications.

- In simulations, the covariate balancing estimator and the normalized version of Tan's (2006) estimator perform relatively well in every setting under consideration.
- In empirical applications, each of the unnormalized estimators appears to be unreliable in at least some cases.