Advance Access publication August 20, 2013

Political Analysis (2013) 21:492–506
doi:10.1093/pan/mpt013

Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable

Peter M. Aronow
Department of Political Science, Yale University, New Haven, CT 06520
e-mail: peter.aronow@yale.edu (corresponding author)

Allison Carnegie
Department of Political Science, Princeton University, Princeton, NJ 08540, and
Department of Political Science, University of Chicago, Chicago, IL 60637
e-mail: acarnegie@uchicago.edu

Edited by Jonathan Katz

Political scientists frequently use instrumental variables (IV) estimation to estimate the causal effect of an endogenous treatment variable. However, when the treatment effect is heterogeneous, this estimation strategy only recovers the local average treatment effect (LATE). The LATE is an average treatment effect (ATE) for a subset of the population: units that receive treatment if and only if they are induced by an exogenous IV. However, researchers may instead be interested in the ATE for the entire population of interest. In this article, we develop a simple reweighting method for estimating the ATE, shedding light on the identification challenge posed in moving from the LATE to the ATE. We apply our method to two published experiments in political science in which we demonstrate that the LATE has the potential to substantively differ from the ATE.

1 Introduction

Instrumental variables (IV) estimation, although commonly used in political science, is subject to an often overlooked limitation. IV estimation (Wald 1940) allows researchers to estimate the causal effect of an endogenous treatment variable through the use of an exogenous variable known as an IV. However, although many researchers intend to estimate the average treatment effect (ATE) for the entire population of interest, IV estimation only recovers the local average treatment effect (LATE) or the ATE for the subpopulation that is influenced by the IV. When treatment effects are heterogeneous across units, the LATE and the ATE may take on different values—potentially causing complications in interpretation.

Several influential papers, including Deaton (2009) and Heckman and Urzua (2010), have expressed concern about the interpretation of the LATE. Both papers argue that the LATE is not usually the parameter of interest, as "we are unlikely to learn much about the process at work" by estimating the LATE (Deaton 2009, 430) and the LATE "is often very difficult to interpret as an answer to an interesting economic question" (Heckman and Urzua 2010, 35). Heckman and Urzua (2010, 35) add that reliance on IV estimation often means that "problems of identification and interpretation are swept under the rug." Even Imbens (2009, 403), who advocates the use of IV

Authors' note: The replication archive for this article is available online at Aronow and Carnegie (2013). The authors acknowledge support from the Yale University Faculty of Arts and Sciences High-Performance Computing facility and staff. Helpful comments from Jake Bowers, John Bullock, Dan Butler, Lara Chausow, Adam Dynes, Ivan Fernandez-Val, Adam Glynn, Holger Kern, Malte Lierl, Mary McGrath, Joel Middleton, Cyrus Samii, two anonymous reviewers, and the participants of the Yale American Politics and Public Policy Workshop, the New Faces in Political Methodology Conference, and the Midwest Political Science Association Conference are greatly appreciated. We also thank Jonathan Katz for helpful editorial guidance. Special thanks to Bethany Albertson, Adria Lawrence, and David Nickerson for generous data sharing and to Dean Eckles, Alan Gerber, Don Green, Greg Huber, and Ken Scheve for particularly helpful conversations. All remaining errors are our own. Supplementary materials for this article are available on the *Political Analysis* Web site.

© The Author 2013. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

estimation, admits that "in many cases the local average treatment effects . . . are *not* the average effects that researchers set out to estimate." Although the authors disagree on the utility of IV estimation, they seem to largely agree on the limitations of the LATE.

The goal of this article is not to dissuade researchers from using IV estimation, particularly for experimental analysis. When properly used and interpreted, IV estimation is a valuable tool, as it allows researchers to overcome common problems of selection bias and unobserved heterogeneity to consistently estimate an internally valid causal quantity for a subset of the population. But it is nevertheless important for researchers to recognize the distinction between the LATE and the ATE. Interpretation of the LATE as a causal quantity of interest is a concern for any researcher who contends with noncompliance, but one that is rarely addressed in applied work in political science. In fact, in a review of thirty-four empirical articles that employ IV estimation from 2004 to 2009 in the *American Political Science Review* and the *American Journal of Political Science*, only two (6%) mention that the causal effect being estimated is the LATE.

So what should applied researchers do when a valid instrument is available and the ATE is the causal quantity of interest? This article articulates a set of assumptions sufficient for moving beyond the LATE to the ATE, and proposes a simple and intuitive reweighting method for operationalizing these assumptions. The method leverages the random assignment of the instrument to achieve a consistent estimator of the ATE for compliers, and then reweights the population so that the compliers have a covariate distribution that matches that of the full population. Consistency for the ATE will follow from nonexperimental assumptions (e.g., ignorability); however, unlike purely observational methods, consistency for *an* average causal effect—for some subpopulation—is preserved under misspecification. The logic of the method sheds light on the connection between the LATE and ATE and the steps necessary to move from the former to the latter.

The article proceeds as follows. We begin by reviewing the assumptions and limitations of IV estimation, first from a structural equation modeling (SEM) perspective and then using the more flexible potential outcomes interpretation (Angrist, Imbens, and Rubin 1996). We next explain our method for estimating the ATE. We provide a maximum likelihood (ML) estimator to estimate each unit's compliance score, or conditional probability of being a complier given covariates. We show researchers how to use the estimated compliance scores to reweight the sample using a method that we refer to as inverse compliance score weighting (ICSW), which is analogous to inverse probability weighting (IPW) for sample correction. We propose sufficient conditions for consistency and detail the consequences of assumption failure. We then apply the method to two published papers in political science and find the potential for substantively meaningful differences between the LATE and ATE.

2 IV Estimation

Political scientists frequently contend with problems of endogeneity, which occur when a variable of interest is systematically related to unobserved causes of the outcome variable. Endogeneity can arise for a variety of reasons, including selection bias, simultaneity, or measurement error. A popular method for dealing with endogeneity is IV estimation, which can be understood through two frameworks: the SEM framework and the potential outcomes framework. The SEM framework is most familiar to political scientists, and so we begin with it. However, IV estimation in the context of this framework introduces the rigid—and perhaps implausible—assumption of constant treatment effects. The potential outcomes model relaxes this assumption, but also changes the causal estimand.

2.1 *IV in an SEM Framework*

The SEM interpretation of IV estimation posits a linear, additive relationship between an outcome variable (Y_i), an endogenous treatment variable (D_i), covariates ($Q_{1i}, Q_{2i}....Q_{Ki}$), and an unobserved error term (u_i) for each unit i, such that

$$Y_i = \beta_0 + \beta_1 D_i + \lambda_1 Q_{1i} + \lambda_2 Q_{2i} + ... + \lambda_K Q_{Ki} + u_i. \quad (1)$$

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

We are able to consistently estimate β_1 using ordinary least squares (OLS) regression when D_i is asymptotically uncorrelated with the error term ($\frac{1}{n}\Sigma D_i u_i \rightarrow_p 0$). This requirement is violated, however, when D_i is endogenous.

A solution to the endogeneity problem is to find an instrument Z_i that is only predictive of Y_i through its impact on D_i. In other words, Z_i affects D_i, but is excluded from equation (1). We can then use IV regression, which is based on a two-equation model. The first equation, or the "first stage," is a regression of the endogenous variable on the instrument and covariates:

$$D_i = \gamma_0 + \gamma_1 Z_i + \delta_1 Q_{1i} + \delta_2 Q_{2i} + ... + \delta_K Q_{Ki} + e_i, \quad (2)$$

where e_i is a second unobserved error term. The coefficients from this first-stage regression generate predicted values of D_i, purging D_i of endogeneity. We can substitute the predicted values of D_i back into equation (1), regressing Y_i on the predicted values of D_i as well as the covariates. This two-stage substitution process is what gives the regression the name two-stage least squares (2SLS). Given the model in equations (1) and (2), the SEM IV framework relies on two main assumptions for consistency. First, the correlation between Z_i and u_i must approach zero as n grows. Second, the covariance between Z_i and D_i must converge to a nonzero quantity as n grows.

Note, however, that whereas the endogenous variable and covariates in equation (1) are indexed by the subscript i, their coefficients (the βs and λs) are not. In other words, the SEM framework assumes constant treatment effects, so that the effect of the treatment is identical for each unit i. The constant effects assumption is strong and may not always be justifiable in social science applications. For example, consider the implications of the constant effects assumption in the context of an experiment designed to assess the effects of negative campaign advertising on political attitudes (Arceneaux and Nickerson 2010). Constant effects implies that a negative advertisement affects attitudes for all subjects, from those who have strong prior political opinions to those who are indifferent, by exactly the same amount.[1] We can avoid making the constant treatment effects assumption using the potential outcomes framework developed by Angrist, Imbens, and Rubin (1996), which shows that the IV estimator will identify the LATE even when treatment effects are heterogeneous. In the following section, we introduce the potential outcomes notation and discuss the implications of identifying the LATE.

2.2 *IV in a Potential Outcomes Framework*

We now move to the Neyman-Rubin potential outcomes framework, focusing on the case of a binary IV (or treatment assignment) Z and a binary endogenous treatment variable (or treatment received) D. Suppose that we have a sample consisting of n units (indexed $i = 1, 2, \ldots, n$) independently drawn with equal probability from some large population of interest. For unit i, let Y_{0i} be the outcome if untreated, $D_i = 0$, and let Y_{1i} be the outcome if treated, $D_i = 1$. The treatment effect for a given unit i is the difference between unit i's outcomes in both possible states of the world, $Y_{1i} - Y_{0i}$. Similarly, we may define D_{0i} as the treatment condition of unit i when assigned to control, $Z_i = 0$, and D_{1i} as the treatment condition of unit i when assigned to treatment, $Z_i = 1$. We will show that the causal estimand recovered by IV estimation will not generally be the ATE, $\mathrm{E}(Y_1 - Y_0)$, but will instead be the LATE, $\mathrm{E}(Y_1 - Y_0 | D_1 > D_0)$.

Following Angrist, Imbens, and Rubin (1996), the population may be divided into four groups: always-takers, never-takers, compliers, and defiers. Always-takers are units that receive treatment regardless of whether they are assigned to treatment or control, so that $D_{0i} = 1$ and $D_{1i} = 1$. Conversely, never-takers do not receive treatment regardless of their treatment assignment, so that $D_{0i} = 0$ and $D_{1i} = 0$. Compliers receive treatment if assigned to treatment and do not receive treatment if assigned to control, so that $D_{0i} = 0$ and $D_{1i} = 1$ (or $D_{1i} > D_{0i}$). Defiers

[1] Other examples include: Hidalgo et al. (2010) show that the effects of economic shocks on conflict are larger in regions with high levels of preexisting inequality; Altonji and Dunn (1996) demonstrate that the returns to education are larger for students in high-quality schools; Arceneaux and Nickerson (2009) find that the mobilizing effects of canvassing are larger for high-propensity voters.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

receive treatment if assigned to control and do not receive treatment if assigned to treatment, so that $D_{0i} = 1$ and $D_{1i} = 0$ (or $D_{0i} > D_{1i}$). In the example of a randomized clinical trial to assess the causal effect of taking a medication, always-takers will take the medication regardless of which group they are assigned to, never-takers will not take the medication regardless of which group they are assigned to, compliers will take the medication if and only if assigned to the treatment group, and defiers will take the medication if and only if assigned to the control group.

Using an IV estimator, we may estimate the LATE using five assumptions, as explicated by Angrist, Imbens, and Rubin (1996). First, we assume that the exclusion restriction is valid, or that Z_i only affects Y_i through D_i. Second, we assume that Z_i to some degree predicts D_i. Third, we invoke the stable unit treatment value assumption, which states that the potential outcomes D_{0i}, D_{1i}, Y_{0i}, and Y_{1i} are fixed and unrelated to the particular arrangement of assignment and treatment statuses. Fourth, we assume that the population contains no defiers, i.e., $\Pr(D_0 > D_1) = 0$. Fifth, we assume that Z_i is randomly assigned and thus independent of all potential outcomes. As no parametric assumptions are necessary, these assumptions are considerably weaker than (and subsumed by) those of the SEM framework.

Angrist, Imbens, and Rubin (1996) demonstrate that these five assumptions imply that the "intention-to-treat" effect of Z_i on Y_i (the average effect of treatment assignment on the outcome) divided by the intention-to-treat (ITT) effect of Z on D (the average effect of treatment assignment on treatment received) is equal to the average causal treatment effect for compliers:

$$\mathrm{E}(Y_1 - Y_0 | D_1 > D_0) = \frac{\mathrm{E}(Y|Z=1) - \mathrm{E}(Y|Z=0)}{\mathrm{E}(D|Z=1) - \mathrm{E}(D|Z=0)}. \tag{3}$$

The LATE, $\mathrm{E}(Y_1 - Y_0 | D_1 > D_0)$, can therefore be estimated with

$$\hat{\tau}_{\mathrm{LATE}} = \frac{\left(\sum_{i=1}^{n} Z_i Y_i\right)\Big/\left(\sum_{i=1}^{n} Z_i\right) - \left(\sum_{i=1}^{n} (1 - Z_i) Y_i\right)\Big/\left(\sum_{i=1}^{n} (1 - Z_i)\right)}{\left(\sum_{i=1}^{n} Z_i D_i\right)\Big/\left(\sum_{i=1}^{n} Z_i\right) - \left(\sum_{i=1}^{n} (1 - Z_i) D_i\right)\Big/\left(\sum_{i=1}^{n} (1 - Z_i)\right)} \tag{4}$$

or, equivalently, the bivariate 2SLS estimator. This estimator is consistent but not unbiased; it is subject to finite sample bias because it is a ratio estimator. When Z_i is randomly assigned (and therefore unrelated to covariates), the addition of covariates to the 2SLS model will also yield consistent estimation of the LATE (Angrist and Pischke 2009).[2] Intuitively, 2SLS rescales the ITT effect of Z_i on Y_i, acknowledging that only compliers will respond to changes in Z_i.

Because IV estimation provides asymptotically unbiased estimates of the LATE even when treatment effects are heterogeneous, IV estimation has become the de facto standard for causal inference in studies with noncompliance in the social sciences. However, as discussed in Section 1, researchers are often interested in the ATE rather than the LATE. In the next section, we present a method that allows for estimation of the ATE, albeit with a stronger set of assumptions for consistency.

3 From LATE to ATE

Although the LATE is a well-defined estimand, it is nevertheless only applicable to a particular subpopulation: compliers. To address this problem, we provide a method to estimate the ATE using the intuition behind IPW, a commonly used method to estimate population-level quantities from nonrepresentative samples. IPW entails weighting units by the inverse of the probability of sample inclusion (Horvitz and Thompson 1952) to asymptotically recover the population quantities. ICSW follows the same logic, but weights units by the inverse of the compliance

[2]We provide a simple intuition for why, under random assignment of Z_i, 2SLS with covariates is consistent for the LATE. The least squares regression of Y_i on Z_i and covariates is consistent for the numerator of equation (3), and the least squares regression of D_i on Z_i and covariates is consistent for the denominator of equation (3). The 2SLS estimator of the effect of D_i will be the ratio of the two least squares estimators and, by Slutsky's theorem, will be consistent for the LATE.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

score, or the conditional probability of being a complier. We next introduce the compliance score and present a new ML estimator to estimate compliance scores even in the presence of two-sided noncompliance (wherein both always-takers and never-takers are present).

3.1 *The Compliance Score*

The compliance score is a latent pretreatment covariate that represents the probability of being a complier, conditional on a unit's pretreatment covariate profile $\mathbf{X}$ (so that $\mathbf{X} \perp\!\!\!\perp Z$). The compliance score for unit i, $P_{Ci} = \Pr(D_1 > D_0 | \mathbf{X} = \mathbf{x}_i)$, where $\mathbf{x}_i$ is the covariate vector for unit i. Although we cannot directly observe each unit's compliance score, each unit's score can be estimated from the data. Consider a randomized experiment in which only one covariate, gender, is available. If we estimate that 75% of the men in the sample are compliers (by estimating the ITT effect of Z_i on D_i for men), we would estimate that $P_{Ci} = 0.75$ for all men. Indeed, when there is a small number of unique covariate profiles (e.g., a small number of binary covariates), we could consistently estimate P_{Ci} for each unique covariate profile simply by estimating the conditional ITT effects of Z_i on D_i. However, when multiple and/or continuous covariates are available, it is useful to put additional structure on the problem to permit estimation.

Given the assumptions outlined in the previous section, we know that the compliance score, as a pretreatment covariate, is independent of Z. The compliance score for a given unit is thus simple to estimate under one-sided noncompliance (when no units assigned to control will ever receive treatment), as $\mathrm{E}(D_1 | Z = 1, \mathbf{X} = \mathbf{x}) - \mathrm{E}(D_0 | Z = 0, \mathbf{X} = \mathbf{x}) = \mathrm{E}(D | Z = 1, \mathbf{X} = \mathbf{x})$. When there are no always-takers or defiers, all units assigned to treatment who take the treatment are compliers. Therefore, for all units in the treatment group, we could run a binomial regression of D on $\mathbf{X}$ to estimate the predicted probabilities of compliance.

We present a new ML estimation technique (similar to that of Yau and Little 2001) that generalizes existing methods of compliance score estimation to the case of two-sided noncompliance. We have not seen an estimator of this form in the literature, as we propose a nested structure for compliance that nevertheless reduces to a binomial ML under one-sided noncompliance and strata means under saturation. For examples of the use of the compliance score under one-sided noncompliance in biostatistics, see Follmann (2000), Joffe and Brensinger (2003), Joffe, Ten Have, and Brensinger (2003), and Roy, Hogan, and Marcus (2008). For Bayesian estimators of the compliance score under two-sided noncompliance, see Imbens and Rubin (1997) and Hirano et al. (2000); however, these methods require specification of a parametric model for the joint distribution of treatment assignment, compliance, and outcomes. In contrast, our method requires only specification of a parametric model for the first stage, and nonparametric extensions are clearly possible via nonparametric ML estimation.

The compliance score may be directly estimated using ML estimation. To begin, we construct a likelihood function. For convenience, we make three easily relaxed parametric assumptions. The first assumption is that the probability of being an always-taker or a complier is a function of covariates with a known distribution:

$$P_{A,C,i} = \Pr(D_1 > D_0 \cup D_0 = 1 | \mathbf{X} = \mathbf{x}_i) = F(\theta_{\mathbf{A},\mathbf{C}} \mathbf{x_i}), \tag{5}$$

where $P_{A,C,i}$ is the conditional probability that unit i is either a complier or an always-taker, $\theta_{\mathbf{A},\mathbf{C}}$ is a vector of coefficients to be estimated, and $F(\cdot)$ is the cumulative distribution function (CDF) for an arbitrary distribution. For the purposes of this article, we use a probit model, so $F(\cdot) = \Phi(\cdot)$, where Φ is the Normal CDF. Second, we similarly specify

$$P_{A|A,C,i} = \Pr(D_0 = 1 | D_1 > D_0 \cup D_0 = 1, \mathbf{X} = \mathbf{x}_i) = F(\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}} \mathbf{x_i}), \tag{6}$$

where $P_{A|A,C,i}$ is the conditional probability that unit i is an always-taker conditional on it being either an always-taker or a complier, and $\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}$ are coefficients to be estimated. Therefore, we may define the compliance score as $P_{C,i} = \Pr(D_{1i} > D_{0i})$. Since we know, by definition, that compliers receive treatment if and only if assigned to treatment and that always-takers always receive treatment,

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

$$\begin{aligned}\Pr(D = 1|\mathbf{X} = \mathbf{x}_i) &= \Pr(D_1 > D_0|\mathbf{X} = \mathbf{x}_i)Z_i + \Pr(D_0 = 1|\mathbf{X} = \mathbf{x}_i) \\ &= P_{A,C,i}(1 - P_{A|A,C,i})Z_i + P_{A,C,i}P_{A|A,C,i}.\end{aligned}$$

We thus have a fully specified model for $\Pr(D = 1|\mathbf{X} = \mathbf{x}_i)$, and this value is strictly bounded within (0,1) since $Z_i \in \{0, 1\}$ and $P_{A,C,i} \in (0, 1)$. This boundedness, along with the binary nature of Z_i, allows us to specify our third assumption: D_i is Bernoulli distributed. Then the likelihood for any unit i,

$$\begin{aligned}L(P_{A|A,C,i}, P_{A,C,i}|D, Z) = (P_{A,C,i}(1 - P_{A|A,C,i})Z_i + P_{A,C,i}P_{A|A,C,i})^{D_i} \\ (1 - P_{A,C,i}(1 - P_{A|A,C,i})Z_i - P_{A,C,i}P_{A|A,C,i})^{1-D_i}.\end{aligned} \tag{7}$$

Combining equations (5), (6), and (7), $L(\theta_{\mathbf{A},\mathbf{C}}, \theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}|D, Z) =$

$$\begin{aligned}\Pi_{i=1}^{n}((F(\theta_{\mathbf{A},\mathbf{C}}\mathbf{x_i})(1 - F(\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}\mathbf{x_i}))Z_i + F(\theta_{\mathbf{A},\mathbf{C}}\mathbf{x_i})F(\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}\mathbf{x_i}))^{D_i} \\ (1 - F(\theta_{\mathbf{A},\mathbf{C}}\mathbf{x_i})(1 - F(\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}\mathbf{x_i}))Z_i - F(\theta_{\mathbf{A},\mathbf{C}}\mathbf{x_i})F(\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}\mathbf{x_i}))^{1-D_i}).\end{aligned}$$

Thus the ML estimate of $\{\theta_{\mathbf{A},\mathbf{C}}, \theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}\}$,

$$\{\hat{\theta}_{\mathbf{A},\mathbf{C}}, \hat{\theta}_{\mathbf{A}|\mathbf{A},\mathbf{C}}\} = \underset{\theta_{\mathbf{A},\mathbf{C}}, \theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}}{\operatorname{argmax}}\big[L(\theta_{\mathbf{A},\mathbf{C}}, \theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}|D, Z)\big].$$

After estimating $\theta_{\mathbf{A},\mathbf{C}}$ and $\theta_{\mathbf{A}|\mathbf{A},\mathbf{C}}$, we may estimate the compliance score for unit i,

$$\hat{P}_{Ci} = \widehat{\Pr}(D_1 > D_0|\mathbf{X} = \mathbf{x}_i) = F(\hat{\theta}_{\mathbf{A},\mathbf{C}}\mathbf{x_i})(1 - F(\hat{\theta}_{\mathbf{A}|\mathbf{A},\mathbf{C}}\mathbf{x_i}))^3.$$

Although the ML estimator imposes parametric assumptions about the relationship between the covariates and the compliance score, these assumptions may be relaxed to arbitrary generality (given data constraints) by including interactions in and nonlinear transformations (e.g., polynomial expansion, splines) of $\mathbf{X}$. A natural extension would be a sieve-type estimator (Geman and Hwang 1982), which would facilitate nonparametric estimation of the compliance score.

3.2 *ICSW*

How does the compliance score help solve the problem of reliance on the LATE even when treatment effects are heterogeneous? Recall that the basic problem addressed by this article is that IV estimation only generates estimates of the ATE for compliers. We propose a reweighting method that we refer to as ICSW.[4] ICSW follows the logic of IPW: if one type of unit is disproportionately represented, we may reweight the sample to reflect the distribution in the target population.

We provide a simple example to illustrate how the method works. Imagine an experiment where gender is the only determinant of compliance, such that 75% of males comply with their treatment assignments but only 10% of females comply with their treatment assignments. Suppose that the treatment effect for males is 0, but the treatment effect for females is 1. If males and females are

[3] As suggested above, in the case of one-sided noncompliance, such that the probability of being an always-taker reduces to zero, the likelihood function reduces to

$$L(\theta_{\mathbf{C}}|D, Z) = \Pi_{i=1}^{n}[(F(\theta_{\mathbf{A},\mathbf{C}}\mathbf{x_i})Z_i)^{D_i} + (1 - F(\theta_{\mathbf{A},\mathbf{C}}\mathbf{x_i})Z_i)^{1-D_i}].$$

This likelihood function is a familiar one: it is the standard likelihood function for binomial regression, if only applied to units in the treatment condition. With one-sided noncompliance and $F(\cdot) = \Phi(\cdot)$, our ML estimator therefore reduces to the probit estimator as applied only to units with $Z_i = 1$.

[4] In work developed contemporaneously with our article, Angrist and Fernandez-Val (2013) present a similar method to reweight covariate-specific LATEs to target populations. Although conceptually similar, the estimation procedures in Angrist and Fernandez-Val (2013) require discrete covariates that take on a finite and fixed number of values. It can also be shown that when ICSW is used with a saturated set of covariates, it reproduces Angrist and Fernandez-Val's (2013) estimator. Angrist and Fernandez-Val (2013) also derive results on analogues to overidentification under the potential outcomes framework given multiple instruments. In addition, for an integrative framework for recovering the ATE along with other causal quantities in the case of one-sided noncompliance, see Esterling, Neblo, and Lazer (2011). Frangakis and Rubin (1999) present a similar, foundational approach involving missing data.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

represented equally in the population, then the ATE is $\frac{0.5\times0+0.5\times1}{0.5+0.5} = 0.5$. The problem, of course, is that we cannot directly estimate the ATE due to noncompliance. IV estimation recovers the LATE, which we compute using a weighted average, $\frac{0.5\times0.75\times0+0.5\times0.10\times1}{0.5\times0.75+0.5\times0.10} = 0.12$. In this example, the LATE is much lower than the ATE because females are underrepresented among compliers. Recovering the ATE with noncompliance entails weighting units by the inverse of the compliance score, then performing IV estimation. Again computing a weighted average, the estimand recovered by this procedure is the ATE: $\frac{0.5\times0.75\times\frac{1}{0.75}\times0+0.5\times0.10\times\frac{1}{0.10}\times1}{0.5\times0.75\times\frac{1}{0.75}+0.5\times0.10\times\frac{1}{0.10}} = \frac{0.5\times0+0.5\times1}{0.5+0.5} = 0.5$. We develop this intuition formally, first providing a general expression for the estimator, then detailing the assumptions necessary for consistency.

3.3 *ICSW Estimator*

Following from our use of the potential outcomes framework, the estimation procedure presented here requires the instrument Z_i and endogenous variable D_i to be binary. This requirement is satisfied in many empirical studies, including numerous randomized experiments with noncompliance.[5] In the bivariate case (when there are no covariates), the estimator is simple. Define $\hat{w}_{Ci} = 1/\hat{P}_{Ci}$. The ICSW estimator $\hat{\tau}_{\text{ATE}} =$

$$\frac{(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i Y_i)/(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i) - (\sum_{i=1}^{n} \hat{w}_{Ci}(1-Z_i) Y_i)/(\sum_{i=1}^{n} \hat{w}_{Ci}(1-Z_i))}{(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i D_i)/(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i) - (\sum_{i=1}^{n} \hat{w}_{Ci}(1-Z_i) D_i)/(\sum_{i=1}^{n} \hat{w}_{Ci}(1-Z_i))}.$$

An equivalent estimator may be derived using weighted 2SLS (with weights $\hat{w}_{Ci}$). As with the simple 2SLS case, covariates may also be included to reduce sampling variability without any consequence for the asymptotic bias of the estimator.

Like other estimators that rely on reweighting, this estimator is typically more variable than its unweighted counterpart, 2SLS, and is potentially subject to greater finite sample bias. The ICSW estimator, like 2SLS, has increased finite sample bias when compliance rates are low, and this problem is exaggerated when some units have very low compliance scores.[6] To characterize uncertainty, we recommend bootstrapping the entire process—from computing the compliance score to weighted IV estimation (Abadie 2002).[7]

3.4 *ICSW Assumptions*

In order to consistently estimate the ATE using ICSW, three additional assumptions are required. In the appendix, we demonstrate that ICSW is consistent under Assumptions 1–3, along with the IV assumptions and suitable regularity conditions.

[5]The method can be generalized to applications using continuous endogenous variables and multivalued (or multiple) instruments. As Angrist and Imbens (1995) demonstrate, with continuous endogenous variables, 2SLS produces a weighted average of grouped data IV estimators. With multiple instruments, 2SLS is a weighted average (efficient under homoskedasticity) of the 2SLS estimator for each instrument. Although detailing such extensions is beyond the scope of the article, under the assumptions outlined here, the ATE could be estimated for each instrument and a weighted average could be produced for the overall ATE.

[6]In order to prevent very small compliance scores from forming in finite samples, we Winsorize our estimates. Define c as the $1/n^{\alpha}$th quantile of $\hat{P}_{Ci}$, where α is a positive calibrating constant. We replace all $\hat{P}_{Ci} < c$ with c. We recommend (and use in this article) $\alpha = 0.275$, which performed well in simulation studies presented in the supplementary appendix. Winsorizing is performed because very low probabilities can introduce instability in the estimates resulting from IPW (Elliott 2009), but Winsorizing may sometimes increase finite sample bias. Asymptotically, assuming that α remains fixed, the Winsorization has no impact on estimation, because $c \to 0$ as $n \to \infty$. In fact, non-Winsorized estimates represent a special case of this Winsorization process whenever $\alpha \to \infty$.

[7]For IPW-style estimators, analytic variance estimators presuming correct model specification may be highly sensitive to model misspecification and inaccurate in finite samples, whereas the bootstrap tends to have better coverage (Funk et al. 2011). In our own simulation studies, assuming $\alpha = 0.275$, we find that the bootstrap has conservative coverage for ICSW in small samples, and proper coverage in larger samples. These studies are presented in the supplementary appendix.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

Assumption 1.
Latent ignorability of compliance with respect to treatment effect heterogeneity (Frangakis and Rubin 1999; Esterling, Neblo, and Lazer 2011) holds. Assumption 1 entails that, for all $\mathbf{x} \in \text{Supp}(\mathbf{X})$, $\text{E}(Y_1 - Y_0|\mathbf{X} = \mathbf{x}) = \text{E}(Y_1 - Y_0|D_1 > D_0, \mathbf{X} = \mathbf{x})$: the expected value of the treatment effect conditional on a covariate profile is equal to the expected value of the treatment effect conditional on a covariate profile and compliance. Put simply, the ATE for all units with a covariate profile $\mathbf{x}$ is equal to the ATE for compliers with a covariate profile $\mathbf{x}$.

Ignorability assumptions, while strong, are necessary for consistent estimation of causal effects under a variety of common methods, including propensity score methods, matching, and even linear regression. However, Assumption 1 may be distinguished from the typical ignorability assumption (Rubin 1978), which requires that both $\text{E}(Y_1|D = 1, \mathbf{X} = \mathbf{x}) = \text{E}(Y_1|D = 0, \mathbf{X} = \mathbf{x})$ and $\text{E}(Y_0|D = 1, \mathbf{X} = \mathbf{x}) = \text{E}(Y_0|D = 0, \mathbf{X} = \mathbf{x})$. Assumption 1 is less restrictive than the typical ignorability assumption as there is no need to account for baseline selection bias, only the determinants of treatment effect heterogeneity. For example, consider a randomized clinical trial of an experimental medicine. Assume the medicine may affect subjects differently depending only on certain known underlying medical conditions. The researcher might not be able to determine or measure all of the causes of take-up that are related to potential outcomes (e.g., wealth or risk-taking), so the typical ignorability assumption would not be satisfied. However, Assumption 1 might still hold; only the determinants of both compliance and heterogeneity need to be included in the conditioning set for consistency.

We can describe sufficient conditions for and consequences of Assumption 1. If all predictors of compliance are known, then Assumption 1 will clearly hold. But what if only some (or even none) of the predictors of compliance are observed? In the case where the treatment effect is a constant value, τ, across all units, Assumption 1 *always* holds: $\text{E}(Y_1 - Y_0|\mathbf{X} = \mathbf{x}) = \text{E}(Y_1 - Y_0|D_1 > D_0, \mathbf{X} = \mathbf{x}) = \tau$. If we were to accept the SEM interpretation of IV, where treatment effects are constant, Assumption 1 would always be satisfied and both 2SLS and ICSW would be consistent for τ. More generally, even if treatment effects are heterogeneous, but still statistically independent of compliance, Assumption 1 would necessarily be satisfied, and both 2SLS and ICSW would be consistent for the ATE.

But what if treatment effects are heterogeneous and we have failed to identify the sufficient conditioning set? As Assumption 1 no longer holds, the ICSW estimator is now biased, even asymptotically. However, ICSW will still recover an internally valid causal effect, not on the full population of interest, but for some population that has a covariate profile that matches that of the full population. To see this, we can compare two expressions. Through simple manipulation of expectations, the LATE may be written as

$$\begin{aligned}&\int_{\Omega} \text{E}(Y_1 - Y_0|D_1 > D_0,\ \mathbf{X} = \mathbf{x}) \frac{\Pr(D_1 > D_0|\mathbf{X} = \mathbf{x}) f(\mathbf{x}) d\mathbf{x}}{\Pr(D_1 > D_0)} \\ &= \int_{\Omega} \text{E}(Y_1 - Y_0|D_1 > D_0,\ \mathbf{X} = \mathbf{x}) g(\mathbf{x}) d\mathbf{x},\end{aligned}$$

where $f(.)$ is the population pdf of $\mathbf{X}$, $g(.)$ is the subpopulation pdf of $\mathbf{X}$ among compliers, and $\Omega = \text{Supp}(\mathbf{X})$. Manipulating equations (10) and (13) from the appendix, the asymptotic value of the ICSW estimator under failure of Assumption 1 will be

$$\int_{\Omega} \text{E}(Y_1 - Y_0|D_1 > D_0,\ \mathbf{X} = \mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

The difference between $g(.)$ and $f(.)$ characterizes the nature of the adjustment performed by ICSW. Much like we would reweight a poll to attempt to ensure that the sample is representative in its observable covariates, ICSW reweights the subpopulation upon which effects are estimated so that it is representative in its observables.

When the ATE is the parameter of interest, the limiting value of the ICSW estimator is thus typically more sensible than that of 2SLS/IV, as ICSW incorporates available covariate

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

information. In fact, 2SLS represents the special case of ICSW where no covariates are included in the compliance score estimation process. Although there are examples where adding covariates can increase asymptotic bias under misspecification (due to complementary imbalances), we concur with Rubin (2009), who argues that the most principled conditioning strategy is one that includes all known pretreatment covariates.[8]

Assumption 2.
Consistency of $\hat{P}_{Ci}$ for P_{Ci}. Assumption 2 amounts to requiring proper specification of the ML estimator. Application of a flexible parametrization of the covariates in $\mathbf{X}$, as is possible with large samples, will necessarily yield consistency, so long as the rate of growth in the number of estimated parameters is sufficiently slow in n. In the case of a saturated model and $F(.)$ as the normal CDF, $\hat{P}_{Ci}$ will be unbiased for the conditional probability of compliance.

Assumption 3.
For all $\mathbf{x} \in \text{Supp}(\mathbf{X})$, $\Pr(D_1 > D_0 | \mathbf{X} = \mathbf{x}) \in (\epsilon, 1]$, where $\epsilon > 0$. Note that IV estimation already requires the assumption of nonzero compliance in the full population, and we generalize this assumption to every covariate profile. This assumption has both theoretical and practical implications. If Assumption 3 fails, then the ATE cannot generally be estimated using the variation induced by Z: the instrument simply does not cause any variation in the treatment received by some types of units. In practice, if a unit has zero probability of compliance, the weighting procedure will produce infinite weights asymptotically, leading to an undefined estimate. Conditional on the validity of Assumptions 1 and 2, Assumption 3 is therefore verifiable in large samples.

4 Applications

We now present two case studies: Green, Gerber, and Nickerson (2003) and Albertson and Lawrence (2009). We select these two articles because they allow us to display the use of our method in the intuitive case of one-sided noncompliance as well as the more complex case of two-sided noncompliance.[9]

4.1 *Green, Gerber, and Nickerson (2003)*

Green, Gerber, and Nickerson (2003) present results from a large-scale ($n = 18,933$) field experiment designed to assess the effect of canvassing on voter turnout. Registered voters in Bridgeport, Columbus, Detroit, Minneapolis, Raleigh, and St. Paul were randomly assigned to treatment or control, where the treatment was encouragement to vote in advance of the November 6, 2001, local elections. Encouragement was delivered in the form of face-to-face contact with nonpartisan student and community organization members. Because some citizens who were assigned to the treatment group could not be contacted, the study features one-sided noncompliance. To get around this limitation, the authors use the attempt to administer the treatment as an instrument for the receipt of the treatment to consistently estimate the LATE. These results are presented in column (1) of Table 1. Among compliers, canvassing is estimated to increase turnout by 5.6 percentage points with a confidence interval of $(2.2, 8.9)$.[10]

[8] Data constraints (e.g., a large number of covariates or small n) may prevent inclusion of all covariates, in which case some dimension reduction or variable selection may be necessary. Fodor (2002) provides a review of such methods.
[9] Replication data for these applications are available at Aronow and Carnegie (2013).
[10] The estimates presented differ slightly from those in the original paper because the original paper used one-way fixed effects for randomization strata, which is inconsistent in the presence of heterogeneous probabilities of assignment to treatment. In general, when using a covariate conditioning strategy, linear regression and IV estimators produce estimates reweighted by the conditional variance of the instrument (Angrist and Pischke 2009; Humphreys 2009). We instead use IPW, which is consistent in this example.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

Table 1 2SLS and ICSW estimates for Green, Gerber, and Nickerson (2003)

Voter turnout in 2001 local elections	*LATE (2SLS)* *(1)*	*ATE (ICSW)* *(2)*
Campaign contact	**5.6**	**6.7**
	(2.2, 8.9)	(3.2, 10.3)
Number of family members	−0.2	0.0
	(−0.8, 0.4)	(−0.6, 0.6)
White	9.0	9.5
	(3.9, 13.7)	(4.5, 13.9)
Black	−1.7	−1.1
	(−7.7, 3.5)	(−6.7, 3.7)
Voted in 2000 general election	23.0	21.0
	(21.9, 24.1)	(19.8, 22.2)
Voted in 1999 general election	30.5	27.2
	(28.5, 32.4)	(25.1, 29.4)
Voted in 2001 primary election	48.7	51.0
	(47.0, 50.5)	(49.2, 52.9)
Age	0.2	0.2
	(0.2, 0.2)	(0.1, 0.2)
Age missing	12.7	17.7
	(−16.5, 44.1)	(−15.9, 50.3)
Democrat	2.6	2.6
	(0.6, 4.7)	(0.7, 04.8)
Republican	3.8	3.2
	(1.3, 6.2)	(0.8, 5.8)
Independent	8.2	6.3
	(5.4, 10.7)	(3.8, 8.8)
City controls	Yes	Yes

Note. Bootstrap 95% confidence intervals in parentheses. Dependent variable is voter turnout in the 2001 general election in percentage points. Covariates include dummy variables for turnout in the 1999 general, 2000 general, and 2001 primary elections, whether the respondent is a Democrat, is a Republican, is an Independent, is black, is white, city of residence, and missing data on age; and integers for the respondent's age and the number of family members in the household

The theoretical implications of the LATE are somewhat unclear in this experiment. Which subjects are compliers is an artifact of the way in which the intervention is deployed. For example: if canvassers attempted to contact subjects multiple times, the contact rate would surely have been higher. There is little theoretical reason to care about subjects who could be contacted in the context of this particular experiment. This shortcoming of the LATE is compounded by the fact that the authors are responding to a broader literature on the determinants of voter mobilization, citing works such as Rosenstone and Hansen (1993), Verba, Schlozman, and Brady (1995), and Putnam (2000). These earlier works focus on the determinants of turnout in the entire population, whereas Green, Gerber, and Nickerson (2003) estimate the determinants of turnout in the complier population, the answer to a potentially different question. Green, Gerber, and Nickerson (2003, 1086) state that a central goal of their study is "to better gauge the average treatment effect of canvassing," but the LATE provides limited information about the ATE. For example, noncompliers may have different education levels, incomes, interest in politics, or other demographic characteristics than compliers, leading them to respond differently to contact.

We begin by estimating compliance scores for the sample. As recommended above, we use all observed pretreatment covariates, which include the covariates listed in Table 1 and dummy variables for the six cities in the sample. Although the set of available covariates may or may not satisfy the ignorability assumption, we will still recover an estimate for a population that has a covariate profile that matches that of the full population. We see that compliers are more likely to have voted previously, be members of a major party, and identify as white (see the supplementary appendix for a correlation matrix between the estimated compliance score and covariates). We now apply ICSW

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

and present the results in column (2) of Table 1. Canvassing is estimated to increase turnout in the entire population by 6.7 percentage points with a confidence interval of (3.2, 10.3), whereas it is estimated to increase turnout among compliers by only 5.6 percentage points. The ATE is therefore estimated to be 19% larger than the LATE.[11] The estimated ATE strengthens Green, Gerber, and Nickerson's (2003, 1094) conclusion that "mobilization campaigns have the potential to increase turnout substantially in local elections," and addresses a population that more closely resembles the population of interest.

4.2 *Albertson and Lawrence (2009)*

Albertson and Lawrence (2009) present findings from an experiment ($n = 507$) in which a representative sample of survey respondents in Orange County, California, were randomly assigned to receive encouragement to view a Fox debate on affirmative action, which would take place on the eve of the 1996 presidential election. Shortly after the election, these respondents were reinterviewed. The postelection questionnaire asked respondents whether they viewed the debate, whether they supported a California proposition (209) to eliminate affirmative action (coded 1 if they supported it and 0 if not), and how informed they felt about the proposition (coded on a scale from 1 to 4 from least to most informed). The authors use a standard IV design to address the fact that some who were not assigned to treatment reported viewing the debate and some who were assigned to treatment did not report viewing the debate. This two-sided noncompliance was nontrivial: 55% of subjects assigned to watch the debate did not report viewing the debate, and 4% of subjects who were not assigned to watch the debate reported viewing it. Albertson and Lawrence's IV regression results show a positive relationship between program viewing and feeling more informed about the issue and a statistically insignificant, negative relationship between program viewing and support for the proposition among compliers. Albertson and Lawrence's original findings are presented in columns (1) and (3) of Table 2.[12]

Is the ATE or the LATE a more appropriate estimand in this case? Our estimand of interest depends on the research question. Albertson and Lawrence (2009, 276) state that the question they seek to answer is whether "civic-minded television ha[s] a lasting impact on those who watch," which responds to an existing literature that "supports the view that television can be expected to inform viewers, make issues more salient, change viewers' attitudes and possibly even affect their behavior." Thus, the population of interest is defined to be viewers, but leaves ambiguous which viewers are of interest. Given the pains taken to achieve a representative sample for the original survey, we may believe that the parameter of theoretical interest is the ATE because it addresses the potential media effects on the population as a whole. Although the ignorability assumption may not hold, ICSW permits estimation of causal effect for a population that, at a minimum, resembles the full population on observables, thus likely better approximating the ATE. Divergence between our estimates of the ATE and LATE implies that effect heterogeneity has the potential to complicate interpretation of the effects of viewership.

We first estimate compliance scores for the sample using the eight covariates used by Albertson and Lawrence (see Table 2 for a description of each of the covariates). The compliance score most strongly correlates with political interest, newspaper reading, and education (see the supplementary appendix for a full correlation matrix). Recall that compliers are substantively unusual: compliers would not ordinarily watch the program but watch only because they were induced by the treatment assignment.

We now estimate the ATE using ICSW; our estimates of the ATE are presented in columns (2) and (4) of Table 2. Although Albertson and Lawrence (2009) find that compliers are 0.28 points

[11]The difference in coefficients, $\hat{\tau}_{ATE} - \hat{\tau}_{LATE}$, has a 95% confidence interval of (-0.2, 2.3) and a 90% confidence interval of (0.0, 2.2).

[12]Note that our replication of their results differs slightly from their original results due to the fact that we use IPW with all covariates as well as the treatment assignment indicator to address missing values in the dependent variable. This IPW step is also included in the bootstrap procedure. As Albertson and Lawrence note, missing values on the outcome variable appear to be missing at random, at least with respect to the treatment indicator.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

Table 2 2SLS and ICSW estimates for Albertson and Lawrence (2009)

	Knowledge		*Opinion*	
	LATE (2SLS) *(1)*	*ATE (ICSW)* *(2)*	*LATE (2SLS)* *(3)*	*ATE (ICSW)* *(4)*
Watching debate	**0.28**	**0.40**	**−0.07**	**−0.05**
	(−0.04, 0.59)	**(0.01, 1.26)**	**(−0.25, 0.13)**	**(−0.31, 0.24)**
Intercept	1.81	1.90	1.01	0.99
	(1.33, 2.26)	(1.34, 2.50)	(0.70, 1.30)	(0.60, 1.29)
Party ID	−0.02	−0.02	−0.08	−0.08
	(−0.05, 0.01)	(−0.06, 0.02)	(−0.10, −0.06)	(−0.10, −0.05)
Political interest	0.25	0.26	−0.03	−0.02
	(0.15, 0.36)	(0.15, 0.38)	(−0.09, 0.03)	(−0.09, 0.06)
Watch news	−0.00	−0.01	0.01	0.00
	(−0.05, 0.05)	(−0.07, 0.05)	(−0.02, 0.04)	(−0.02, 0.04)
Education	0.00	0.00	−0.01	−0.01
	(−0.02, 0.03)	(−0.03, 0.03)	(−0.03, 0.00)	(−0.03, 0.01)
Read news	0.11	0.10	−0.01	−0.01
	(0.07, 0.15)	(0.05, 0.15)	(−0.03, 0.02)	(−0.03, 0.02)
Female	−0.05	−0.06	−0.02	−0.00
	(−0.19, 0.07)	(−0.23, 0.10)	(−0.10, 0.06)	(−0.09, 0.12)
Income	−0.01	−0.02	0.01	0.01
	(−0.04, 0.02)	(−0.05, 0.01)	(−0.01, 0.02)	(−0.01, 0.02)
White	0.06	0.03	0.17	0.14
	(−0.11, 0.24)	(−0.19, 0.27)	(0.06, 0.29)	(−0.01, 0.27)

Note. Bootstrap 95% confidence intervals in parentheses. Dependent variables are knowledge (coded on a scale from 1 to 4 from least to most informed) and opinion (coded 1 if the respondent supported the proposition and 0 otherwise). Covariates include: television news watching habits (coded on a seven-point scale from never watches to watches every day), newspaper reading habits (coded on seven-point scale from never reads to reads every day), interest in politics and national affairs (coded on a four-point scale from low interest to high interest), party ID (coded on an eleven-point scale), income (coded on a scale from 1 to 11 from poorest to richest), female (coded 1 if the respondent is female and 0 otherwise), education (coded on a thirteen-point scale from least to most educated), and white (coded 1 if the respondent is white and 0 otherwise)

more informed after viewing the program, we find that viewers in the entire subject population are 0.40 points more informed after viewing. We therefore estimate that the ATE is approximately 45% larger than the LATE. Turning to our analysis of the effect of program viewing on support for the measure, we see that Albertson and Lawrence's (2009) 2SLS estimate and our ICSW estimate are nearly identical at −0.07 points and −0.05 points, respectively. We therefore find that the effect of viewing the debate on support for compliers is very similar to the estimated ATE, suggesting two possible interpretations. Since we cannot reject the null hypothesis of no treatment effect for either the LATE or the ATE, we may suspect that there is no effect on attitudes resulting from the treatment. Alternatively, if we believe there is a small treatment effect, we find no evidence that the LATE diverges from the ATE.[13]

5 Conclusion

IV methods have grown in popularity in recent years, due to their ability to overcome problems of endogeneity and selection bias. However, our results suggest that researchers should pay careful attention to the estimand when using this popular method, as the LATE can differ considerably from the ATE. When the ATE is the target parameter, ICSW may be a more sensible estimator than is standard IV. As the use of IV methods continues to expand, this type of recognition and discussion of the causal quantity of interest will only grow in importance.

[13]The difference in coefficients for knowledge, $\hat{\tau}_{ATE} - \hat{\tau}_{LATE}$, has a 95% confidence interval of (−0.02, 0.90) and a 90% confidence interval of (0.00, 0.52). The difference in coefficients for opinion has a 95% confidence interval of (−0.12, 0.16) and a 90% confidence interval of (−0.08, 0.12).

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

Appendix: ICSW Consistency

We derive expressions for the inverse compliance score weighted numerator (average ITT effect) and denominator (average probability of compliance) of equation (3), thus obtaining the asymptotic value of the IV estimator after ICSW. For clarity, we remove i subscripts to denote all units subject to any conditional probabilities articulated. As in equation (3), $\text{ITT} = \Pr(D_1 > D_0)\text{E}(Y_1 - Y_0|D_1 > D_0)$, which we may express as a weighted average of the conditional ITTs:

$$\text{ITT} = \int_{\Omega} \Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})\text{E}(Y_1 - Y_0|D_1 > D_0, \mathbf{X} = \mathbf{x})f(\mathbf{x})d\mathbf{x}. \tag{8}$$

With the equations above, we are now able to apply ICSW, multiplying by the inverse of the compliance score: $\frac{1}{\Pr(D_1 > D_0|\mathbf{X}=\mathbf{x})}$. We then divide by the average weight across all units (to normalize the weight). We define the average weight $\overline{w_c} = \int_{\Omega} \frac{1}{\Pr(D_1 > D_0|\mathbf{X}=\mathbf{x})} f(\mathbf{x})d\mathbf{x}$.

We can write the weighted ITT as follows:

$$\text{ITT}^w = \frac{1}{\overline{w_c}} \int_{\Omega} \frac{\Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})\text{E}(Y_1 - Y_0|D_1 > D_0|\mathbf{X} = \mathbf{x})}{\Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})} f(\mathbf{x})d\mathbf{x}. \tag{9}$$

Since the $\Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})$ terms cancel, equation (9) reduces to

$$\text{ITT}^w = \frac{1}{\overline{w_c}} \int_{\Omega} \text{E}(Y_1 - Y_0|D_1 > D_0,\ \mathbf{X} = \mathbf{x})f(\mathbf{x})d\mathbf{x}. \tag{10}$$

Applying Assumption 1, we can rewrite equation (10) as $\text{ITT}^w = \frac{1}{\overline{w_c}} \int_{\Omega} \text{E}(Y_1 - Y_0|\mathbf{X} = \mathbf{x})f(\mathbf{x})d\mathbf{x}$. By the law of total probability, the reweighted numerator of the IV estimator is

$$\text{ITT}^w = \frac{1}{\overline{w_c}} \int_{\Omega} \text{E}(Y_1 - Y_0|\mathbf{X} = \mathbf{x})f(\mathbf{x})d\mathbf{x} = \frac{1}{\overline{w_c}} \text{E}(Y_1 - Y_0). \tag{11}$$

We can now reweight the denominator of the IV estimator by expanding the definition of the complier population, weighting, and simplifying:

$$\Pr(D_1 > D_0) = \int_{\Omega} \Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})f(\mathbf{x})d\mathbf{x}, \tag{12}$$

$$\Pr(D_1 > D_0)^w = \frac{1}{\overline{w_c}} \int_{\Omega} \frac{\Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})}{\Pr(D_1 > D_0|\mathbf{X} = \mathbf{x})} f(\mathbf{x})d\mathbf{x} = \frac{1}{\overline{w_c}} \int_{\Omega} f(\mathbf{x})d\mathbf{x} = \frac{1}{\overline{w_c}}. \tag{13}$$

Dividing equation (11) by equation (13), we have the asymptotic value of the ICSW estimator,

$$\frac{\text{ITT}^w}{\Pr(D_1 > D_0)^w} = \frac{\frac{1}{\overline{w_c}}\text{E}(Y_1 - Y_0)}{\frac{1}{\overline{w_c}}} = \text{E}(Y_1 - Y_0), \tag{14}$$

which is the ATE. By Slutsky's theorem, a consistent estimator of $\frac{\text{ITT}^w}{\Pr(D_1 > D_0)^w}$ will be

$$\frac{\left(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i Y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i\right) - \left(\sum_{i=1}^{n} \hat{w}_{Ci}(1 - Z_i) Y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{w}_{Ci}(1 - Z_i)\right)}{\left(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i D_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{\hat{w}}_{Ci} Z_i\right) - \left(\sum_{i=1}^{n} \hat{w}_{Ci}(1 - Z_i) D_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{w}_{Ci}(1 - Z_i)\right)},$$

as

$$\left(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i Y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{w}_{Ci} Z_i\right) - \left(\sum_{i=1}^{n} \hat{w}_{Ci}(1 - Z_i) Y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{w}_{Ci}(1 - Z_i)\right) \to_p \text{ITT}^w$$

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

and

$$\left(\sum_{i=1}^{n}\hat{w}_{Ci}Z_iD_i\right)\Big/\left(\sum_{i=1}^{n}\hat{w}_{Ci}Z_i\right)-\left(\sum_{i=1}^{n}\hat{w}_{Ci}(1-Z_i)D_i\right)\Big/\left(\sum_{i=1}^{n}\hat{w}_{Ci}(1-Z_i)\right)\rightarrow_p \Pr(D_1>D_0)^w$$

under suitable regularity conditions.

References

Abadie, Alberto. 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association* 97(457):284–92.

Albertson, Bethany, and Adria Lawrence. 2009. After the credits roll: The long-term effects of educational television on public knowledge and attitudes. *American Politics Research* 37(2):275–300.

Altonji, J. G., and T. A. Dunn. 1996. The effects of family characteristics on the return to education. *Review of Economics and Statistics* 78:692–704.

Angrist, Joshua D., and Guido W. Imbens. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90(430):431–42.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91:444–55.

Angrist, Joshua D., and Ivan Fernandez-Val. 2013. ExtrapoLATE-ing: External validity and overidentification in the LATE framework. In *Advances in economics and econometrics: Tenth world congress,* Vol. 3, eds. Manuel Arellano Daron Acemoglu, and Eddie Deke, 401–36. Cambridge: Cambridge University Press.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Arceneaux, K., and D. W. Nickerson. 2009. Who is mobilized to vote? A re-analysis of 11 field experiments. *American Journal of Political Science* 53(1):1–16.

Arceneaux, Kevin, and David W. Nickerson. 2010. Comparing negative and positive campaign messages. *American Politics Research* 38(1):54–83.

Aronow, Peter M., and Allison Carnegie. 2013. Replication data for: Beyond LATE: Estimation of the average treatment effect with an instrumental variable. *Dataverse Network*. http://hdl.handle.net/1902.1/21729 (accessed August 1, 2013).

Deaton, Angus. 2009. Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. *Proceedings of the British Academy, 2008 Lectures* 162:123–60.

Elliott, Michael R. 2009. Model averaging methods for weight trimming in generalized linear regression models. *Journal of Official Statistics* 25(1):1–20.

Esterling, Kevin M., Michael A. Neblo, and David M. J. Lazer. 2011. Estimating treatment effects in the presence of noncompliance and nonresponse: The generalized endogenous treatment model. *Political Analysis* 19(2):205–26.

Fodor, Imola.K. 2002. A survey of dimension reduction techniques. *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory*, Vol. 9: 1–18.

Follmann, Dean A. 2000. On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association* 95(452):1101–9.

Frangakis, Constantine E., and Donald B. Rubin. 1999. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 86(2):365–79.

Funk, Michele Jonsson, Daniel Westreich, Chris Wiesen, Til Strmer, M. Alan Brookhart, and Marie Davidian. 2011. Doubly robust estimation of causal effects. *American Journal of Epidemiology* 173(7):761–7.

Geman, Stuart, and Chii-Ruey Hwang. 1982. Nonparametric maximum likelihood estimation by the method of sieves. *Annals of Statistics* 10(2):401–14.

Green, Donald P., Alan S. Gerber, and David W. Nickerson. 2003. Getting out the vote in local elections: Results from six door-to-door canvassing experiments. *Journal of Politics* 65(4):1083–96.

Heckman, James J., and Sergio Urzua. 2010. Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics* 156(1):27–37.

Hidalgo, F. D., S. Naidu, S. Nichter, and N. Richardson. 2010. Economic determinants of land invasions. *Review of Economics and Statistics* 92(3):505–23.

Hirano, Keisuke, Guido W. Imbens, Donald B. Rubin, and Xiao-Hua Zhou. 2000. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 1(1):69–88.

Horvitz, D. G., and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47(260):663–85.

Humphreys, Macartan. 2009. *Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities*. Working paper.

Imbens, Guido W. 2009. *Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)*. NBER Working paper.

Imbens, Guido W., and Donald B. Rubin. 1997. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* 25(1):305–27.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press

Joffe, Marshall M., and Colleen Brensinger. 2003. Weighting in instrumental variables and G-estimation. *Statistics in Medicine* 22(1):1285–303.

Joffe, Marshall M., Thomas R. Ten Have, and Colleen Brensinger. 2003. The compliance score as a regressor in randomized trials. *Biostatistics* 4(3):327–40.

Putnam, Robert C. 2000. *Bowling alone: The collapse and renewal of American Community*. New York: Simon and Schuster.

Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, participation, and democracy in America*. New York: Macmillan Publishing Company.

Roy, Jason, Joseph W. Hogan, and Bess H. Marcus. 2008. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* 9(2):277–89.

Rubin, Donald B. 1978. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6(1):34–58.

———. 2009. Author's reply: Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Statistics in Medicine* 28(9):1420–23.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.

Wald, A. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11:284–300.

Yau, Linda H. Y., and Roderick J. Little. 2001. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* 96:1232–44.

https://doi.org/10.1093/pan/mpt013 Published online by Cambridge University Press