

# Coarsening Bias: How Coarse Treatment Measurement Upwardly Biases Instrumental Variable Estimates

John Marshall

*Department of Government, Harvard University, Cambridge, MA 02138*

*e-mail: jlmarsh@fas.harvard.edu (corresponding author)*

Edited by Jonathan Katz

Political scientists increasingly use instrumental variable (IV) methods, and must often choose between operationalizing their endogenous treatment variable as discrete or continuous. For theoretical and data availability reasons, researchers frequently coarsen treatments with multiple intensities (e.g., treating a continuous treatment as binary). I show how such coarsening can substantially upwardly bias IV estimates by subtly violating the exclusion restriction assumption, and demonstrate that the extent of this bias depends upon the first stage and underlying causal response function. However, standard IV methods using a treatment where multiple intensities are affected by the instrument—even when fine-grained measurement at every intensity is not possible—recover a consistent causal estimate without requiring a stronger exclusion restriction assumption. These analytical insights are illustrated in the context of identifying the long-run effect of high school education on voting Conservative in Great Britain. I demonstrate that coarsening years of schooling into an indicator for completing high school upwardly biases the IV estimate by a factor of three.

## 1 Introduction

Instrumental variables (IVs) are becoming a standard component of the political scientist's methodological toolkit (Sovey and Green 2011). IV analyses have illuminated the effects of political institutions on economic development (Acemoglu, Johnson and Robinson 2001), international trade agreements on foreign direct investment (Büthe and Milner 2008), and campaign spending on election outcomes (Gerber 1998). Given that an appropriate instrument can identify important causal relationships that cannot be easily disentangled, it is not surprising to find that the number of articles published in the *American Journal of Political Science* (AJPS) and the *American Political Science Review* (APSR) using IV techniques has almost doubled over the last decade (see Fig. A1 in the Online Appendix).

Although best practice for using IV methods is now receiving greater scrutiny (e.g., Angrist and Pischke 2008; Dunning 2008; Sovey and Green 2011), this article illustrates a potentially severe but rarely-recognized source of bias: how coarsening a continuous or multi-valued (endogenous) treatment variable can substantially upwardly bias IV estimates.<sup>1</sup> This bias is particularly important given that such coarsening may often appear appealing to applied researchers in cases where: (1) the researcher believes that coarsening aids interpretation, or avoids imposing linearity on a statistical relationship; (2) theory suggests that the treatment effect may be non-linear; or (3) more granular measures of the treatment are unavailable. Since 2005, 36% of AJPS and APSR publications using IV methods instrument for binary or ordinal treatments with few categories, and may be susceptible to coarsening bias.<sup>2</sup> Furthermore, coarsening bias counters the perceived wisdom—based on the

---

*Authors' note:* I thank Matt Blackwell, John Bullock, Anthony Fowler, Andy Hall, Torben Iversen, Horacio Larreguy, Rakeen Mabud, Daniel Moskowitz, Arthur Spirling, Brandon Stewart, Dustin Tingley, Tess Wise, the editor and two anonymous referees for illuminating discussions or useful comments. Replication materials are available online as Marshall (2016). Supplementary materials for this article are available on the *Political Analysis* Web site.

<sup>1</sup>Following Angrist, Imbens, and Rubin (1996), I refer to the endogenous variable as the “treatment.”

<sup>2</sup>I count ordinal treatments with five or fewer categories.

case of classical measurement error—that IV methods necessarily help to correct for measurement error in the treatment variable.<sup>3</sup>

Building on Angrist and Imbens's (1995) analysis of IV methods when the treatment can take multiple ordered values (or intensities), I show analytically and demonstrate empirically how coarsening an endogenous treatment variable can substantially bias IV estimates. Intuitively, coarsening bias arises when an instrument affects the intensity of the underlying treatment, which in turn affects the outcome of interest, but does not register a change in the treatment status when it is operationalized as a binary variable.<sup>4</sup> For example, a dummy for completing high school would fail to register the effects of school leaving laws that encouraged some students to stay in school for an additional year *without* completing high school. But by grouping together multiple years of schooling (or treatment intensities) that each affect the outcome, coarsening falsely creates the impression that the sum of the effects at each intensity can be attributed to completing high school. Coarsening bias therefore violates the exclusion restriction assumption underpinning IV identification strategies (e.g., Angrist, Imbens, and Rubin 1996), because the instrument affects the outcome through an avenue not captured by the treatment when operationalized as a dichotomous variable.

I demonstrate that coarsening bias is especially large when: (1) any change in treatment intensity affects the outcome, for example, if the treatment's true effect is linear; and (2) when the instrument has large effects, or a large "first stage," on receiving treatment intensities other than the coarsened treatment threshold.<sup>5</sup> In general, only when the effect of a treatment is concentrated at a single intensity such as completing high school (among the intensities affected by the instrument) *and* the researcher is able to both recognize and measure the value of the treatment where this occurs, will the IV estimate associated with a coarsened treatment variable be consistently estimated. These conditions constitute a demanding assumption that I refer to as the *strong exclusion restriction*.

In many observational studies, this strong exclusion restriction may not hold. Consider the case of Pierskalla and Hollenbach (2013), who use communication regulations as instruments to identify the effect of a binary measure of local cell phone coverage on collective action. If favorable regulations increase cell phone usage without affecting their coarse measure of cell phone coverage, coarsening bias could explain why their IV estimates are twenty times larger than the corresponding OLS estimates. Similarly, Milligan, Moretti, and Oreopoulos (2004) use the compulsory school laws of U.S. states to instrument for completing high school, and report that completing high school increases turnout by 30–40 percentage points and the likelihood that an individual follows politics by 40–85 percentage points. These estimates are upwardly biased if these laws increased education levels below the point of completing high school, and such lower levels of education also affect political outcomes.

In experimental studies, the extent of coarsening bias depends upon the type of endogenous variable that a randomized instrument affects. Experiments inducing respondents to uptake truly binary treatments are not affected by this bias. For example, get-out-the-vote canvassing is unlikely to impact respondents that did not answer the door (e.g., Gerber and Green 2000). However, for Gerber, Huber, and Washington (2010), who in one specification instrument for an individual's partisan identification and find effects fifteen times larger than their corresponding OLS estimates, upward bias may occur if their randomized mailing causes voters to move toward a political party without passing the threshold required to register a new partisan identification.<sup>6</sup>

<sup>3</sup>Kane, Rouse, and Staiger (1999) make a similar point for categorical treatment variables measured with non-classical error. However, this article shows that IV estimates are upwardly biased even when a categorical treatment variable is correctly measured.

<sup>4</sup>The argument equally applies when only a single value of an ordinal treatment is affected by the instrument because the treatment effectively serves as a dummy variable.

<sup>5</sup>Angrist and Imbens (1995) only flag the second condition, and thus do not discuss the key role of the CRF, or its interaction with the first stage. This article also shows how coarsening bias arises from a subtle exclusion restriction and how identification requires a stronger exclusion restriction depending upon both the first stage and underlying CRF. I also exploit a rare applied opportunity to disentangle and quantify coarsening bias.

<sup>6</sup>Gerber, Huber, and Washington (2010) also consider a seven-point partisanship scale, which is unlikely to be biased.

However, a consistent causal estimate can still be recovered without requiring that the strong exclusion restriction holds. When the treatment is coded as a multi-valued variable, where the instrument affects multiple treatment intensities, standard IV methods consistently estimate the *local average per-unit treatment effect* (LAPTE), or the average causal effect of a unit increase in the treatment among compliers (Angrist and Imbens 1995). My analysis offers novel justifications for this linearization, which is typically used when the treatment is believed to linearly affect the outcome. First, although a different quantity of interest is returned, this approach always provides a consistent estimate when the instrument affects at least two levels of the treatment—even when the treatment is non-linearly related to the outcome. Second, without the strong exclusion restriction, the researcher need not correctly specify the functional form relating the treatment to the outcome. Finally, under certain conditions, this approach is robust even without observing all treatment intensities. Nevertheless, when the treatment’s effects are highly non-linear across intensities, the intent to treat (ITT) effect estimated by the reduced form may be preferred to an IV estimate relying on assumptions about how the treatment effect varies by intensity for its interpretation.

While coarsening bias is clear in theory, its threat to inference in applied settings is difficult to gauge. I illustrate its importance in the context of Marshall’s (forthcoming) study of late high school education’s downstream effect on vote choice in Great Britain. While the study uses the consistent linearization approach, this article demonstrates that instrumenting for an indicator for completing high school would have overestimated the causal effect of late high school education on voting Conservative later in life by a factor of three.<sup>7</sup> Several properties ensure that this example offers a rare opportunity to differentiate coarsening bias from other potential explanations for large IV estimates. The results suggest that coarsening bias can explain why IV estimates are often orders of magnitude larger than the corresponding OLS or reduced form estimates. To avoid drawing biased inferences, which could cause policy-makers to mistakenly adopt relatively ineffective policies, this article thus demonstrates that it is imperative that political scientists become aware of how coarsening an endogenous treatment variable can dramatically inflate IV estimates.

## 2 IVs with Coarsened Treatments

Given that many interesting treatment variables cannot be easily randomized, IV approaches offer an appealing means for researchers to estimate causal relationships where they suspect that the treatment is subject to endogeneity (Sovey and Green 2011).<sup>8</sup> The key idea is that a researcher finds an “instrument” which (1) affects the level of the (endogenous) treatment variable, but (2) does not itself affect the outcome except by affecting the treatment. In my running example of schooling, raising the minimum school leaving age is likely to keep students in school longer, but is unlikely to affect political preferences except by keeping students in school longer.

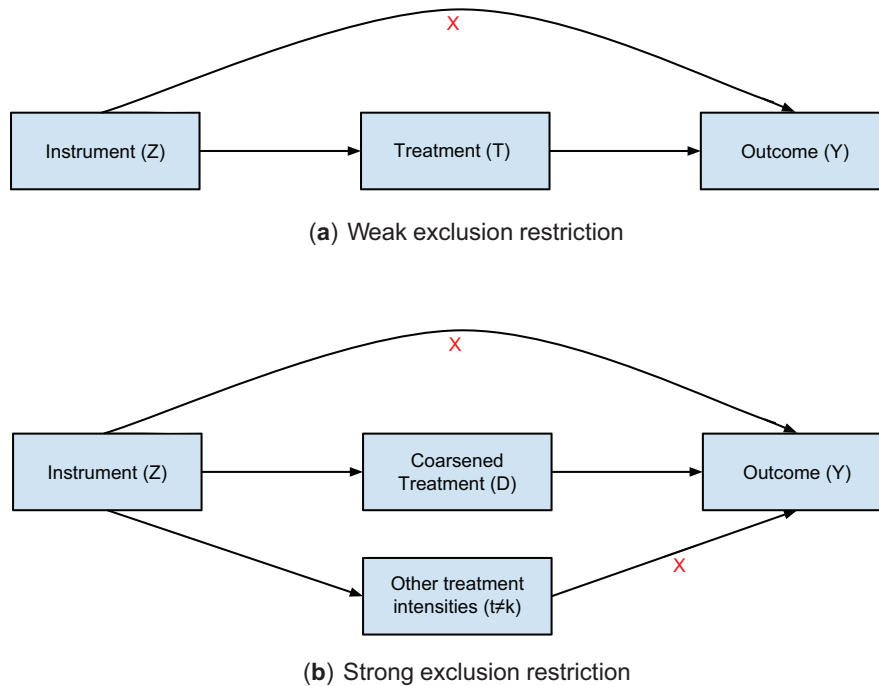
After identifying a plausible instrument, researchers must decide how to operationalize their endogenous treatment variable. For example, should education be coded as years of schooling or an indicator for completing high school? This article shows that the decision to coarsen a treatment can undermine the goal of IV estimation—consistent estimation of a treatment’s causal effect. Failing to consistently estimate an effect means that, even as the sample size becomes large, the IV estimate will not converge to the true value and confidence intervals will exclude the true value with probability approaching 1.<sup>9</sup> This section analyzes this inconsistency, focusing on the simplest case where there is a binary instrument and the treatment is coarsened into a binary indicator.<sup>10</sup>

<sup>7</sup>Replication data and code can be found on the *Political Analysis* dataverse (Marshall 2016).

<sup>8</sup>Another motivation for using IV is to address classical measurement error, which attenuates OLS estimates. This article shows that this motivation does not apply to non-classical measurement error induced by coarsening.

<sup>9</sup>IV estimators are biased but consistent in finite samples (Staiger and Stock 1997).

<sup>10</sup>All results naturally extend to multi-valued instruments, multiple instruments, multi-valued coarsened treatments, and the inclusion of control variables.



**Fig. 1** Graphical representation of weak and strong exclusion restrictions.

## 2.1 Setup and Notation

IV analyses start from observing an instrument, a treatment, and an outcome. The binary instrument, for each observation  $i = 1, \dots, N$ , is denoted by  $Z_i \in \{0, 1\}$ . The endogenous treatment intensity of observation  $i$ ,  $T_i \in \{1, \dots, J\}$ , takes one of  $J$  ordered intensities. As  $J$  becomes large, the analysis generalizes to essentially continuous treatments. Finally,  $Y_i$  is  $i$ 's observed outcome of interest.

Following Angrist, Imbens, and Rubin (1996), I examine the IV estimator in terms of heterogeneous potential outcomes. Before defining potential outcomes, we first assert that the stable unit treatment value assumption (SUTVA) holds, which implies that  $i$ 's potential outcomes are not related to the instrument and the treatment values received by any other individual. Given SUTVA, we may write  $T_i(Z_i) \equiv T_{iZ}$  as  $i$ 's potential outcomes of  $T_i$  for any value of the instrument  $Z_i$  that  $i$  receives, and  $Y_i(Z_i, T_i(Z_i)) \equiv Y_{iZT_Z}$  correspondingly denotes  $i$ 's potential outcomes of  $Y_i$ .<sup>11</sup>

In addition to SUTVA (A1), IV estimation relies on four additional assumptions. First, that the instrument is independent of potential outcomes and potential treatment intensities (A2). This follows from the (conditional) random assignment of an instrument, such as a leaving age reform that alters incentives to remain in school across otherwise similar cohorts. Second, that there exists a “first stage” (A3), such that the instrument affects the intensity of the treatment  $i$  receives. Third, monotonicity (A4) requires that, for all individuals, the instrument either never decreases or never increases the treatment intensity. This rules out “defiers” that leave school earlier in response to raising the leaving age. Fourth, the *weak* exclusion restriction requires that  $Z_i$  only affects  $Y_i$  through the treatment  $T_i$  (A5). Consequently,  $Y_{izt} = Y_{it}$  for any  $z$  and  $t$ . The diagram in Fig. 1a represents this weak exclusion restriction graphically, marking by “X” the exclusion of the possibility that a schooling leaving age reform affects political preferences through a path other than the additional education that an individual receives.

These standard assumptions, which are explained in greater detail elsewhere (e.g., Angrist, Imbens, and Rubin 1996; Angrist and Pischke 2008; Sovey and Green 2011) are formalized below:

<sup>11</sup>Given SUTVA, observed outcomes relate to potential outcomes through  $T_i = Z_i T_{i1} + (1 - Z_i) T_{i0}$  and  $Y_i = Z_i Y_{i1T_i} + (1 - Z_i) Y_{i0T_0}$ .

- A1. Stable unit treatment value assumption: for all  $i = 1, \dots, N$ , and where  $Z_{-i}$  and  $T_{-i}$  represent the vectors of instrument and treatment assignments for all observations except  $i$ , (1) for all  $w$  and  $w'$ ,  $T_i(Z_i, Z_{-i} = w) = T_i(Z_i, Z_{-i} = w')$ ; and (2) for all  $w, w', v$ , and  $v'$ ,  $Y_i(Z_i, T_i(Z_i), Z_{-i} = w, T_{-i}(w) = v) = Y_i(Z_i, T_i(Z_i), Z_{-i} = w', T_{-i}(w') = v')$ .
- A2. Instrument independence: for all  $i = 1, \dots, N$ ,  $T_{i0}, T_{i1}, Y_{i0T_0}$ , and  $Y_{i1T_1}$  are jointly independent of  $Z_i$ .
- A3. First stage:  $\mathbb{E}[T_i|Z_i = 1] - \mathbb{E}[T_i|Z_i = 0] \neq 0$ .
- A4. Monotonicity: for all  $i = 1, \dots, N$ ,  $T_{i1} - T_{i0} \geq 0$  or  $T_{i1} - T_{i0} \leq 0$ .
- A5. Weak exclusion restriction: for all  $i = 1, \dots, N$ ,  $Y_{izt} = Y_{iz't}$  for any  $t$  and all  $z$  and  $z'$ .

Without loss of generality, I analyze the case where  $T_{i1} - T_{i0} \geq 0$ .

For any value  $t$  of the treatment, the local average treatment effect (LATE) of moving from treatment intensity  $t - 1$  to treatment intensity  $t$  because of the instrument is defined as  $\beta_t \equiv \mathbb{E}[Y_{it} - Y_{it-1}|T_{i1} \geq t > T_{i0}]$ .  $Y_{it} - Y_{it-1}$  defines the difference in the outcome resulting from receiving treatment intensity  $t$  as opposed to intensity  $t - 1$ , while  $T_{i1} \geq t > T_{i0}$  defines this effect for the set of compliers that only reach intensity  $t$  when they receive the instrument. Collecting the LATE across all intensities defines the *causal response function* (CRF), and thus describes how each level of the treatment affects the outcome among compliers. The shape of the CRF in any given empirical application will prove crucial for understanding when coarsening is problematic.

## 2.2 Coarsening Bias

Coarsening bias may occur when the researcher, whether by choice or constrained by data availability, coarsens their measure of the treatment. In particular, in the hope of identifying the LATE of obtaining intensity  $k$  beyond only obtaining the preceding level  $k - 1$ , or  $\beta_k$ , the treatment intensity  $T_i$  is partitioned by defining the indicator  $D_{ik} \equiv 1(T_i \geq k)$ .<sup>12</sup> This coarsened binary variable indicates whether an individual receives at least treatment intensity  $k$ , and could represent the effect of completing high school beyond completing the penultimate grade of high school.

The coarsened IV estimator for the effect of  $D_{ik}$  on  $Y_i$ , denoted  $\hat{\beta}_k^{IV}$ , divides the reduced form effect of  $Z_i$  on  $Y_i$  by the first-stage effect of  $Z_i$  on  $D_{ik}$  (see e.g., Angrist and Pischke 2008). Separate causal estimates of the reduced form and first stage are identified under assumptions A1 and A2. Further utilizing assumptions A4 and A5, the probability limit of the coarsened IV estimator can be expressed as the weighted sum of the causal effect for compliers moving from intensity  $t - 1$  to  $t$ ,  $\beta_t$ , for each such interval (see Angrist and Imbens [1995] and the proof in the Online Appendix):

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_k^{IV} \equiv \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_{ik}|Z_i = 1] - \mathbb{E}[D_{ik}|Z_i = 0]} = \frac{\sum_{t=2}^J p_t \beta_t}{p_k}, \quad (1)$$

where  $\text{plim}_{N \rightarrow \infty}$  denotes the probability limit as the sample sizes approaches  $\infty$ . Importantly,  $p_t \equiv \Pr(T_{i1} \geq t > T_{i0}) = \mathbb{E}[D_{it}|Z_i = 1] - \mathbb{E}[D_{it}|Z_i = 0]$  denotes the probability that an individual only reaches treatment intensity  $t$  because they received the instrument  $Z_i = 1$ , and thus represents the proportion of compliers at treatment intensity  $t$  in the population.<sup>13</sup> For example,  $p_{11}$  captures the probability that raising the school leaving age induces an individual that would have stayed in school for less than 11 years without the instrument to stay in school for at least 11 years. Each  $p_t$ , including  $p_k$ , can be estimated by regressing an indicator for  $T_i \geq t$  on  $Z_i$  (Angrist and Imbens 1995).

The probability limit in equation (1) is well-defined under A3, provided  $p_k > 0$ . In the case of schooling, equation (1) shows that the coarsened IV estimator for the effect of completing high school converges to the sum of the effects of reaching *each* additional year of schooling—not just

<sup>12</sup>Angrist and Imbens (1995) describe this as an “incorrectly coded binary treatment.”

<sup>13</sup>Where the instrument induces  $i$  to receive more than one additional intensity,  $i$  is counted in all first stages, so contributes once to each  $p_t$  (Angrist and Imbens 1995). When estimating the LAPTE (see below), the effect is thus effectively split between intensities.

completing high school—weighted by the number of individuals completing each additional year of schooling only because they received the instrument.

However, under the standard assumptions, the coarsened IV estimator rarely consistently estimates our quantity of interest: the LATE of obtaining treatment intensity  $k$ . Inspection of equation (1) shows that  $\hat{\beta}_k^{IV}$  converges to  $\beta_k$  when  $\sum_{t \neq k} p_t \beta_t = 0$ . Therefore,  $\hat{\beta}_k^{IV}$  only consistently estimates  $\beta_k$  in four special cases when:

1. The instrument only affects reaching intensity  $k$ ; or  $p_t = 0, \forall t \neq k$ .
2. The LATE at all intensities other than  $k$  is zero; or  $\beta_t = 0, \forall t \neq k$ .
3. One of the preceding conditions holds for each intensity  $t$ , ensuring that  $p_t \beta_t = 0, \forall t \neq k$ .
4. The direction of the effects (weighted by  $p_t$ ) differs across intensities, but ultimately cancel out; or  $\sum_{t \neq k} p_t \beta_t = 0$ .

These special cases are often implausible in practice. First, as the number of intensities increases, or as  $T_i$  becomes effectively continuous, it becomes increasingly implausible that only one intensity is affected. Second, as the discussion of the examples in the introduction suggests, it is often hard to believe that only one particular intensity of the treatment affects the outcome.

The following proposition summarizes these insights:

**Proposition 1 (Coarsening Bias).** *Assume A1, A2, A3, A4, and A5 hold, and that  $p_k > 0$ . The probability limit of the coarsened IV estimator  $\hat{\beta}_k^{IV}$  can be expressed as:*

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_k^{IV} = \beta_k + \frac{\sum_{t \neq k} p_t \beta_t}{p_k}. \quad (2)$$

*Provided  $\text{sign}(\beta_k) = \text{sign}(\beta_t)$  for all  $t \neq k$  where  $p_t \neq 0$ , the limit of the coarsened IV estimator accentuates the true causal effect:  $|\beta_k| \leq |\text{plim}_{N \rightarrow \infty} \hat{\beta}_k^{IV}|$ .*

**Proof.** All proofs are provided in the Online Appendix. ■

Provided that each treatment intensity causes  $Y_i$  to shift in the same direction (i.e., the CRF is monotonic in  $T_i$ ), Proposition 1 further demonstrates that the coarsened IV estimator is *upwardly* biased in magnitude. This upward bias is increasing in both  $p_t/p_k$  and  $|\beta_t|$ , for any  $t \neq k$ . In other words, coarsening bias is greater when the first stage is relatively large at intensities other than  $k$  and the LATE at intensities other than  $k$  is large. If  $\beta_{t \neq k} \in [a, b]$ , for all  $t \neq k$ , and  $p_t \geq 0$ , then the bounds on the true value of  $\beta_k$  are  $\beta_k \in [\hat{\beta}_k^{IV} - sb, \hat{\beta}_k^{IV} - sa]$ , where  $s \equiv \sum_{t \neq k} p_t/p_k$  is the relative strength of the first stage at intensities other than  $k$ .

The general inconsistency of the coarsened IV estimator reflects a subtle exclusion restriction violation that arises from coarsening  $T_i$  into  $D_{ik}$ . Assumption A5 requires that  $Z_i$  does not affect  $Y_i$  through avenues other than  $T_i$ . However, coarsening  $T_i$  into  $D_{ik}$  allows for  $T_i$  to affect  $Y_i$  without going through  $D_{ik}$ , but without violating A5. Consequently, while any effect of  $Z_i$  on  $Y_i$  through  $T_i$  is registered in the reduced form (the numerator of equation (1)), the first stage (the denominator) only registers cases where  $Z_i$  induces  $i$  to pass the threshold used to define the treatment indicator  $D_{ik}$  (i.e., moving from intensity  $k - 1$  to  $k$ ). Coarsening bias thus arises if values of  $T_i$  other than  $k$ , such as years of high school before completing high school, also affect  $Y_i$ , because such changes in the reduced form are not captured in the first stage.

Consistent estimation of the effect of  $D_{ik}$  requires a stronger assumption. The *strong* exclusion restriction (A5\*) that  $Z_i$  only affects  $Y_i$  through  $D_{ik}$  is sufficient:<sup>14</sup>

**A5\*.** Strong exclusion restriction: for all  $i = 1, \dots, N$ , all  $t$  such that  $p_t \neq 0$ , and all  $z$  and  $z'$ , (1)  $Y_{izt} = Y_{iz't'}$  for all  $t, t' \geq k$ ; and (2)  $Y_{izt} = Y_{iz't'}$  for all  $t, t' < k$ .

<sup>14</sup>This captures the first, second, and third of the four special cases described above. The second is probably the most empirically relevant, given that all  $\beta_t$ s typically go in the same direction and most instruments induce  $p_t \neq 0$  for some  $t \neq k$ .

As illustrated in Fig. 1b, this assumption is more demanding than A5 because, in addition to requiring that the instrument only affects the outcome by altering the intensity of the treatment, it also requires that the researcher coarsen the treatment such that  $\beta_t = 0$  for all intensities  $t \neq k$  that are affected by the instrument.

The next proposition establishes the consistency of the coarsened IV estimator when A5\* holds, and thus demonstrates the importance of the seemingly minor distinction between assumptions A5 and A5\* when  $T_i$  is coarsened into  $D_{ik}$ .

**Proposition 2 (Consistency under the Strong Exclusion Restriction).** *If  $p_k > 0$  and assumptions A1, A2, A3, A4, and A5\* hold,  $\hat{\beta}_k^{IV}$  consistently estimates  $\beta_k$ .*

If we cannot find an instrument that only affects completion of high school, Proposition 2 shows that consistent estimation of the effect of completing high school using the coarsened IV estimator is only possible when other levels of schooling impacted by the instrument do not affect the outcome.

### 2.3 Linearization as an Alternative to Coarsening

If coarsening an endogenous treatment can substantially upwardly bias IV estimates, what alternative strategies are available? In general, the CRF cannot be identified (see Abadie 2003). Although there exist non-linear and semi-parametric IV estimation strategies (e.g., Abadie 2003; Newey and Powell 2003), such approaches rely on stronger assumptions, require large amounts of data, and are technically challenging to estimate. However, a simple alternative to coarsening is to operationalize the treatment as a linear intensity, and thus produce a complier-weighted linear approximation to the CRF. Although this is the default approach for many researchers, its desirable theoretical properties, and the cases when it performs best, have not been clearly articulated.

The linearized IV estimator replaces  $D_{ik}$  with  $T_i$  in the first stage. The first stage could thus regress years of education, rather than an indicator for completing high school, on  $Z_i$ . However, by using  $T_i$  in the first stage to re-scale the reduced form estimate, Angrist and Imbens (1995) show that the estimand changes and instead identifies the LAPTE:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{\text{LAPTE}}^{IV} \equiv \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[T_i|Z_i = 1] - \mathbb{E}[T_i|Z_i = 0]} = \frac{\sum_{t=2}^J p_t \beta_t}{\sum_{t=2}^J p_t}. \quad (3)$$

The LAPTE is thus a weighted linear approximation: the causal effect at each treatment intensity is weighted by the proportion of compliers at that intensity. Therefore, if the instrument primarily induces students to complete high school, the LAPTE disproportionately weights the effect of completing high school.

The LAPTE estimator is a more robust approach than coarsening in three important respects, which provide important justifications for researchers already using this approach. First, only the weaker exclusion restriction (A5) is required for consistent estimation of the LAPTE. The weaker assumption required reflects the fact that the LAPTE “corrects” the first stage associated with the coarsened estimator by capturing all effects of the instrument on the endogenous treatment. We can see this by noting that only the denominator differs between equations (1) and (3). Provided that the instrument increases or decreases all affected treatment intensities, the signs of the LAPTE and coarsened estimators are always the same. Second, and consequently, even when the true CRF is not exactly linear, the LAPTE estimator represents a consistent linear approximation.

Third, the linearized IV approach can be robust even without observing all intensities. If the  $J$  observed intensities represent a coarsening of the true intervals (e.g., because  $T_i$  is continuous), a

linear causal effect can still be recovered provided that the intervals between intensity categories are equally spaced.

**Proposition 3 (Robustness of LAPTE to Missing Intensity Categories).** *Let only  $J$  equally spaced categories of  $T_i$  be observed when there are in fact  $\alpha J$  equally spaced categories, where  $\alpha > 1$  is finite and  $\alpha J$  is an integer. Denote  $\hat{\beta}_{LAPTE}^{IV,J}$  and  $\hat{\beta}_{LAPTE}^{IV,\alpha J}$ , respectively, as the IV estimators in the observed sample (denoted by superscript  $J$ ) and unobserved sample (denoted by superscript  $\alpha J$ ). Let assumptions A1–A5 hold, and assume  $p_t > 0$  for at least two intensities. If the effect of  $T_i$  is linear such that  $\beta_j^J = \tau$  for all intervals  $j$ , then  $\text{plim}_{N \rightarrow \infty} \hat{\beta}_{LAPTE}^{IV,J} = \alpha \text{plim}_{N \rightarrow \infty} \hat{\beta}_{LAPTE}^{IV,\alpha J}$ .*

Consequently, obtaining the coefficient on the quantity of interest only requires an adjustment by factor  $\alpha$  to identify the average linear causal effect for any unit interval.

In Proposition 3, a first stage is required for at least two treatment intensities to ensure that the IV estimate averages across coarsened categories. Without estimates at two different intensities to “draw a line through,” IV estimates would be equally susceptible to coarsening bias. Consequently, if treatment measurement is sufficiently coarse that the instrument only affects a single intensity, researchers may require a new data set, or a second data set drawn from the same population that can separately estimate the first stage (see Angrist and Pischke 2008). If such alternatives are not feasible, researchers should focus on estimating the reduced form ITT effect.

## 2.4 Coarsening in Practice: CRFs, First-Stage Weights, and Bias

Proposition 1 demonstrated that the extent of coarsening bias depends upon the first stage and the shape of the CRF. Specifically, bias depends upon the LATE at different treatment intensities ( $\beta_t$ ) weighted by the relative effects of the instrument on the treatment intensities that do not define the coarsened treatment ( $\sum_{t \neq k} p_t/p_k$ ). Since the CRF is not known in advance, it is essential to understand the types of causal relationships for which the strong exclusion restriction required for Proposition 2 is tenable, and what causal quantities the linear approximation can recover when only the weak exclusion restriction is tenable. This subsection compares the coarsened and linearized IV approaches to estimating “single jump,” linear, and non-linear CRFs.

### 2.4.1 Single jumps in the CRF

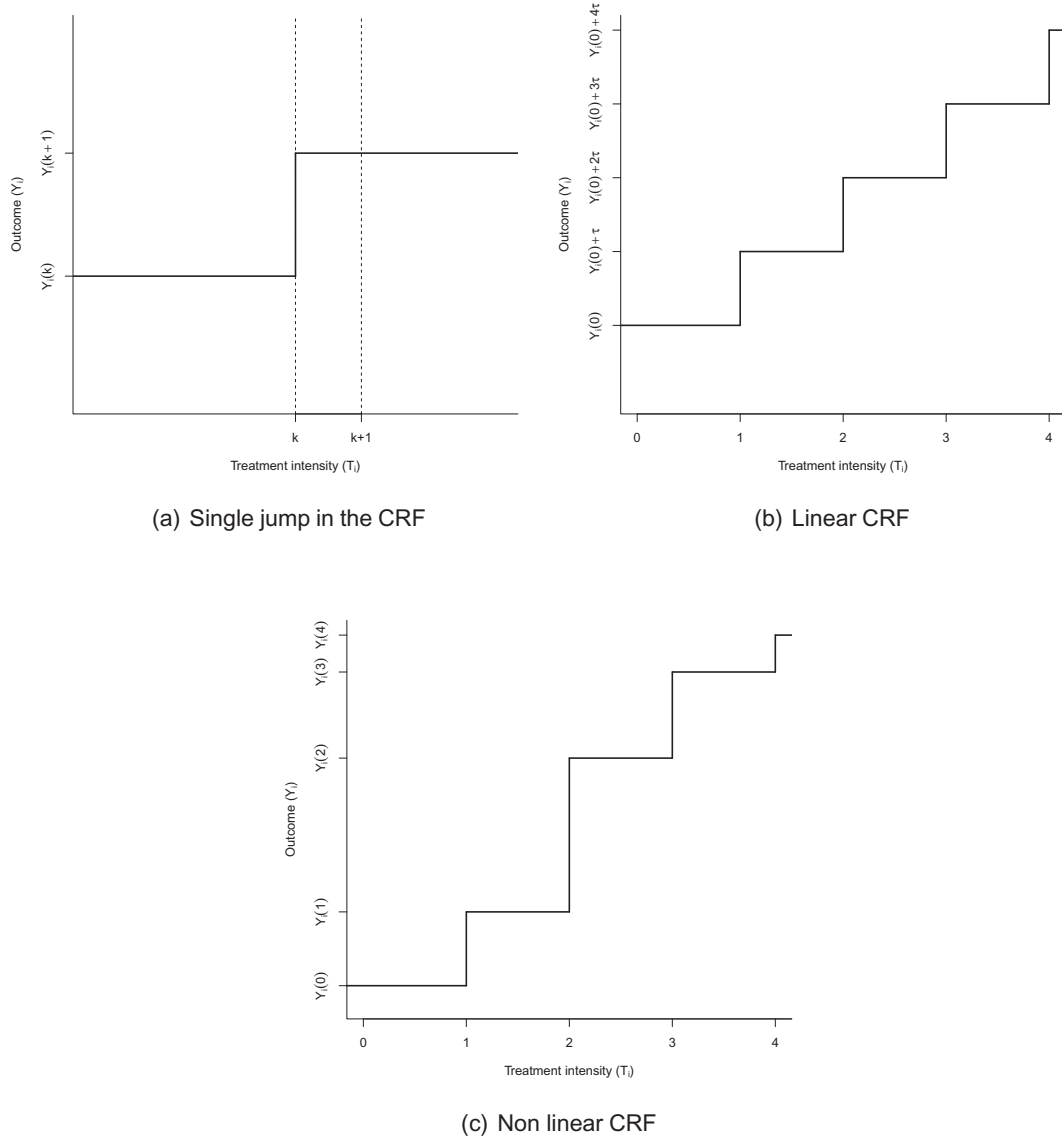
When the CRF exhibits a single jump, coarsening can be appropriate. Figure 2a depicts the case where no level of schooling other than completing high school (i.e., reaching treatment intensity  $k$ ) affects the outcome. Provided that the researcher is able to correctly identify intensity  $k$ —the only point at which there is a (positive) causal effect in the figure—as the key jump, then  $\beta_k$  can be consistently estimated using the coarsened IV estimator  $\hat{\beta}_k^{IV}$  when a suitable instrument ensures  $p_k > 0$ . This works because  $\beta_t = 0$  for all  $t \neq k$  for this CRF, and thus the coarsened IV estimator is consistent regardless of whether  $p_t > 0$  for some other  $t \neq k$ .

In practice, however, it is hard to know whether  $k$  correctly captures the true jump in the CRF. In general, tipping points are not straight-forward to predict. If the researcher incorrectly surmises that  $k + 1$  is the correct threshold, *at best* they fail to detect the existence of the effect of intensity  $k$  but consistently identify no effect at  $k + 1$ . In the simple example of Fig. 2a, where  $\beta_t = 0$  for all  $t \neq k$ , the researcher correctly concludes that  $\beta_{k+1} = 0$  only if their instrument does not induce subjects to reach intensity  $k$ . In other words,  $p_k = 0$  ensures a consistent estimate of a quantity that was probably not of primary interest. However, when  $p_k > 0$  and  $p_{k+1} > 0$ , the coarsened IV estimator will produce the following inconsistent estimate of the LATE at intensity  $k + 1$ :

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{k+1}^{IV} = \frac{p_k \beta_k}{p_{k+1}} > 0. \quad (4)$$

Although approximately correct in the sense that there is a causal effect *nearby*, this estimator both





**Fig. 2** Examples of CRFs.

wrongly attributes the effect to intensity  $k + 1$  and does not even consistently estimate the effect at intensity  $k$  unless  $p_{k+1} = p_k$ .

In experiments where subjects cannot be partially treated it is easier to determine clear cutoffs. As noted above, in the case of Gerber and Green’s (2000) get-out-the-vote canvassing, knocking on a door is only likely to affect respondents that opened the door to receive the treatment. But even experiments can be hard to evaluate if there is partial compliance, such that individuals can experience some of the treatment without being designated as treated. This could occur, for example, if subjects partially learn about a treatment by reading an email heading, but are not registered as treated because they did not open the email. If such information affects political behavior, then the strong exclusion restriction is violated.

The linearized IV estimator with an uncoarsened treatment faces different issues when the CRF is singularly discontinuous. The linearization introduces a trade-off between consistent estimation and the quantity of interest. On the one hand,  $\hat{\beta}_{LAPTE}^{IV}$  provides a consistent causal estimate regardless of the CRF, and is thus robust to mis-specifying the threshold of the discontinuity. On the

other hand, weighting the effect at each intensity by its relative first stage will underestimate the effect of the sharp jump at intensity  $k$  when intensities other than  $k$  do not affect the outcome. If additional years of schooling do not affect an outcome except upon completion of high school, the LAPTE will positively weight the effect of completing high school *and* the zero effects at years other than completing high school. While this remains a well-defined causal quantity, it may differ from the researcher's primary quantity of interest. However, CRFs that approximate sharp jumps may be empirically rare, and hard to compellingly identify in practice.

### 2.4.2 Linear CRFs

The bias associated with the coarsened IV estimator for category  $k$  can be particularly large when the true CRF is linear. As depicted in Fig. 2b, a linear CRF entails a casual effect of size  $\beta_i = \tau$  at each treatment intensity interval. The probability limit of the coarsened IV estimator is:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_k^{IV} = \beta_k + s\tau, \quad (5)$$

where the coarsening bias in the second term is increasing in  $s$ , the first stage at intensities other than  $k$  relative to the first stage at  $k$ . In particular, more than one half of all compliers must reach intensity  $k$  for the coarsened IV estimator to converge upon an estimate less than double the size of the true effect.<sup>15</sup> This concern also increases with how close the treatment intensity categories are to one another (i.e., increases in the number of categories  $J$ ), because it becomes increasingly implausible that any instrument could only increase the likelihood of receiving a particular intensity, so  $s$  becomes increasingly large.

However, when the CRF is linear and  $T_i$  is observable, the LAPTE estimator is highly desirable. From inspection of equation (3), it is easy to see that when the true effect is  $\tau$  at each interval,  $\tau$  is recovered exactly by  $\hat{\beta}_{LAPTE}^{IV}$  as the sample becomes large. Moreover, Proposition 3 shows that even when some categories of  $T_i$  are unobserved, the LAPTE still consistently estimates  $\tau$ . Except in the rare instance where  $p_t = 0$ , for all  $t \neq k$ , the LAPTE strictly improves upon the inconsistent coarsened IV estimator. Although this is not surprising, this fact is important because many multi-valued treatments such as education are believed to exert relatively linear effects. Furthermore, similar advantages apply when the CRF is non-linear.

### 2.4.3 Non-linear CRFs

In practice, many CRFs likely reflect more complex non-linear relationship than a single jump at a given intensity. For example, each additional unit of schooling increases subsequent income, but certain thresholds such as completing high school or university may have particularly large effects. Figure 2c illustrates such a case, although innumerable other non-linear CRFs can be imagined. When are the coarsened and LAPTE estimators appropriate approximations to non-linear CRFs?

As demonstrated above, the coarsened IV estimate will generally be upwardly biased whenever the CRF is not discontinuous at a single coarsened treatment intensity. Conversely, the LAPTE remains a consistent estimate of the complier-weighted average across intervals regardless of the CRF's functional form. Therefore, even when the causal effect varies substantially across intensities, the complier-weighted LAPTE is informative about the average effect among compliers in the population that the sample represents, even though it may not provide the non-linear quantity the researcher would ideally estimate. Furthermore, in the limit, the LAPTE always yields a coefficient at least as small in magnitude as the coarsened estimator, and may thus be regarded as a more conservative approach (Angrist and Imbens 1995). Given the extent of upward bias that coarsening can introduce, and the difficulty of efficiently implementing non-linear IV estimates, this trade-off may be preferable in many applications.

<sup>15</sup>Note that  $s = \sum_{t \neq k} p_t / p_k = (\bar{p} - p_k) / p_k < 1$ , only when  $p_k > \bar{p} / 2$ , where  $\bar{p} \equiv \sum_{t=2}^J p_t$ .

However, in some cases, the ITT effect of the instrument may be the only informative estimate. When the instrument affects a wide range of intensities and the CRF significantly departs from linearity, the coarsened IV estimator may be substantially biased, and plausible bounds may be hard to establish. Moreover, an estimate of the LAPTE provides misleading estimates of the treatment's effect when the instrument affects many intensities of a non-monotonic CRF. In an extreme example, a null LAPTE estimate could mask the impact of a treatment where the LATE changes sign across intensities. In contrast, the ITT—estimated by the reduced form—captures the overall impact of the instrument in the population. The ITT does not make any claims about the endogenous treatment's effect, but its interpretation does not rely on assumptions about the CRF. Moreover, as in the case of a school leaving age reform, the ITT is often the policy-relevant quantity of interest. In such a context, and given that policy-makers often weight different types of compliers differently, researchers should report estimates of all the  $p_i$ s alongside the ITT. Although this cannot identify the LATE in different parts of the CRF, it may help readers to appropriately scale the ITT.

## 2.5 Implications for Applied Research

The preceding analysis shows that coarsening an endogenous treatment can substantially upwardly bias IV estimates, except in the special cases where the instrument only induces subjects to reach the intensity where the treatment is coarsened or the CRF only registers an effect at a single and correctly identified intensity. Therefore, as the simulation analysis in the Online Appendix illustrates, estimating the LAPTE is more appropriate for most CRFs. Furthermore, even when the CRF is not linear, treating the endogenous treatment as linear has three important advantages: the LAPTE (1) does not suffer from coarsening bias, and always provides a consistent causal estimate; (2) does not rely on researchers correctly specifying the intensity where the effect is large; and (3) is often robust when the researcher cannot observe all treatment intensities. Nevertheless, the easily-interpretable and often policy-relevant ITT estimate of the instrument's effect may become increasingly appealing when researchers are faced with a non-linear CRF for which average effects across intensities are difficult to interpret.

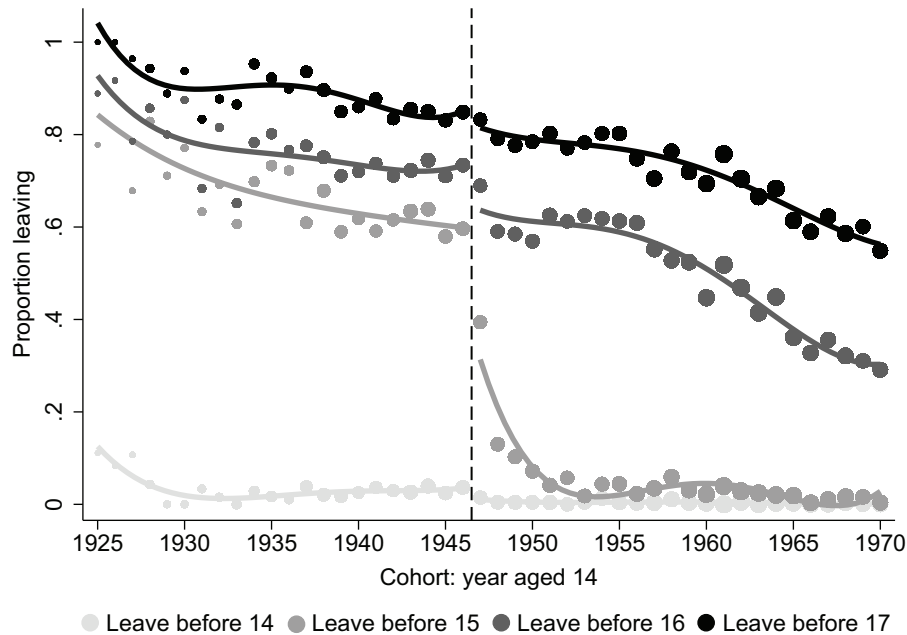
In general, researchers seeking to implement an IV analysis must rely on their theoretical intuitions and descriptive data—including the reduced form relationship, separate first-stage regressions, and the (endogenous) OLS relationship—to determine the appropriate specification in the standard case where only a single instrument is available. Under special circumstances where multiple instruments are available, I show below that a sharper empirical assessment is possible. This will allow me to quantify the coarsening bias associated with instrumenting for a dichotomous measure of schooling.

## 3 Observational Application: High School Education's Effect on Voting Conservative in Great Britain

Building on a recent study by Marshall (forthcoming), this section illustrates the risk of coarsening bias in an applied setting. Marshall (forthcoming) uses Britain's 1947 school leaving reform to instrument for measures of schooling, and finds that each additional year of late high school significantly increases the probability that an individual votes for the right-wing Conservative party later in life. Figure 3 shows that the reform, which increased the school leaving age in Britain from 14 to 15 years for students who reached the age of 14 years after April 1, 1947, dramatically increased the proportion of students staying in school until age 15 years. Importantly, the reform also increased the proportion leaving at age 16 years, and thus completing high school, but did not impact higher levels of education.<sup>16</sup>

The study's findings are not vulnerable to coarsening bias, because it focuses on the LAPTE. However, several particular features of the study's setting enable me to quantify the coarsening bias

<sup>16</sup>Both the reform and research design are discussed in detail in Marshall (forthcoming).



**Fig. 3** 1947 compulsory schooling reform and student leaving age by cohort.

*Notes:* Data are from the British Election Survey. Curves represent fourth-order polynomial fits. Gray dots are birth-year cohort averages, and their size reflects their weight in the sample.

that would arise if the study instead instrumented for an indicator for completing high school. By differentiating the effects of coarsening from other factors that could explain large IV estimates, the objective of this section is to demonstrate that coarsening can substantially bias IV estimates.

### 3.1 Estimation: Coarsening and the LAPTE

I use British Election Survey data from 1979 to 2010, and follow the research design described in Marshall (forthcoming).<sup>17</sup> The outcome is an indicator for the 34% of respondents that reported voting for the Conservative party at the most recent election. The instrument is an indicator for the 49% of respondents that turned 14 in 1947 or later, and were thus affected by the higher school leaving age. Departing from the original study, schooling—the endogenous treatment variable in this application—is measured *both* as a linear intensity and a coarsened indicator. The linear intensity is the number of years of schooling. The coarsened treatment is an indicator for the 64% of respondents that completed high school. Since the purpose of this application is to analyze the decision to coarsen, not assess the results substantively, operationalization is described in the Online Appendix.

To estimate the effect of late high school education on voting Conservative, Marshall (forthcoming) uses Britain's 1947 leaving age reform as an instrument for schooling in the context of a "fuzzy" regression discontinuity (RD) design. In particular, whether an individual was affected by the reform is a discontinuous function of their birth year cohort. However, since the reform could not force every student to remain in school for (at least) an additional year, the 1947 reform is used as an instrument that discontinuously increases the probability that a student received greater schooling. Figure 3 indicates a strong and monotonic first stage, while Marshall (forthcoming) provides evidence that Conservative voting is continuous across cohorts in all covariates other than

<sup>17</sup>The sample differs slightly from Marshall (forthcoming) because I focus on the eight elections from 1979 to 2010 for which data on the outcome, instrument, and treatment measures are available, and restrict attention to working age respondents (aged below 70 years).

**Table 1** Estimates of schooling's effect on voting Conservative

	<i>Years of schooling</i> LLR (1)	<i>Completed high school</i> LLR (2)	<i>Vote Con.</i> LLR IV (3)	<i>Vote Con.</i> LLR IV (4)
Post 1947 reform	0.614*** (0.135)	0.150*** (0.030)		
Years of schooling			0.113** (0.054)	
Completed high school				0.464** (0.209)
Observations	4820	4820	4820	4820
First-stage <i>F</i> -statistic	20.7	24.8	20.7	24.8

*Notes:* Specifications (1) and (2) present the first-stage estimates where years of schooling and completing high school are, respectively, the endogenous treatment variable. Specifications (3) and (4) are, respectively, the IV estimates for years of schooling and completing high school. All specifications are estimated using LLR with a triangular kernel and the Imbens and Kalyanaraman (2012) optimal bandwidth of 11.514. Robust standard errors are given in parentheses. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

the school leaving age at the reform discontinuity (implying assumption A2 above) and that the reform only affected Conservative voting by increasing years of schooling.<sup>18</sup>

The design entails estimating the following structural equation for individuals  $i$  from cohort  $c$ :

$$\text{Vote Conservative}_{ic} = \beta \text{Schooling}_{ic} + f(\text{Birth year}_c) + \varepsilon_{ic}, \quad (6)$$

where  $f$  is a flexible function of birth year cohort. Specifically, following Marshall (forthcoming), I estimate local linear regressions (LLRs) using observations within the Imbens and Kalyanaraman (2012) optimal bandwidth (of 11.514 cohorts) and weight using a triangular kernel. The corresponding first stage is given by:

$$\text{Schooling}_{ic} = \alpha \text{Post 1947 reform}_c + f(\text{Birth year}_c) + \varepsilon_{ic}. \quad (7)$$

Conditioning on  $f(\text{Birth year}_c)$ , the combination of equations (6) and (7) approximately estimate equations (3) and (1), depending upon the measure of  $\text{Schooling}_{ic}$ . Instrumenting for years of schooling corresponds to the LAPTE in equation (3), while instrumenting for the indicator for completing high school corresponds with the coarsened IV estimator in equation (1).

### 3.2 Illustrating Coarsening Bias

The first-stage estimates in Table 1 verify that the 1947 reform instrument substantially increased schooling. Column (1) shows that, on average, the reform increased years of schooling by 0.61 years. The large  $F$ -statistic confirms the strong first stage. Column (2) shows that—in addition to keeping students in school until 15 years—the reform also increased the probability of completing high school by 15 percentage points. This increase is also highly statistically significant ( $F$ -statistic of 24.8), and thus confirms that any bias in the IV estimates does not simply reflect weak instruments (Staiger and Stock 1997).

I now examine the IV estimates of schooling's effect on voting behavior. I first instrument for years of schooling to estimate the LAPTE. Reinforcing Marshall's (forthcoming) findings, the estimate of the LAPTE in column (3) shows that each additional year of late high school significantly increases a complier's probability of voting Conservative in later life by 11 percentage points.

To evaluate the impact of coarsening the treatment variable on the IV estimates, I turn to the case where schooling is dichotomized as an indicator for completing high school. Column (4) reports that voters induced to complete high school by the reform are fully 46 percentage points more likely to vote Conservative in later life. This statistically significant estimate seems implausibly

<sup>18</sup>The weak exclusion restriction need not hold for the purposes of identifying the effects of coarsening, since the exclusion restriction similarly impacts both approaches to defining the treatment.

large, particularly when considering that compliers represent a relatively large share of the student population (so are unlikely to be an anomalous group experiencing unusually large effects), and given that the average effect of an additional year's schooling is four times smaller.

Furthermore, the conditions under which coarsening bias is large—highlighted in equation (2)—appear to be satisfied. First, the reform predominantly kept students in school until age 15 years, but did not compel most students to complete high school (i.e., large  $p_t/p_k$ ). Second, there are reasons to believe that an additional year of late high school, without completing high school, imparts skills that ultimately affect political preferences (i.e.,  $\beta_t \neq 0, t \neq k$ ). Third, since both IV estimates rely on the weak exclusion restriction, differences cannot be explained by the 1947 reform affecting vote choice through channels other than schooling. There are thus good reasons to believe that the estimate for completing high school suffers from substantial coarsening bias.

Nevertheless, although such large effects of completing high school are surprising, they are not completely implausible. The huge effect of completing high school could perhaps be squared with the smaller average effect of an additional year of schooling if the effect of staying until 15 years is actually zero. To confirm that the large estimate actually reflects coarsening bias, I also exploit Britain's 1972 reform raising the school leaving age from 15 to 16 years to separate the effects of the penultimate and final year of high school.<sup>19</sup> As the Online Appendix shows, the 1972 reform also significantly increased the proportion of students leaving school at age 16 years without affecting further education. Consequently, I can use both reforms as instruments to identify the effects of both the penultimate year of high school and completing high school.<sup>20</sup> Given that the availability of two instruments that only affect two levels of the treatment is uncommon, this application represents a rare opportunity to demonstrate and quantify coarsening bias.

The results, described in detail in the Online Appendix to save space, show that completing the penultimate year of high school increases the probability of voting Conservative later in life by 10 percentage points, while completing high school similarly increases Conservative voting by a further 17 percentage points. This implies that the effect of additional late high school on Conservative voting is relatively linear. These estimates are consistent with the LAPTE estimate of 11 percentage points for each additional year in column (3) of Table 1. However, by failing to return the coarsened IV estimate for completing high school of 0.46, the results demonstrate that the strong exclusion restriction fails to hold. In fact, coarsening is responsible for upwardly biasing the IV estimate by a factor of nearly three.

#### 4 Conclusion

This article shows that coarsening bias should be a major concern for applied researchers using IV methods. Coarsening bias arises where a treatment with multiple intensities is transformed into a binary indicator or short scale. Except in special cases that are likely to be rare in empirical applications, such coarsening subtly violates the exclusion restriction requiring that the instrument only affects the outcome through the measured treatment. Although IV methods can address classical measurement error in a treatment variable, coarsening bias shows that non-classical measurement error can dramatically inflate IV estimates.

As demonstrated in the case of high school's long-run effects on vote choice in Britain, the upward bias can be considerable. I show that using a binary indicator for completing high school overestimated the causal effect of the final year of high school on voting Conservative by nearly three times. The extent of coarsening bias clearly has important implications for the interpretation of academic findings, but could also substantially mislead policy-makers seeking to choose the most efficacious policy. As IV methods become a standard part of a political scientist's methodological toolkit, it is thus essential that researchers become aware of the risks of coarsening bias.

<sup>19</sup>Using years of schooling also represent a coarsening of days. However, the estimates below do not suggest a bias when estimating the penultimate and ultimate years of high school separately.

<sup>20</sup>This is because  $p_t \approx 0$  for all  $t$  other than leaving school at 15 years and leaving school at 16 years.

Coarsening bias may be common in applied research. Coarsening an endogenous treatment variable may appear to be theoretically appealing or easier to interpret, to avoid strong linearity assumptions, or be necessitated by data unavailability. However, in general, only when the treatment's effects are concentrated at a certain level of the treatment *and* the researcher is able to pinpoint the specific threshold where this causal effect occurs can coarsening consistently estimate the desired causal quantity. In most cases, it is better to avoid invoking the strong exclusion restriction required to coarsen an endogenous treatment, and instead consistently estimate the LAPTE by measuring the treatment intensity using a linear treatment variable. Unlike coarsening bias, which arises when the true CRF is not discontinuous, this approach provides consistent causal estimates even when the underlying CRF is not linear. Moreover, because this approach is also robust when some categories cannot be observed, it provides an effective solution in the frequent case that a good measure of the treatment is not available. Nevertheless, where the true CRF is highly non-linear, this estimate may be misleading, and researchers are instead advised to report the ITT effect and the first stages at treatment intensity.

*Conflict of interest statement.* None declared.

## References

- Abadie, Alberto. 2003. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 113(2):231–63.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91(5):1369–401.
- 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(434):444–55.
- Angrist, Joshua D., and Jörn-Steffan Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Büthe, Tim, and Helen V. Milner. 2008. The politics of foreign direct investment into developing countries: Increasing FDI through international trade agreements? *American Journal of Political Science* 52(4):741–62.
- Dunning, Thad. 2008. Model specification in instrumental-variables regression. *Political Analysis* 16(3):290–302.
- Gerber, Alan. 1998. Estimating the effect of campaign spending on senate election outcomes using instrumental variables. *American Political Science Review* 92(2):401–11.
- Gerber, Alan S., and Donald P. Green. 2000. The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review* 94(3):653–63.
- Gerber, Alan S., Gregory A. Huber, and Ebonya Washington. 2010. Party affiliation, partisanship, and political beliefs: A field experiment. *American Political Science Review* 104(4):720–44.
- Imbens, Guido W., and Karthik Kalyanaraman. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3):933–59.
- Kane, Thomas J., Cecilia Elena Rouse, and Douglas Staiger. 1999. *Estimating returns to schooling when schooling is misreported*. NBER working paper 7235.
- Marshall, John. 2016. *Replication Data for: Coarsening bias: How coarse treatment measurement upwardly biases instrumental variable estimates*, Harvard Dataverse, V1. <http://dx.doi.org/10.7910/DVN/J7HUX3>.
- . Forthcoming. Education and voting Conservative: Evidence from a major schooling reform in Great Britain. *Journal of Politics*.
- Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos. 2004. Does education improve citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics* 88:1667–95.
- Newey, Whitney K., and James L. Powell. 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71(5):1565–78.
- Pierskalla, Jan H., and Florian M. Hollenbach. 2013. Technology and collective action: The effect of cell phone coverage on political violence in Africa. *American Political Science Review* 107(2):207–24.
- Sovey, Allison J., and Donald P. Green. 2011. Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science* 55(1):188–200.
- Staiger, Douglas, and James H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65(3):557–86.