When is TSLS *Actually* LATE?[*]

Christine Blandhol[†] John Bonney[‡]

Magne Mogstad[§] Alexander Torgovitsky[¶]

January 14, 2025

Abstract

Linear instrumental variable estimators, such as two-stage least squares (TSLS), are commonly interpreted as estimating non-negatively weighted averages of causal effects, referred to as local average treatment effects (LATEs). We examine whether the LATE interpretation actually applies to the types of TSLS specifications that are used in practice. We show that if the specification includes covariates—which most empirical work does—then the LATE interpretation does not apply in general. Instead, the TSLS estimator will, in general, reflect treatment effects for both compliers *and* always/never-takers, and some treatment effects for the always/never-takers will *necessarily* be negatively weighted. We show that the only specifications that have a LATE interpretation are "saturated" specifications that control for covariates nonparametrically, implying that such specifications are both sufficient and *necessary* for TSLS to have a LATE interpretation, at least without additional parametric assumptions. This result is concerning because, as we document, empirical researchers almost never control for covariates nonparametrically, and rarely discuss or justify parametric specifications of covariates. We apply our results to thirteen empirical studies and find strong evidence that the LATE interpretation of TSLS is far from accurate for the types of specifications actually used in practice. We offer concrete recommendations for practice motivated by our theoretical and empirical results.

[*]We thank Richard Boylan, Deniz Dutz, Sergio Firpo, Mikkel Gandil, Tymon Słoczyński, Chris Walters, and Thomas Wiemann for helpful comments.

[†]Department of Economics, Princeton University.

[‡]Department of Economics, Stanford University. Research supported by National Science Foundation Graduate Research Fellowship under grant DGE-1656518.

[§]Kenneth C. Griffin Department of Economics, University of Chicago; Statistics Norway; NBER.

[¶]Kenneth C. Griffin Department of Economics, University of Chicago. Research supported by National Science Foundation grant SES-1846832.

1 Introduction

Instrumental variable (IV) strategies are widely used for causal inference in economics, political science, sociology, epidemiology, and other fields. Since the work of Imbens and Angrist (1994), it has been increasingly common to interpret linear IV estimators as estimating a local average treatment effect (LATE), or at least a non-negatively weighted average of LATEs.

The LATE interpretation is most commonly derived for simplified IV specifications that do not include covariates. We examine whether the LATE interpretation extends to the types of linear IV specifications that are used in practice. We show that if the IV specification includes covariates—which most empirical work does—then the LATE interpretation does not apply in general. Instead, the linear IV estimand with covariates is generally composed of treatment effects for both compliers and always-takers, and some always-taker treatment effects are necessarily negatively weighted.

Our finding challenges the claim by Angrist and Pischke (2009, pg. 173) that

> *2SLS with covariates produces an average of covariate-specific LATEs... These results provide a simple casual* [typo in original] *interpretation for 2SLS in most empirically relevant settings.*

Their assertion is based on a "saturated" two stage least squares (TSLS) specification that controls for covariates nonparametrically, described by Angrist and Pischke (2009) as the "saturate and weight approach" (Theorem 4.5.1; originally Theorem 3 in Angrist and Imbens, 1995). Our results show that this type of saturated specification is not only sufficient for TSLS with covariates to be interpretable as an average of covariate-specific LATEs, it is also *necessary*, at least without additional parametric assumptions.

In Section 2, we report the results of a survey on the specification of linear IV estimators in published empirical papers in economics. Of the 99 papers in our survey that use a linear IV estimator with covariates, we found only five papers that used a saturated specification at least once and only a single paper that exclusively used saturated specifications. The implication of our results for the 98 other papers is that they may not be estimating an average of covariate-specific LATEs. In fact, they may be estimating a quantity that doesn't even satisfy the minimal requirement of being a non-negatively weighted average of subgroup-specific treatment effects, a property we describe as weakly causal.

Section 2 also contains an exposition of our main findings in the special case of a binary treatment and binary instrument. This case exposes the central intuition: if the covariates are not specified flexibly, then the TSLS estimand depends not only on treatment *effects*, but also on potential outcome *levels*. We call this phenomenon level dependence. Because the TSLS estimand is generally level dependent, it does not necessarily have a unique decomposition into a weighted average of subgroup treatment effects. Consequently, analyzing whether the TSLS estimand is weakly causal is more complicated than simply checking for non-negative weights.[1]

In Section 3, we tackle this challenge by providing a conceptual definition of a weakly causal estimand that is separated from the form that the estimand takes. We then provide sufficient and necessary conditions for an estimand to be weakly causal. The characterization has two components. First, a weakly causal estimand cannot be level dependent. Second, a weakly causal estimand should not apply negative weight to the treatment effects for any subgroup.

In Section 4, we specialize this definition to TSLS estimands. We show that a *necessary* condition for the TSLS estimand to be weakly causal is that the TSLS specification has rich covariates, meaning that it exactly reproduces the conditional mean of the instrument. Specifications that are saturated in covariates, such as the Angrist and Pischke (2009) "saturate and weight" specification, will always have rich covariates. But a non-saturated TSLS specification only has rich covariates if an implicit parametric functional form assumption happens to be correct. Saturated specifications can be extremely data hungry, which may explain why they were so seldom used in our survey of empirical papers.

Kolesár (2013) provided the most general sufficient conditions for the TSLS estimand to be equal to a non-negatively weighted average of LATEs. His conditions maintain rich covariates as an *assumption*. Our results show that rich covariates is also *necessary* for the TSLS estimand to have even a weakly causal interpretation, let alone an interpretation as a non-negatively weighted average of LATEs.

The implication of our results is that the Angrist and Pischke (2009) interpretation of TSLS as a non-negatively weighted average of LATEs is fragile. In particular, it depends on rich covariates, which is an implicit parametric functional form assumption that appears to always be left unstated in empirical work. Although our survey turned up only a single paper that used a TSLS specification guaranteed to satisfy rich covariates, we found numerous papers that nevertheless invoked the widespread LATE interpretation. Our results draw this interpretation into question.

In Section 5, we consider alternatives to TSLS. One alternative is to change the TSLS specification to be saturated. However, as our empirical survey suggested, and as our simulations confirm, this is often impractical due to the large number of regressors

[1] In this sense our results and analysis are quite different from the recent literature on two-way fixed effects models (e.g. Goodman-Bacon, 2021; Sun and Abraham, 2021), which point out interpretation problems that arise in event studies if there are heterogeneous treatment effects due to observables (in particular, cohorts). When analyzed without covariates, these estimands are not level dependent, but may have negative weights. A consequence is that the problems we point to remain even with constant treatment effects (Section 4.5), unlike in the two-way fixed effects literature.

produced by saturating. An alternative is to use double/debiased machine learning (Chernozhukov et al., 2018, "DDML") to estimate a partially linear IV (PLIV) modification of TSLS that controls for covariates in an additive but nonparametric way. This frees the researcher of the need to choose a parameterization of the covariates, but comes at a computational cost. It also estimates a quantity that, while weakly causal, might still be hard to interpret.

When the instrument is binary, a related and potentially more attractive alternative is to estimate an unconditional average causal response (ACR), which reduces to an unconditional LATE when the treatment is also binary. This can be done either nonparametrically with DDML or semi-parametrically using instrument propensity score weighting (e.g. Tan, 2006; Uysal, 2011; Słoczyński et al., 2024). A third potential alternative for the binary instrument, binary treatment case is Abadie's (2003) κ weighting approach. The implementation of κ weighting requires explicitly parameterizing the conditional mean of the instrument given the covariates, the same object that we show needs to be implicitly assumed to be correctly specified for TSLS to be weakly causal. However, we show that κ weighting too will only be weakly causal if rich covariates is satisfied, the same necessary condition as for TSLS.

In Section 6, we compare the TSLS estimator with these alternatives in thirteen empirical papers. We find strong evidence that rich covariates is often not satisfied in practice. DDML PLIV estimates can be nearly as different from TSLS as TSLS is different from its comparable OLS estimate. The Ramsey (1969) RESET test tends to do a good job detecting when failures of rich covariates lead to sizable differences between TSLS and a DDML PLIV estimate. DDML PLIV estimates can still be dramatically different from DDML or semi-parametric estimates of the unconditional ACR/LATE.

In Section 7, we provide some concluding remarks and recommendations for practice, all of which can be implemented in Stata or `R` with mature software packages. These recommendations show that it is still possible to estimate an unconditional ACR/LATE or a weighted average of conditional LATEs in the presence of covariates. But not with the types of TSLS specifications that are currently being used in practice.

Słoczyński (2020, 2024) has recently made a different critique of the interpretation of TSLS estimators. He maintains rich covariates as an assumption and shows that the TSLS estimand can still fail to be weakly causal if the direction of monotonicity varies with covariates but the TSLS specification does not include instrument-covariate interactions in the first stage. In contrast, our analysis focuses on the necessity of the rich covariates condition under a stronger, unconditional monotonicity condition. Słoczyński (2024) also makes the important theoretical point that even when rich covariates is satisfied, the resulting linear IV estimand may be quite different from the type of unconditional LATE that practitioners might expect. We do not discuss any theory about this point, although we do illustrate it our empirical applications.

Rich covariates remains necessary under the weaker monotonicity condition considered by Słoczyński (2020, 2024). Taken together, our paper and Słoczyński (2024) show that two conditions are necessary for the TSLS estimand to be interpretable as a non-negatively weighted average of LATEs: (i) rich covariates, and (ii) a first stage equation flexible enough to capture changes in the direction of monotonicity across covariate values. The necessity of these conditions provides a definitive answer to the question: "When is TSLS *actually* LATE?" That answer: probably not often.

2 Overview

In this section, we demonstrate our main results in the special case of a binary treatment and a binary instrument.

2.1 Linear IV with covariates is not LATE

Let $T \in \{0,1\}$ be a binary treatment and $Z \in \{0,1\}$ be a binary instrument. The outcome is Y with potential outcomes $Y(0)$ and $Y(1)$ related via $Y = (1-T)Y(0) + TY(1)$. Potential treatment states are $T(0)$ and $T(1)$ with $T = (1-Z)T(0) + ZT(1)$. The vector of covariates is X.

Assume that Z is conditionally exogenous in the sense of being independent of $(Y(0), Y(1), T(0), T(1))$ conditional on X. Suppose that the Imbens and Angrist (1994) monotonicity condition holds so that $\mathbb{P}[T(1) \geq T(0)] = 1$. The monotonicity condition implies that the group variable $G \equiv (T(0), T(1))$ can take three values with non-zero probability: $G = (0,0) \equiv \textsc{nt}$ are the never-takers, $G = (0,1) \equiv \textsc{cp}$ are the compliers, and $G = (1,1) \equiv \textsc{at}$ are the always-takers.

Consider a linear IV regression with outcome variable Y, endogenous variable T, excluded instrument Z, and a vector of control variables X that includes a constant. The Frisch-Waugh-Lovell Theorem can be used to show that the IV estimand (the population coefficient on T) is given by

$$\beta_{\text{iv}} = \frac{\mathbb{E}[Y\tilde{Z}]}{\mathbb{E}[T\tilde{Z}]}, \quad \text{where} \quad \tilde{Z} \equiv Z - \mathbb{L}[Z|X] \tag{1}$$

are the residuals from a regression of Z on X, and

$$\mathbb{L}[Z|X] \equiv X' \, \mathbb{E}[XX']^{-1} \, \mathbb{E}[XZ]$$

are the population fitted values from regressing (linearly projecting) Z onto X. The IV estimand, β_{iv}, is often interpreted as reflecting a non-negatively weighted average of treatment effects for only the compliers. The following proposition shows that this is not true in general.

Proposition 1. Suppose that $\mathbb{E}[Y(t)|X] = \eta_t' X$ for some unknown parameters η_t, $t = 0, 1$.[2] Let $\Delta(\textsc{cp}, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = \textsc{cp}, X = x]$ and $\Delta(\textsc{at}, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = \textsc{at}, X = x]$ denote the conditional average treatment effects for the compliers and always-takers, respectively. Then

$$\beta_{\text{iv}} = \mathbb{E}[\omega(\textsc{cp}, X)\Delta(\textsc{cp}, X)] + \mathbb{E}[\omega(\textsc{at}, X)\Delta(\textsc{at}, X)], \tag{2}$$

$$\text{where} \quad \omega(\textsc{cp}, X) \equiv \mathbb{E}[Z|X]\,(1 - \mathbb{L}[Z|X])\,\mathbb{P}[G = \textsc{cp}|X]\,\mathbb{E}[\tilde{Z}T]^{-1}$$

$$\text{and} \quad \omega(\textsc{at}, X) \equiv \mathbb{E}[\tilde{Z}|X]\,\mathbb{P}[G = \textsc{at}|X]\,\mathbb{E}[\tilde{Z}T]^{-1}.$$

If $\mathbb{E}[\tilde{Z}T] > 0$, then the complier weights $\omega(\textsc{cp}, X)$ are negative if and only if $\mathbb{L}[Z|X] > 1$. The always-taker weights $\omega(\textsc{at}, X)$ are strictly negative with positive probability unless $\mathbb{E}[\tilde{Z}|X] = 0$ deterministically.

Proposition 1 shows that, in general, β_{iv} reflects not only the compliers, but also the always-takers. The monotonicity condition implies that the first stage coefficient is positive, so $\mathbb{E}[\tilde{Z}T] > 0$. The weights on the always-takers therefore have the same sign as the random variable $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X]$. Because X contains a constant, $\mathbb{E}[\tilde{Z}] = \mathbb{E}[\mathbb{E}[\tilde{Z}|X]] = 0$, implying that $\mathbb{E}[\tilde{Z}|X]$ is either always equal to zero, or else it has positive probability of taking both positive and negative values. As a consequence, whenever $\mathbb{L}[Z|X] \neq \mathbb{E}[Z|X]$, the IV estimand incorporates negatively weighted treatment effects for some groups, which means that it fails to satisfy even a minimal condition for "being causal."

This reasoning shows that in order for the LATE interpretation to hold, it is necessary that $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$, a condition we call rich covariates. Specifications that are saturated in covariates, such as "saturate and weight" (Angrist and Pischke, 2009), have rich covariates. If Z and X are independent, as can be the case in some controlled and natural experiments, then any specification with a constant will have rich covariates.[3] Outside these two cases, having rich covariates is a parametric assumption. If it fails, then the IV estimand β_{iv} reflects not just compliers, but also negatively weighted always-takers.

There is no reason to expect, a priori, that the weights on the always-taker treatment effects in (2) will be small in magnitude. In many applications, the proportion of always-takers, $\mathbb{P}[G = \textsc{at}|X]$, can be expected to be considerably larger than the proportion of

[2] This additional linearity assumption is made in order to simplify the weights. Removing the assumption only *amplifies* the negative interpretation issues exposed by Proposition 1. Our general results in Section 4 do not maintain this assumption.

[3] In Appendix SA.1, we discuss the case in which Z is randomly assigned conditional on a subset of X, as might occur in a stratified experiment.

compliers, $\mathbb{P}[G = \text{CP}|X]$. As a consequence, even negative values of $\mathbb{E}[\tilde{Z}|X]$ that are small in magnitude can produce large negative weights on the always-taker treatment effects.

Decomposition (2) is not the only one possible. Instead of interpreting β_{iv} as a weighted average of compliers and always-takers, one can interpret it as a weighted average of compliers and never-takers, or of all three groups, as shown in the next proposition.

Proposition 2. Suppose that $\mathbb{E}[Y(t)|X] = \eta_t' X$ for some (unknown) parameters η_t, and both $t = 0, 1$. Let $\Delta(\text{NT}, x) \equiv \mathbb{E}[Y(1)-Y(0)|G = \text{NT}, X = x]$ denote the conditional average treatment effect for the never-takers. Then for any real number ϵ,

$$
\begin{aligned}
\beta_{\text{iv}} &= \mathbb{E}[\omega_\epsilon(\text{CP}, X)\Delta(\text{CP}, X)] + \mathbb{E}[\omega_\epsilon(\text{AT}, X)\Delta(\text{AT}, X)] + \mathbb{E}[\omega_\epsilon(\text{NT}, X)\Delta(\text{NT}, X)], \\
\text{where} \quad & \omega_\epsilon(\text{CP}, X) \equiv \Big(\epsilon\, \mathbb{E}[\tilde{Z}|X] + \mathbb{L}[Z|X](1 - \mathbb{E}[Z|X])\Big)\, \mathbb{P}[G = \text{CP}|X]\, \mathbb{E}[\tilde{Z}T]^{-1}, \\
& \omega_\epsilon(\text{AT}, X) \equiv \epsilon\, \mathbb{E}[\tilde{Z}|X]\, \mathbb{P}[G = \text{AT}|X]\, \mathbb{E}[\tilde{Z}T]^{-1}, \\
\text{and} \quad & \omega_\epsilon(\text{NT}, X) \equiv (\epsilon - 1)\, \mathbb{E}[\tilde{Z}|X]\, \mathbb{P}[G = \text{NT}|X]\, \mathbb{E}[\tilde{Z}T]^{-1}.
\end{aligned}
$$

Each choice of ϵ in Proposition 2 provides a different interpretation of β_{iv}, with Proposition 1 corresponding to $\epsilon = 1$. However, unless rich covariates holds, so that $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X] = 0$, any choice of ϵ still leads to an interpretation that involves either the always-takers or the never-takers, or both, with negative weights for some values of X, as well as potentially negative weights for the compliers. Only in specifications with rich covariates is β_{iv} a non-negatively weighted average among compliers alone.[4]

2.2 Intuition

The intuition behind Propositions 1 and 2 can be seen by writing the numerator of β_{iv} as

$$
\mathbb{E}[Y\tilde{Z}] = \mathbb{E}\left[\mathbb{E}\left[Y\tilde{Z}|X\right]\right] = \mathbb{E}\Big[\overbrace{\mathbb{C}[Y, Z|X]}^{\text{only contains complier treatment effects}}\Big] + \mathbb{E}\Big[\underbrace{\mathbb{E}[Y|X]}_{\text{contains all three groups}}\mathbb{E}[\tilde{Z}|X]\Big], \tag{3}
$$

where $\mathbb{C}$ denotes covariance. The first term in (3) is the average of the numerator of a nonparametric IV specification that *conditions* on X. The argument in Imbens and

[4]Proposition 2 shows that a causal interpretation can be partially salvaged if there is one-sided non-compliance. For example, if there are no always-takers, so that $\mathbb{P}[G = \text{AT}|X] = 0$, then one can take $\epsilon = 1$, so that β_{iv} is a weighted average among compliers alone. The same is true if there are no never-takers by taking $\epsilon = 0$. The complier weights can still be negative in these cases if $\mathbb{L}[Z|X]$ does not lie in $[0, 1]$, but rich covariates is stronger than necessary to rule this out. However, this conclusion depends on the simplifying linearity assumption that $\mathbb{E}[Y(t)|X] = \eta_t' X$, which we do not maintain in our general results in Section 4.

Angrist (1994) shows that this term is equal to an average of scaled LATEs, which only reflects treatment effects for the compliers. It is the second term of (3) that causes problems. This term reflects the difference between nonparametric conditioning and linear projection.[5]

When covariates are not rich, so that $\mathbb{E}[\tilde{Z}|X] \neq 0$, the second term in (3) generally depends on $\mathbb{E}[Y|X]$, a quantity which is determined not only by compliers, but also by always-takers and never-takers. This creates level dependence in β_{iv} because the always-takers always have $Y = Y(1)$ and the never-takers always have $Y = Y(0)$: β_{iv} depends on the levels of the always-taker and never-taker potential outcomes. As we show in Section 3, level dependent estimands do not have a causal interpretation because the levels can lead β_{iv} to have the "wrong sign."

The expression in Proposition 1 arises from centering the term $\mathbb{E}[Y|X]$ in (3) around $\mathbb{E}[Y(0)|X]$. The simplifying linearity assumption implies that $\mathbb{E}[Y(0)|X] = \eta_0'X$ is uncorrelated with $\mathbb{E}[\tilde{Z}|X]$. Since never-takers always have $Y = Y(0)$, the centering removes the average untreated outcome for the never-takers, leaving only a weighted average of the complier and always-taker treatment effects. Alternatively, we can center around $\mathbb{E}[Y(1)|X] = \eta_1'X$, which leaves a weighted average of the complier and never-taker treatment effects. Both decompositions are equally valid ways to rewrite a single number, β_{iv}, as a weighted average of $\Delta(\textsc{cp}, X)$, $\Delta(\textsc{at}, X)$, and $\Delta(\textsc{nt}, X)$. Taking an ϵ–weighted average of these two decompositions yields the expression in Proposition 2, which creates a family of equally-valid decompositions.

The theory we develop in Section 3 is designed to handle this type of non-uniqueness in decomposition and determine, in a general setting, necessary conditions for the existence of *some* "good" decomposition. For the simplified case considered here, with a binary treatment, a binary instrument, and the linearity assumption $\mathbb{E}[Y(t)|X] = \eta_t'X$, this type of analysis can be done directly, as in Proposition 2. Our analysis of more general TSLS specifications in Section 4 shows that the necessity of rich covariates for a causal interpretation is a conclusion that applies more broadly.

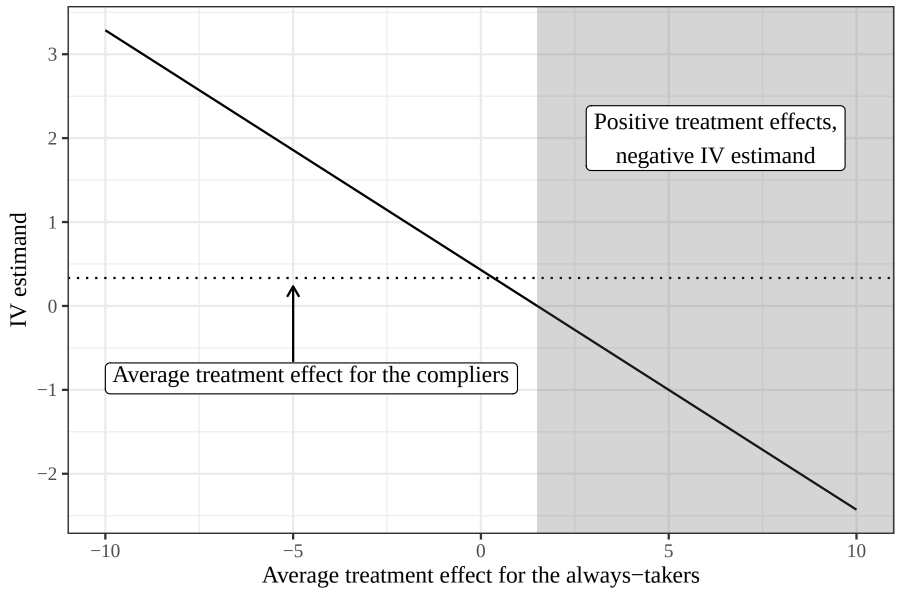
2.3 Numerical illustration

As a simple illustration of these results, suppose that $X \in \{(1, -1), (1, 0), (1, 1)\}$ with equal probability, where the first component corresponds to a constant. Then suppose that

$$\mathbb{E}[Z|X = x] = \mathbb{P}[Z = 1|X = (1, x)] = \begin{cases} 4/5 & \text{if } x \in \{-1, 1\} \\ 2/5 & \text{if } x = 0 \end{cases}.$$

[5] Firpo et al. (2020) make a similar point in the context of balance tests for stratified experiments.

Figure 1: IV with covariates is not LATE



Regressing Z onto X yields the constant regression line:

$$\mathbb{L}[Z|X] = X' \, \mathbb{E}[XX']^{-1} \, \mathbb{E}[XZ] = 2/3,$$

so that $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X] \neq 0$ and is both positive and negative with non-zero probability.

Suppose that the conditional group share probabilities are given by:

$$\begin{aligned} &\text{(never-takers)} && \mathbb{P}[G = \textsc{nt}|X = (1,x)] = 1/3 \\ &\text{(compliers)} && \mathbb{P}[G = \textsc{cp}|X = (1,x)] = 1/6 + |x|/6 \\ &\text{(always-takers)} && \mathbb{P}[G = \textsc{at}|X = (1,x)] = 1/2 - |x|/6. \end{aligned}$$

Simplifying the algebra in Proposition 1 yields

$$\omega(\textsc{cp},(1,x)) = \begin{cases} 12/7, & \text{if } |x| = 1 \\ 3/7, & \text{if } x = 0 \end{cases} \quad \text{and} \quad \omega(\textsc{at},(1,x)) = \begin{cases} 6/7, & \text{if } |x| = 1 \\ -18/7, & \text{if } x = 0 \end{cases}.$$

For simplicity, assume that $Y(0) = 0$, so that treatment effects are determined solely by $Y(1)$, and that $\mathbb{E}[Y(1)|G = \textsc{cp}, X = x] \equiv \mu(\textsc{cp})$ and $\mathbb{E}[Y(1)|G = \textsc{at}, X = x] = \mu(\textsc{at})$ do not depend on x. Then Proposition 1 shows that

$$\beta_{\text{iv}} = \frac{9}{7}\mu(\textsc{cp}) - \frac{2}{7}\mu(\textsc{at}).$$

Table 1: IV papers by journal and type

	(1)	(2)	(3)	(4)
	All papers	Papers using TSLS	Papers using TSLS with covariates	Papers using TSLS with covariates, referring to LATE
American Economic Review	100% 44	95% 42	82% 36	27% 12
Quarterly Journal of Economics	100% 28	93% 26	86% 24	14% 4
Journal of Political Economy	100% 23	91% 21	83% 19	30% 7
Econometrica	100% 15	73% 11	73% 11	27% 4
Review of Economic Studies	100% 12	100% 12	75% 9	25% 3
All	100 % 122	92% 112	81% 99	25% 30

Figure 1 shows the value of β_{iv} as a function of $\mu(\text{AT})$, keeping $\mu(\text{CP}) = 1/3$. If it were true that LATE only reflects the compliers, then we would expect to see a flat line, so that the IV estimand doesn't depend on the treatment effect for the always-takers. Not only is the line not flat, it slopes down. This means that the IV estimand can be negative even when both the compliers and the always-takers have positive treatment effects.

2.4 Survey on IV specifications used in empirical work

Propositions 1 and 2 show that using an IV specification that is saturated in covariates is needed for the LATE interpretation asserted by Angrist and Pischke (2009). To get a sense of how common it is to saturate in covariates, we surveyed the specifications used in the empirical economics literature.

Our sample was constructed by searching the Web of Science Database for articles published between January 2000 and October 2018 containing the words "instrument" or "instrumental variable" in the abstract, title, or topic words. We restricted the search to the following five journals: *Journal of Political Economy*, *American Economic Review*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Econometrica*. In total, 266 articles matched our search criteria.

We restricted our attention to papers that use at least one IV specification in an empirical application. This produced 122 papers; the other 144 papers not included were either methodological papers without an empirical application, or were papers that used the word "instrument" in a different context, such as to describe a policy or financial instrument. Column (1) of Table 1 tabulates the papers used in our survey by the journal in which they were published.

Column (2) shows that over 92% of the papers in our survey use TSLS (including exactly identified linear IV) for at least some of their results. Column (3) counts the subset of the papers in column (2) for which *all* TSLS specifications in the main body of the paper include at least one covariate, or the authors explicitly state the exogeneity assumption for the instrument as conditional on covariates.[6] Comparing columns (2) and (3) shows that using covariates in TSLS is extremely common practice; only 13 out of the 112 papers that use TSLS include any specifications without covariates. Column (4) shows that almost a third of the papers that use TSLS with covariates also explicitly use the phrases "compliers," "local average treatment effect," or "LATE" to describe their results.

In Table 2, we categorize the papers in column (3) of Table 1 by the TSLS specifications they use. Column (2) shows that only 5% of the papers use any specification that is saturated in covariates. These are typically preliminary specifications with only a set of fixed effects. Column (3) shows that every paper uses at least one specification that is *not* saturated in covariates, with only one exception. The one exception is Chamberlain and Imbens (2004). Column (4) shows that those authors also saturate the first stage in both the covariates and the instruments, as prescribed by Angrist and Pischke's (2009) "saturate and weight approach."

2.5 Implications for empirical practice

Avoiding the conclusion of Propositions 1 and 2 requires choosing a specification with rich covariates, that is, one that ensures $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$.

The saturate and weight (SW) specification (Angrist and Pischke, 2009) is saturated in covariates, so has rich covariates. However, it also uses a first stage that is fully saturated in both the covariates *and* the instruments, meaning that the regressors are indicators for all possible instrument-covariate combinations. This results in many excluded variables and potential many instruments bias, which may explain why the SW specification was used by only a single paper in the survey. In fact, that one paper (Chamberlain and Imbens, 2004) is a methodological consideration of many instruments bias.

However, the interactions between covariates and instruments used in the SW specification may not be necessary for the LATE interpretation. Excluded interactions

[6] Another possible justification for including covariates is to improve statistical precision. This motivation was rarely stated explicitly in the papers in our survey. While it is difficult to infer researchers' unstated reasons for choosing particular specifications, it seems unlikely that they would *only* use specifications with covariates if covariates were only being used to improve precision.

Table 2: TSLS papers with covariates by journal and empirical specification

	(1)	(2)	(3)	(4)
		At least one specification		
	Papers using TSLS with covariates	Saturated in covariates	Not saturated in covariates	Saturated in instruments and covariates
American Economic Review	100% 36	0% 0	100% 36	0% 0
Quarterly Journal of Economics	100% 24	4% 1	100% 24	0% 0
Journal of Political Economy	100% 19	16% 3	100% 19	0% 0
Econometrica	100% 11	9% 1	91% 10	9% 1
Review of Economic Studies	100% 9	0% 0	100% 9	0% 0
All	100 % 99	5% 5	99% 98	1% 1

Notes: This table classifies the papers from column (3) of Table 1 by TSLS specification.

were not used in (1) and yet Propositions 1 and 2 show that if covariates are rich, then β_{iv} will be composed of only non-negatively weighted complier effects. The reason is that we assumed that the Angrist and Imbens (1995) monotonicity condition went in the same direction for every covariate group. In contrast, the SW specification is premised on a weaker version of the monotonicity assumption that allows the direction of monotonicity to vary with covariates. Słoczyński (2020, 2024) shows that including the instrument-covariate interaction terms used in SW is necessary when considering this weaker monotonicity condition.

Our results show that flexibly controlling for covariates is important for ensuring that TSLS has a causal interpretation. If a flexible covariate specification cannot be used, then another response is to test the null hypothesis that $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$. The most well-known test is Ramsey's (1969) RESET test (e.g. Wooldridge, 2010, pp. 137–138), which is straightforward to implement in either Stata or `R`. No papers in our survey reported such a test. If Z is binary, then it is also sensible to check that the fitted values $\mathbb{L}[Z|X]$ lie between 0 and 1, which is necessary for $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$. Alternatively, researchers can consider using an estimator other than TSLS. We discuss alternative estimators in Section 5 and apply them in Section 6.

3 Definition and characterization of weakly causal estimands

In this section we define a weak property that an estimand should satisfy in order to "be causal." We do this because, as Proposition 2 showed, if rich covariates fails, then the TSLS estimand might have multiple equally valid decompositions. Alternatively, if the simplifying linearity assumption maintained in Proposition 2 is dropped, the TSLS estimand might not have any decomposition in terms of only treatment effects. These complications motivate a more abstract definition of a weakly causal estimand that is separated from the functional form that the estimand takes. We develop the weakly causal property in the context of a nonparametric IV model using potential outcomes notation (e.g. Angrist et al., 1996) with an ordered treatment and a multivalued instrument. The results generalize the special case of a binary treatment and binary instrument discussed in Section 2.

3.1 The nonparametric instrumental variables model

A discrete, ordered treatment variable T takes values in $\mathcal{T} \equiv \{t_0, t_1, \ldots, t_J\}$, listed in increasing order. We are interested in the causal effects that T has on an outcome variable, Y. We observe a scalar- or vector-valued instrumental variable (IV) Z that takes values in a set $\mathcal{Z} \equiv \{z_0, z_1, \ldots, z_K\}$. The case in Section 2 corresponds to $\mathcal{T} = \{0, 1\}$ and $\mathcal{Z} = \{0, 1\}$. There is a vector of covariates X with support $\mathcal{X}$.

Associated with each level of the IV is a potential treatment choice, $T(z)$. Associated with each level of the treatment is a potential outcome, $Y(t)$, which does not directly depend on the instrument due to the usual exclusion restriction. The potential and actual treatments and outcomes are related through

$$T = \sum_{z \in \mathcal{Z}} \mathbb{1}[Z = z]T(z) \quad \text{and} \quad Y = \sum_{t \in \mathcal{T}} \mathbb{1}[T = t]Y(t).$$

We maintain the following standard nonparametric exogeneity condition throughout our analysis.

Assumption EXO. (Exogeneity) $(\{T(z)\}_{z \in \mathcal{Z}}, \{Y(t)\}_{t \in \mathcal{T}}) \perp\!\!\!\perp Z|X$.

We assume that each of T, Z, and X are discretely distributed with finite support. This is just for mathematical simplicity. Our theoretical results can be extended to allow for T to be a continuous scalar, and both X and Z to be vectors with continuous components. The changes required essentially involve replacing sums with integrals and finite indices with function arguments. We also assume throughout that the expectation of Y exists.

Our analysis uses a partition of individuals into mutually exclusive and exhaustive groups based on their potential treatment choices. Let $G \equiv (T(z_0), T(z_1), \ldots, T(z_K))$

denote an individual's choice group, that is, their configuration of potential treatment choices under each of the instrument values. Let $\mathcal{G}$ denote the values that G can take. In the binary treatment, binary instrument case, G takes values in $\mathcal{G} = \{(0,0),(1,1),(0,1),(1,0)\}$, corresponding to the groups Angrist et al. (1996, Table 1) called the never-takers, always-takers, compliers, and defiers, respectively. Using the group notation, Assumption EXO can be equivalently written as follows.

Assumption EXO. (Exogeneity, group form) $(G, \{Y(t)\}_{t \in \mathcal{T}}) \perp\!\!\!\perp Z | X$.

3.2 Definition of a weakly causal estimand

Consider the *group treatment responses* (GTRs)

$$\mu_j(g,x) \equiv \mathbb{E}[Y(t_j)|G = g, X = x],$$

which are the expected potential outcomes across choice and covariate groups.[7] We collect the GTRs as $\mu \equiv \{\mu_j(g,x) : j = 0,1,\ldots,J, g \in \mathcal{G}, x \in \mathcal{X}\}$, which takes values in $\mathbb{R}^{d_\mu}$. Let β be a quantity whose value depends on μ. We use the following definition as a minimal requirement for β to be interpreted as "causal."

Definition WC. β is **weakly causal** if both of the following statements are true for any μ:

$$\begin{aligned}
&\text{If } \mu_j(g,x) - \mu_{j-1}(g,x) \geq 0 \text{ for all } j \geq 1, \text{ all } g \in \mathcal{G}, \text{ and every } x \in \mathcal{X}, \text{ then } \beta \geq 0. \\
&\text{If } \mu_j(g,x) - \mu_{j-1}(g,x) \leq 0 \text{ for all } j \geq 1, \text{ all } g \in \mathcal{G}, \text{ and every } x \in \mathcal{X}, \text{ then } \beta \leq 0. \quad (4)
\end{aligned}$$

Definition WC is a natural requirement to place on an estimand. The requirement is merely that *if* the causal effect of the treatment has the same sign for every treatment contrast, and every choice and covariate subgroup, then the summary estimand β also has that sign. That is, β is weakly causal if it is not a systematically misleading measure of the sign of the underlying group- and covariate-specific treatment effects.

Definition WC is intended to be an extremely weak criterion. An estimand can be weakly causal and still be completely uninteresting. For example, the trivial estimand $\beta = 0$ is weakly causal. However, it seems unlikely that an estimand that fails to be weakly causal could still reasonably be described as reflecting the causal effect of T on Y, since it may not even have the right sign. As minimal as Definition WC is, we have already seen in Figure 1 that a linear IV estimand can fail to satisfy it, even if the instrument satisfies exclusion and exogeneity (Assumption EXO).

[7] As a minor abuse of notation, we assume that $\mu_j(g,x)$ is well-defined for all (g,x), even if g is not in the support of G given X, so that $\mathbb{P}[G = g, X = x] = 0$. This convention has no impact on our results.

3.3 Weak causality and non-negatively weighted averages

We consider estimands that can be written as

$$\beta = \mathbb{E}[b(T, X, Z)Y] \tag{5}$$

for some function b. For example, β_{iv} in Section 2 satisfies (5) with $b(T, X, Z) = \tilde{Z}/\mathbb{E}[T\tilde{Z}] = (Z - \mathbb{L}[Z|X])/\mathbb{E}[T(Z - \mathbb{L}[Z|X])]$. The following proposition decomposes these estimands into GTRs.

Proposition 3. Suppose that β has form (5), and that Assumption EXO holds. Then

$$\beta = \sum_{g,x} \omega_0(g, x)\mu_0(g, x) + \sum_{g,x}\sum_{j=1}^{J} \omega_j(g, x) \left(\mu_j(g, x) - \mu_{j-1}(g, x)\right), \tag{6}$$

where $\omega_j(g, x) \equiv \mathbb{E}\left[\mathbb{1}[T \geq t_j]b(t_j, x, Z)|G = g, X = x\right] \mathbb{P}[G = g, X = x]$ for all $j \geq 0$.

The next proposition shows that an estimand that is weakly causal can be written as a non-negatively weighted average of subgroup-specific treatment effects. This criterion is widely-used (e.g. Angrist, 1998; Lee, 2008; Angrist and Pischke, 2009; Card et al., 2015; Goodman-Bacon, 2021; Sun and Abraham, 2021; Goldsmith-Pinkham et al., 2024). The proposition also shows that there are two reasons that an estimand can fail to be weakly causal: either it places negative weights on treatment effects or it depends on the levels of the potential outcomes (or both).

Proposition 4. Suppose that β has the form (5) and that Assumption EXO holds. Then β is weakly causal if and only if:

- **(Non-negative weights)** $\omega_j(g, x) \geq 0$ for all $j \geq 1$, and all g and x.
- **(Level independence)** $\omega_0(g, x) = 0$ for all g and x.

If these conditions are satisfied, then

$$\beta = \sum_{g,x}\sum_{j=1}^{J} \omega_j(g, x) \left(\mu_j(g, x) - \mu_{j-1}(g, x)\right) \tag{7}$$

for non-negative weights $\omega_j(g, x) \geq 0$.

Proposition 3 shows that if β has form (5), then β can always be written as (6). Proposition 4 uses that representation to show that if β cannot also be written like (7) with weights that are non-negative, then one of two things must be true: either β only reflects treatment effects, but some of these effects are negatively weighted, or else β reflects not just treatment effects but also the levels of potential outcomes. The first situation violates the non-negative weights requirement, which is naturally necessary for β to be weakly causal (recall Figure 1). The second situation violates the level independence requirement. Level independence is necessary for β to be weakly causal because it prevents the possibility that all treatment effects are positive, even while the levels of the GTRs are such that $\beta < 0$.

4 When is TSLS weakly causal?

In this section we specialize the general results of the previous section to a class of TSLS estimands.

4.1 TSLS specifications and estimands

A TSLS specification is characterized by four components: (i) the outcome variable; (ii) the endogenous variables that are included in the second stage and are regressands in the first stage; (iii) the excluded variables that are excluded from the second stage but are regressors in the first stage; and (iv) the included variables that are regressors in both stages. The nonparametric IV model specifies the outcome variable, Y, but not which combinations of T, Z, and X go in the first and second stages. A TSLS specification produces a TSLS estimator, the probability limit of which is called the TSLS estimand.[8]

We consider TSLS specifications where there is a single endogenous variable, T, a single scalar excluded variable, Z, and a vector of included variables, X. For this case, the TSLS estimand is the same as the linear IV estimand because Z and T have the same dimension. In our working paper (Blandhol et al., 2022), we considered more general TSLS specifications and reached similar conclusions subject to the complications discussed ahead in Section 4.4. In what follows, we reserve the phrase TSLS for specifications with strictly more excluded variables than endogenous variables.

The coefficient on T for the linear IV (née TSLS) estimand with a single endogenous variable and a single excluded variable is given by

$$\beta_{\text{iv}} = \frac{\mathbb{E}[\tilde{Z}Y]}{\mathbb{E}[\tilde{Z}T]} = \mathbb{E}\left[\left(\frac{\tilde{Z}}{\mathbb{E}[\tilde{Z}T]}\right)Y\right]. \tag{8}$$

Proposition 3 shows that β_{iv} can be written as (6) with

$$\omega_j(g,x) = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\mathbb{1}[T \geq t_j]\tilde{Z}|G = g, X = x\right]\mathbb{P}[G = g, X = x]. \tag{9}$$

[8] Our definition of the TSLS estimand presumes the standard asymptotic framework where the number of observations is growing and the dimension of X is fixed. Kolesár (2013) and Evdokimov and Kolesár (2019) consider alternative frameworks that allow for the dimensions of either or both of these vectors to also be growing.

Proposition 4 shows that whether β_{iv} is weakly causal is determined by $\omega_j(g, x)$.

4.2 Main result

Monotonicity conditions are essential for TSLS estimands to have weakly causal interpretations. We maintain the following monotonicity condition.

Assumption MON. (Monotonicity) Label the values of Z in increasing order as $z_0 \leq z_1 \leq \cdots \leq z_K$. Then

$$\mathbb{P}[T(z_0) \leq T(z_1) \leq \cdots \leq T(z_K) | X = x] = 1 \quad \text{for all } x.$$

Assumption MON means that increasing the instrument weakly increases treatment for all individuals. This is a strong form of monotonicity because it operates in the same direction conditional on $X = x$ for all values of x. Results under weaker forms of monotonicity can be found in our working paper (Blandhol et al., 2022).[9]

Our main result is Theorem 1, which uses the following definition.

Definition RC. Let $\mathbb{L}[Z|X = x] \equiv \mathbb{E}[ZX'] \, \mathbb{E}[XX']^{-1}x$ be the population fitted value at $X = x$ from regressing Z onto X. An IV specification has **rich covariates** if $\mathbb{E}[Z|X = x] = \mathbb{L}[Z|X = x]$ for every $x \in \mathcal{X}$.

Theorem 1. Suppose that Assumptions EXO and MON are satisfied. Then β_{iv} is weakly causal if *and only if* the IV specification has rich covariates.

Theorem 1 shows that rich covariates is both sufficient and necessary for the linear IV estimand to have a weakly causal interpretation. The sufficient direction is a special case of Kolesár (2013, Theorem 1), who explicitly maintained rich covariates as an *assumption* (Kolesár, 2013, Assumption L). The necessary direction shown here is novel. It shows that rich covariates is an essential assumption.

As Kolesár (2013, pp. 10–11) notes, there are two important special cases in which an IV specification will have rich covariates. One is when X represents a saturated specification consisting of a vector of indicators for a set of exclusive and mutually exhaustive events. The other is when Z is mean independent of X so that $\mathbb{E}[Z|X = x] = \mathbb{E}[Z]$ is constant in x, which implies that

$$\mathbb{L}[Z|X = x] \equiv \mathbb{E}[ZX'] \, \mathbb{E}[XX']^{-1}x = \mathbb{E}[Z] \, \mathbb{E}[1X'] \, \mathbb{E}[XX']^{-1}x = \mathbb{E}[Z],$$

[9] In particular, Assumption MON corresponds to what we called "ordered strong monotonicity" in Blandhol et al. (2022). See Section 4.4 and Appendix C of Blandhol et al. (2022) for a discussion of why requiring the instrument to be ordered can be important.

because X contains a constant. Outside these two special cases, the claim that an IV specification has rich covariates is an implicit parametric assumption that should be defended.

4.3 Implications for OLS under selection on observables

Theorem 1 also applies to selection on observables by taking $Z = T$, under which Assumption MON is immediately satisfied.

Corollary 1. Suppose that $Z = T$ and that Assumption EXO is satisfied. Let β_{ols} denote the coefficient on T for the OLS estimand generated by the ordinary least squares regression of Y on T and X. Then β_{ols} is weakly causal if and only if $\mathbb{L}[T|X] = \mathbb{E}[T|X]$.

Angrist (1998) proposed implementing a selection on observables strategy using the OLS estimator described in Corollary 1 with a saturated specification of covariates. He described the difference between this regression coefficient and nonparametric matching as "partly cosmetic" (Angrist, 1998, pg. 255). Based on these results, Angrist and Pischke (2009, Section 3.3.1) argue that "the differences between regression and matching are unlikely to be of major empirical importance."

However, Corollary 1 shows that Angrist's (1998) argument cannot be extrapolated beyond the saturated case that he considered. The result implies that any deviation from full saturation will mean that the OLS estimand fails to be weakly causal unless one assumes that the propensity score $\mathbb{P}[T = 1|X = x] = \mathbb{E}[T|X] = \mathbb{L}[T|X]$ is actually linear in X. Moreover, whenever Angrist's (1998) saturated specification can actually be implemented, the overlap condition $\mathbb{P}[T = 1|X = x] \in (0, 1)$ must hold for every x, or else there would be perfect collinearity.

The implication of Corollary 1 then is that there are only two situations in which Angrist's (1998) linear regression implementation of selection on observables will be weakly causal. First, when the propensity score is implicitly assumed to be linear. Second, when it is also possible to nonparametrically estimate conditional average treatment effects x-by-x. The first case involves a parametric assumption, while in the second case one could just as well weight the x-by-x treatment effects into a parameter such as the average treatment effect that is not only weakly causal but also has a clear counterfactual interpretation. Outside these two cases, β_{ols} is not weakly causal.

4.4 Specifications with more general excluded variables

Słoczyński (2020, 2024) considers the interpretation of TSLS estimands with a binary treatment and a binary instrument under the assumption that the specification has rich covariates. He considers both Assumption MON, which he calls "strong" monotonicity, and a "weak monotonicity" counterpart in which the direction of monotonicity can vary with x. Słoczyński (2020, 2024) shows that if Assumption MON fails, but weak monotonicity is satisfied, then β_{iv} will not be weakly causal even if the specification has rich covariates. The problem is that the IV specification includes only a single excluded variable, Z, so is not flexible enough to pick up changes in monotonicity in the first stage. Słoczyński (2020, 2024) shows that this problem can be resolved by using the "saturate and weight" TSLS specification in Angrist and Pischke (2009), which includes interactions between X and Z as excluded variables.

Kolesár (2013) provided general sufficient conditions for the TSLS estimand to be interpreted as a non-negatively weighted average of treatment effects under weak monotonicity. As noted above, Kolesár (2013) maintained rich covariates as an assumption, whereas we show that rich covariates is a necessary condition. Kolesár (2013) showed that given rich covariates, the TSLS estimand can be written as a weighted average of treatment effects. Whether the weights are non-negative depends on whether the first stage equation is able to sufficiently well approximate the nonparametric propensity score. In our working paper (Blandhol et al., 2022), we provided a lower-level characterization of when the weights are non-negative in terms of the first stage specification being "monotonicity-correct." The takeaway from that characterization confirmed Słoczyński's intuition that even when rich covariates is satisfied, an additional necessary condition for TSLS to be weakly causal is that the first stage is sufficiently flexible to reproduce the direction of monotonicity across covariate groups.

The rich covariates condition extends readily to more general types of excluded variables. Suppose that the excluded variables are a vector $i(Z, X)$ with population first stage coefficient vector γ. Let $\dot{Z} \equiv \gamma' i(Z, X)$. In Blandhol et al. (2022) we show that a necessary condition for the resulting TSLS estimand to be weakly causal is that $\mathbb{E}[\dot{Z}|X = x] = \mathbb{L}[\dot{Z}|X = x]$ for all x, a condition that naturally generalizes the case considered here with $i(Z, X) = Z$ scalar. This condition has basically the same content as Definition RC, but involves the aggregate $\dot{Z}$ instead of just Z itself.

4.5 Constant, linear treatment effects

Suppose that we assume that treatment effects are constant and linear.

Assumption CLE. (Constant, linear effects) There exists a constant Δ such that $\mu_j(g, x) - \mu_{j-1}(g, x) = \Delta(t_j - t_{j-1})$ for every $j \geq 1$, $g \in \mathcal{G}$ and $x \in \mathcal{X}$.

Theorem 1 continues to hold under Assumption CLE, except Assumption MON no longer needs to be maintained.

Proposition 5. Suppose that Assumptions EXO and CLE are satisfied. Then β_{iv} is weakly causal if *and only if* the IV specification has rich covariates.

The assumptions of Proposition 5 allow for a simple illustration of the level dependence phenomenon. Assumption CLE implies that $Y = Y(t_0) + \Delta T$, so

$$\beta_{\text{iv}} = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}(Y(t_0) + \Delta T)] = \Delta + \overbrace{\mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}Y(t_0)]}^{\text{depends on } Y(t_0)}. \tag{10}$$

Using Assumption EXO, the potentially level dependent term can be written as

$$\mathbb{E}[\tilde{Z}Y(t_0)] = \mathbb{E}\left[\mathbb{E}[\tilde{Z}|X]\,\mathbb{E}[Y(t_0)|X]\right]. \tag{11}$$

The nonparametric IV model does not restrict $\mathbb{E}[Y(t_0)|X]$ at all. Level dependence will therefore happen whenever $\mathbb{E}[\tilde{Z}|X] \neq 0$ with positive probability, which in turn happens whenever the IV specification does not have rich covariates.

Proposition 5 shows that the necessity of rich covariates does not have to do with heterogeneous or nonlinear treatment effects per se. Rather, it is a fundamental consequence of the exercise started by Imbens and Angrist (1994) of interpreting a linear IV estimand through a nonparametric IV model. The linear IV estimator was designed for the linear IV model; giving it a causal interpretation within a nonparametric IV model requires additional parametric assumptions.

Instead of that additional parametric assumption being rich covariates, one can maintain a parametric assumption on a conditional mean of the potential outcomes.

Assumption LIN. (Linear potential outcome mean) $\mathbb{E}[Y(t_j)|X = x] = \eta' x$ for some η and some j.

Proposition 6. Suppose that Assumptions EXO, CLE, and LIN are satisfied. Then $\beta_{\text{iv}} = \Delta$, so β_{iv} is weakly causal.

Assumption LIN—or something similar—is explicitly stated in classical and textbook treatments of IV models, e.g. Heckman and Robb (1985, pp. 184–186) or Wooldridge (2010, pg. 939). But it is not part of the nonparametric IV model that is used to justify the widely-invoked "LATE interpretation" of the linear IV estimator (Angrist and Imbens, 1995). As Abadie (2003, pg. 247) points out, an undesirable implication of Assumption LIN is that one can have $\beta_{\text{iv}} = \Delta$ even if the excluded "instrument" Z is actually some nonlinear function of X alone, an example of what Angrist and Pischke (2009, pg. 191) describe as "back-door identification."

A higher-level alternative to having rich covariates or imposing Assumption LIN is to directly assume that the left-hand side of (11) is zero. This assumption appears in Wooldridge's (2010, pg. 937) discussion of the binary treatment case as the assumption that $\mathbb{L}[Y(t_0)|X, Z]$ does not depend on Z. If we put aside knife-edge balancing cases, (11) shows that this assumption either requires rich covariates or Assumption LIN.

However, considering the high-level assumption usefully exposes the fundamental problem with using the nonparametric IV model to justify linear IV: Assumption EXO by itself *does not* imply that regressing $Y(t_0)$ onto X and Z would yield a zero coefficient on Z, even though this condition is essential for giving the linear IV estimand a causal interpretation.

Assumption LIN was also maintained in Propositions 1 and 2, which showed that β_{iv} is not weakly causal without rich covariates. This does not contradict Proposition 6 because of the addition of constant, linear treatment effects (Assumption CLE). When Assumption CLE is removed to allow for heterogeneous treatment effects, Assumption LIN no longer suffices as a substitute for rich covariates.

5 Alternatives to linear IV

5.1 Partially linear IV

Theorem 1 shows that β_{iv} is weakly causal if and only if the IV specification has rich covariates. If weak covariates is satisfied, it follows from (8) that $\beta_{\text{iv}} = \beta_{\text{rich}}$, where

$$\beta_{\text{rich}} \equiv \frac{\mathbb{E}\left[Y(Z - \mathbb{E}[Z|X])\right]}{\mathbb{E}\left[T(Z - \mathbb{E}[Z|X])\right]} = \frac{\mathbb{E}\left[\mathbb{C}[Y, Z|X]\right]}{\mathbb{E}\left[\mathbb{C}[T, Z|X]\right]}. \tag{12}$$

If rich covariates is not satisfied, then it may be that $\beta_{\text{iv}} \neq \beta_{\text{rich}}$, however we can still consider β_{rich} as what the IV estimand would have been had rich covariates actually been satisfied. Given Assumptions EXO and MON, β_{rich} always satisfies the minimal requirement of being weakly causal.

One way to estimate β_{rich} is to use a richer linear IV specification that controls for covariates so flexibly that rich covariates must be satisfied. If X is discrete, then this is the same as using a saturated specification with one dummy variable for each discrete value of X. These types of specifications are discussed in Angrist (1998) and Angrist and Pischke (2009), but were rarely used in the IV papers in our survey (Table 2). They take an extreme position on the bias-variance trade-off that is difficult to defend for settings in which X takes many values.

Chernozhukov et al. (2018) show how machine learning (ML) methods can be used to estimate a modification of the classical linear IV model where the linear function of covariates has been replaced by an unknown function. They describe the model as partially linear IV (PLIV). It is straightforward to show that the coefficient on treatment in their model is equal to β_{rich}. Chernozhukov et al. (2018) show how to construct the Neyman orthogonal score for the PLIV model, which depends on the treatment coefficient as well as the functions

$$\nu(x) \equiv (\mathbb{E}[Y|X=x], \mathbb{E}[T|X=x], \mathbb{E}[Z|X=x]). \tag{13}$$

They then show how to use the orthogonality of the score in conjunction with cross-fitting to construct consistent and asymptotically normal estimators of the treatment coefficient under nonparametric assumptions about the unknown functions that comprise ν. They suggest estimating ν using supervised ML algorithms such as random forests, gradient boosted trees, and neural networks.

5.2 Unconditional average causal response

Proposition 4 showed that any weakly causal estimand, such as β_{rich}, can be written as a non-negatively weighted average of subgroup treatment effects. Given rich covariates, the general form of the weights in (9) becomes

$$\omega_j(g,x) = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{C}\left[\mathbb{1}[T \geq t_j], Z|G=g, X=x\right]\mathbb{P}[G=g, X=x]. \tag{14}$$

While (14) has a reasonable statistical interpretation—larger groups and contrasts that covary more with the instrument get more weight—it does not appear to have a more concrete counterfactual interpretation. One obstacle is that the instrument can be multivalued, which even without covariates turns the linear IV estimand into a weighted average of treatment effects across different complier groups (Imbens and Angrist, 1994). If the instrument $Z \in \{0,1\}$ is binary, then a parameter that does have a concrete counterfactual interpretation is the unconditional average causal response (ACR) (Angrist and Imbens, 1995):

$$\beta_{\text{acr}} \equiv \mathbb{E}\left[Y(T(1)) - Y(T(0))|T(1) > T(0)\right]. \tag{15}$$

Note that β_{acr} is the LATE when $T \in \{0,1\}$ is binary, so that $T(1)=1$ and $T(0)=0$.

As Słoczyński (2020, 2024) points out, the difference between β_{rich} and β_{acr} can be large. To see why, let $\beta_{\text{acr}}(x) \equiv \mathbb{E}[Y(T(1)) - Y(T(0))|T(1) > T(0), X=x]$ be the ACR conditional on $X=x$. Then iterating expectations shows that

$$\beta_{\text{acr}} = \mathbb{E}\left[\beta_{\text{acr}}(X)\frac{\mathbb{P}[T(1) > T(0)|X]}{\mathbb{P}[T(1) > T(0)]}\right], \tag{16}$$

whereas, with a bit of algebra, it can also be shown that

$$\beta_{\text{rich}} = \mathbb{E}\left[\beta_{\text{acr}}(X)\frac{\mathbb{P}[T(1) > T(0)|X]\,\mathbb{V}[Z|X]}{\mathbb{E}[\mathbb{P}[T(1) > T(0)|X]\,\mathbb{V}[Z|X]]}\right]. \tag{17}$$

The difference between β_{acr} and β_{rich} arises because the latter puts extra weight on values of X with more variation in Z. Słoczyński (2020, 2024) argues that β_{acr} is likely what empirical researchers have in mind, and he shows that the difference in weights can make β_{rich} misleading. So, even if rich covariates holds, β_{tsls} may not be what an empirical researcher expects. In Section 6, we find empirical evidence that β_{rich} and β_{acr} can be quite different.

We produce this evidence by directly estimating β_{acr} using two different approaches. Both approaches are based on a finding due to Tan (2006) and Frölich (2007) that

$$\beta_{\text{acr}} = \frac{\mathbb{E}[\mathbb{E}[Y|Z=1,X] - \mathbb{E}[Y|Z=0,X]]}{\mathbb{E}[\mathbb{E}[T|Z=1,X] - \mathbb{E}[T|Z=0,X]]}. \tag{18}$$

The first approach comes from Chernozhukov et al. (2018), who show how to estimate β_{acr} using DDML. The orthogonal score that they derive involves the five functions

$$\nu(x) \equiv \big(\mathbb{E}[Y|Z=0,X=x], \mathbb{E}[Y|Z=1,X=x], \mathbb{E}[T|Z=0,X=x], \\ \mathbb{E}[T|Z=1,X=x], \mathbb{E}[Z|X=x]\big),$$

all of which can be estimated nonparametrically using ML algorithms. The second approach comes from Uysal (2011), Heiler (2022), and Słoczyński et al. (2024), who exploit the connection that (18) has with propensity score weighting: the numerator looks like the average treatment effect of Z on Y and the denominator like the average treatment effect of Z on T. They propose a weight-normalized inverse propensity score estimator and derive its asymptotic properties. Implementing the estimator requires parameterizing the instrument propensity score, $\mathbb{E}[Z|X]$.[10]

5.3 Abadie's (2003) κ

Abadie (2003, Section 4.2.1) and Angrist and Pischke (2009, pp. 179–180) suggest using a weighted regression to control for covariates when both $T \in \{0,1\}$ and $Z \in \{0,1\}$ are binary. As Abadie (2003) showed, a weighted linear regression of Y on T and X with weights given by

$$\kappa \equiv 1 - \frac{T(1-Z)}{1-\mathbb{E}[Z|X]} - \frac{(1-T)Z}{\mathbb{E}[Z|X]} \tag{19}$$

is, in the population, the same as an unweighted linear regression of Y on T and X among the subpopulation of compliers. Abadie (2003) showed that if rich covariates hold, then the κ-weighted estimate of the coefficient on T is numerically equivalent to

[10] See also MaCurdy et al. (2011), Donald et al. (2014), Ogburn et al. (2015), Sun and Tan (2022), and Singh and Sun (2024) for related estimators.

the linear IV estimate, so estimates a weakly causal estimand.

The next proposition shows that rich covariates turns out to also be *necessary* for the κ-weighted estimand to be weakly causal.

Proposition 7. Suppose that Assumptions EXO and MON are satisfied and that both T and Z are binary. Let β_{abadie} be the population estimand for a weighted linear regression of Y on T and X with weights given by κ. Then β_{abadie} is weakly causal if and only if the linear IV specification has rich covariates, so that $\mathbb{E}[Z|X] = \mathbb{L}[Z|X]$. When this is true, $\beta_{\text{abadie}} = \beta_{\text{iv}} = \beta_{\text{rich}}$.

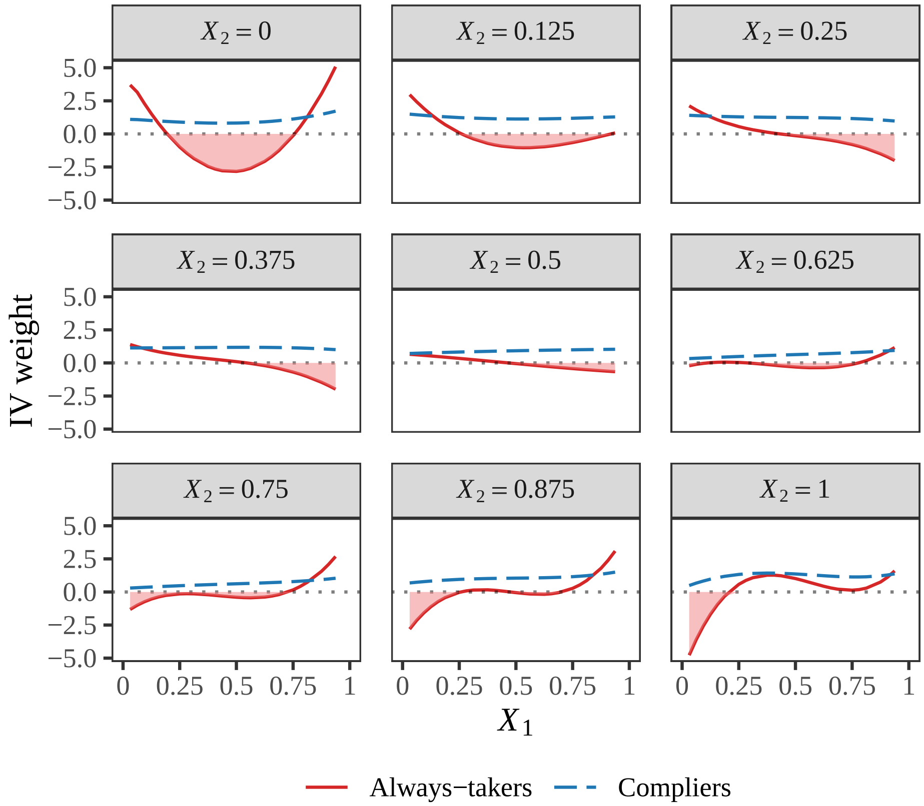
Proposition 7 shows that the κ-weighting estimate is vacuous in the same way that Corollary 1 showed that Angrist's (1998) interpretation of the OLS estimand was vacuous. When rich covariates holds, the κ-weighted estimate is numerically equal to the linear IV estimate, and so there's no reason to use it. When rich covariates does not hold, the κ-weighted estimate is not weakly causal.

Angrist and Pischke (2009, pp. 180–181) use Angrist's (2001) reanalysis of Angrist and Evans (1998) as an example to dismiss the relevance of Abadie's (2003) approach. Yet Angrist (2001, pg. 12) also reports that "the covariates are not highly correlated with the twins instruments..." Our findings show why it is misleading to extrapolate the Angrist and Pischke (2009) argument to other empirical settings: the case when Z is mean independent of X is one where *any* covariate specification is rich. If Z and X are dependent—as is often the case when covariates are used in an IV analysis—then the linear IV estimand will not have a complier interpretation unless $\mathbb{E}[Z|X = x]$ is modeled correctly. At the same time, Proposition 7 also implies that the implementation of Abadie's κ proposed by Angrist and Pischke (2009, pp. 180–181) *only* has a causal interpretation when the IV specification has rich covariates.

5.4 Monte Carlo simulation

In this section, we report the results of a Monte Carlo simulation based on a data generating process (DGP) calibrated to Card's (1995) data on the returns to schooling, which we reanalyze in the next section. We use covariates $X \equiv (X_1, X_2)$, where X_1 takes a number of values that we vary across simulations, while X_2 always takes nine values. We generate the binary instrument Z—presence of a nearby college in Card's application—by specifying $\mathbb{E}[Z|X = x]$ to be an interacted cubic polynomial fit to the Card data with X_1 as experience and X_2 as region indicators (Figure SA.1). We generate the binary treatment T (college attendance) so that $\mathbb{P}[T = 1|Z = z]$ matches its estimated counterpart in Card's data. Then, we generate the outcome Y (log wages) using an optimization procedure that matches several estimates in Card's data while also ensuring that Assumption LIN is satisfied, as in the simplified case discussed in

Figure 2: Weights for β_{iv} in the simulation DGP



Notes: The figure shows the weights in Proposition 1 for the linear IV estimand when X_1 takes 24 values. The weights vary by both choice group and $X = (X_1, X_2)$. The weights for the compliers are always non-negative, but the weights for the always-takers are often negative, as shown in shaded red. The decomposition underlying the figure is not unique (Proposition 2). Figure SA.2 shows the analogous figure for a decomposition involving only compliers and never-takers.

Section 2. See Appendix SA.2 for more details.

We use this DGP to compare the performance of five estimators.

The first is a linear IV estimator that controls for X_1 linearly while including a full set of indicators for X_2, but omits any nonlinear or interaction terms. This specification does not satisfy rich covariates, so is not weakly causal. Figure 2 illustrates the weights for the estimand β_{iv} of this estimator using the Proposition 1 decomposition into compliers and always-takers, for the case when X_1 takes 24 values. All complier groups are positively-weighted. However, always-takers receive considerable weight, both positive and negative. The overall value of β_{iv} is .660, which reflects the sum of .391 from positively-weighted compliers, .614 from positively-weighted always-takers, and $-.345$ from negatively-weighted always-takers. Figure SA.2 shows that writing β_{iv}

in terms of complier and never-takers instead of always-takers also leads to negative weights, as shown in Proposition 2. The second estimator is Abadie's κ-weighted estimator using the same covariate specification, which Proposition showed is also not weakly causal. The estimand for the κ-weighted estimator, β_{abadie}, is very similar to that for β_{iv} regardless of how many values X_1 takes (Figure SA.3).

The third estimator is a linear IV estimator that includes nonlinear and interaction terms, so that rich covariates is satisfied. We call this estimator "correctly specified," as it assumes that X has been chosen so that $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$. The fourth estimator is a linear IV estimator that is saturated in X. This estimator also satisfies rich covariates, but the number of regressors it uses increases with the support of X_1, which we will vary in the simulation. The fifth estimator is a DDML estimator for the PLIV model using an ensemble of a random forest with 1000 trees, gradient boosted trees with 1000 stages, and a neural network with two neurons.[11] Each of these three estimators can be viewed as estimating $\beta_{\text{rich}} = .430$, which is a weakly causal estimand comprised of only non-negatively weighted complier effects.

The top row of Figure 3 compares the means of the five estimators.[12] The column varies the sample size while the x-axis varies the size of the support of X_1. The linear IV estimator converges to the negatively-weighted estimand β_{iv}, so it exhibits a bias for β_{rich} that does not decrease with the sample size. The correctly specified and saturated estimators both converge to β_{rich}, however when the sample size is small relative to the number of covariate values, the saturated estimator exhibits substantial bias. The bias of the DDML-PLIV estimator for β_{rich} is larger than the correctly specified estimator, but decreases as the sample size increases.

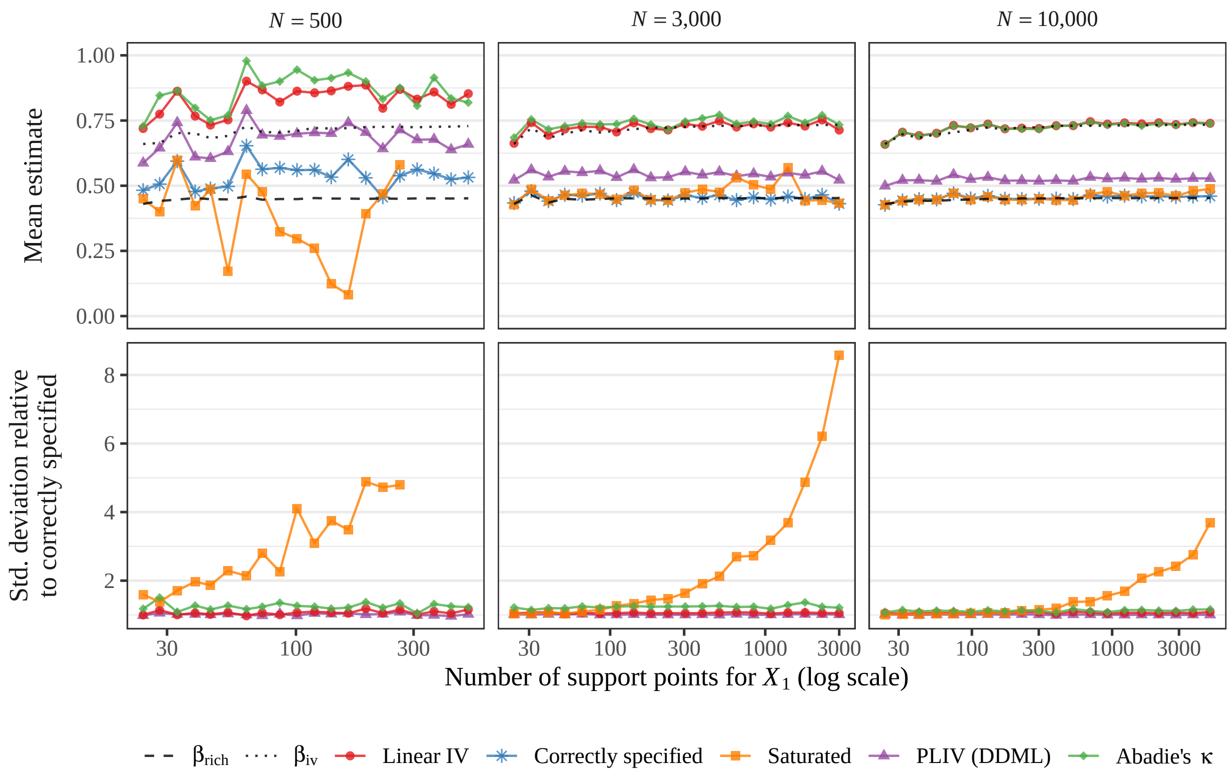
The bottom row of Figure 3 compares the standard deviations of the five estimators. The comparison is taken relative to the correctly-specified estimator to keep the magnitude comparable across sample sizes. The linear, κ-weighted, correctly specified, and DDML-PLIV estimators all exhibit broadly similar standard deviations across sample sizes and number of covariate values. The flexibility of these estimators does not depend on the number of values that the covariates take, so increasing the support of X_1 does not have a large impact on their standard deviations. In contrast, the standard deviation of the saturated linear IV estimator explodes as the number of covariates increases.

Overall, the simulation results show that saturated specifications may be too flexible to yield usefully-precise estimates. This is likely the reason that saturated specifica-

[11] The ensemble is formed by short-stacking with convex weights chosen through non-negative least squares, as advocated by Ahrens et al. (2023, 2024b). The DDML estimates are random due to the sample splits used in cross-fitting. We repeat each estimate across five sample splits and report the resulting median, as recommended by Chernozhukov et al. (2018).

[12] More detailed tabular results are reported in Table SA.1.

Figure 3: Simulation results



Notes: Each point is constructed from 500 draws from the DGP discussed in Appendix SA.2. Means and standard deviations are computed after trimming the top and bottom .025 quantiles of the distribution. Results for the saturated estimator for $N = 500$ and more than 300 support points have many undefined draws so are excluded.

tions are used so infrequently in the literature (Section 2.4). The ideal, but infeasible, solution would be to use a linear IV estimator in which $\mathbb{E}[Z|X]$ is known to be correctly specified, so that rich covariates is satisfied. Without such knowledge, this solution entails an assumption that the specification is in fact correct. If assuming rich covariates is unattractive, then the DDML-PLIV estimator provides a feasible alternative that can viewed as nonparametrically estimating the weakly causal quantity β_{rich}. In our simulation, it leads to estimates that have similar precision as linear IV.

6 Applications

In this section, we use our findings to reanalyze several empirical studies. We begin with Card's (1995) estimates of the returns to education as a classic and familiar example. Next, we turn to the papers by Nunn and Wantchekon (2011) and Dube and Harish (2020) as more modern examples of how linear IV is applied and interpreted in practice. Finally, we reanalyze the main estimates for ten studies from our survey in Section 2.4.

For all studies, we reproduce the original linear IV estimates alongside their comparable OLS estimates. We conduct a Ramsey (1969) RESET test of the null hypothesis that $\mathbb{E}[Z|X]$ is linear in X, that is, of rich covariates.[13] We then estimate β_{rich} with DDML using the same ensemble as in Section 5.4. Standard errors for all estimators are heteroskedasticity and/or cluster-robust depending on the original application.[14]

6.1 Card (1995)

Card (1995) used a sample of 24–34-year-old men from the 1976 interview of the NLSY to estimate the returns to education. The outcome Y is log hourly wage. The treatment T is years of education. The instrument Z is a binary indicator for the presence of an accredited four-year college in the local labor market when the respondent was 14 years old. In his main results, Card (1995, Table 3A, column (5)) includes the following covariates as X: a quadratic in years of potential experience, a race indicator for Black, geography indicators for living in the South and in an urban area, a set of indicators for region of residence in 1966, and an indicator for residence in an urban area in 1966. All of these terms enter additively, so the covariate specification is not saturated and might not satisfy rich covariates.

Column (1) of Table 3 reproduces Card's IV estimate of the returns to education and OLS estimates of the comparable OLS estimand, β_{ols}, that instruments for T with itself. The original linear IV estimate of .132 uses covariates and increases by about 30% if the covariates are omitted. The RESET test overwhelmingly rejects the null hypothesis that the specification has rich covariates. By Theorem 1, this is the same as rejecting the null hypothesis that β_{iv} has a weakly causal interpretation. However, it doesn't necessarily imply that β_{iv} is not equal to β_{rich}, a quantity which does have a weakly causal interpretation.

The DDML estimate of β_{rich} reported in the fourth row is modestly smaller than the IV estimate of β_{iv}, with a similar standard error. Some perspective on the magnitude of this difference is given in the row titled relative specification bias, where we report an estimate of $|\beta_{\text{iv}} - \beta_{\text{rich}}|/|\beta_{\text{iv}}|$ at about .076, or roughly an 8% difference. The subsequent row reports an estimate of $|\beta_{\text{iv}} - \beta_{\text{rich}}|/|\beta_{\text{ols}} - \beta_{\text{rich}}|$, which at roughly 21% shows that the difference between β_{iv} and β_{rich} represents a sizable fraction of the "selection bias" between OLS and the DDML estimate.

The sixth row reports DDML estimates of β_{acr}. While both β_{rich} and β_{acr} are weakly causal, the DDML estimate of β_{acr} is roughly half the size of the DDML estimate of β_{rich}, with a comparable standard error. This difference likely reflects the difference in

[13] We implement the RESET test using the second and third orders of the fitted values.

[14] For the DDML estimates, we still report the median estimate, but we use 100 random sample splits instead of five as in the simulations. The DDML standard errors include an adjustment for uncertainty over splits (Chernozhukov et al., 2018, pg. C30).

Table 3: Comparison of IV estimates for three applications

	(1)	(2)	(3)
	Card (1995)	Nunn & Wantchekon (2011)	Dube & Harish (2020)
OLS	0.075 (0.004)	–0.203 (0.033)	0.115 (0.035)
IV, no covariates	0.188 (0.026)	–0.190 (0.111)	1.011 (0.522)
IV, with covariates	0.132 (0.054)	–0.271 (0.088)	0.400 (0.211)
PLIV (DDML)	0.122 (0.053)	–0.071 (0.091)	0.318 (0.240)
Abadie's κ	—	—	–0.970 (7.330)
LATE/ACR (DDML)	0.067 (0.046)	—	0.203 (0.141)
LATE/ACR (IPSW)	0.085 (0.049)	—	1.362 (0.356)
Ramsey RESET test p-val. (H_0 : rich covariates)	0.000	0.000	0.000
Relative specification bias	0.076	0.738	0.204
Specification vs. selection bias	0.213	1.515	0.400
Outcome variable	log(hourly wage)	Trust in neighors	At war
Outcome variable, mean	6.262	1.732	0.296
Treatment variable	Years of education	log(1 + slave exports)	Queen ruling
Treatment variable, mean	13.263	0.621	0.160
Included variables	14	99	66
Sample size	3,010	16,679	3,586

Notes: Heteroskedasticity- or cluster-robust standard errors are reported in parentheses. Standard errors for Abadie's κ are bootstrapped with the top and bottom .5% of the bootstrap distributed trimmed. LATE estimates are not reported in column (2) because the instrument is not binary. Relative specification bias is an estimate of $|\beta_{\text{iv}} - \beta_{\text{rich}}|/|\beta_{\text{iv}}|$ and specification vs. selection bias is an estimate of $|\beta_{\text{iv}} - \beta_{\text{rich}}|/|\beta_{\text{ols}} - \beta_{\text{rich}}|$.

weights discussed in Section 5.2, providing an empirical illustration of a critique made by Słoczyński (2024). In the sixth row, we report an alternative estimate of β_{acr} that uses the normalized instrument-propensity score weighting (IPSW) estimator proposed by Uysal (2011), Heiler (2022), and Słoczyński et al. (2024). The estimate and standard error are quite similar to those for DDML.

6.2 Two modern applications

We now turn to two more recent examples. These examples are explicit in their use of an extensive set of covariates to justify the exogeneity of the instrument.

We first consider an influential paper by Nunn and Wantchekon (2011), who estimate the effect of the slave trade on modern day measures of trust in Africa using data from the 2005 Afrobarometer survey. The outcome Y is the respondent's reported level of trust in their neighbors. The treatment T is the natural log of (one plus) total historical slave exports for the respondent's ethnic group. The instrument Z is the historical distance of the respondent's ethnic group from the nearest coast. In their main results, Nunn and Wantchekon (2011) include as covariates X a set of country fixed effects, a set of demographic controls for the respondent, measures of ethnic homogeneity for the respondent's district, a set of variables intended to proxy for the amount of European influence, distance of the ethnic group's historical homeland to the Saharan slave trade, and a historical measure of the ethnic group's reliance on fishing. In total, there are 93 covariates. The authors are quite explicit that their motivation for incorporating these covariates is to help ensure the exogeneity of their instrument (Nunn and Wantchekon, 2011, pg. 3239).

Column (2) of Table 3 reproduces the IV estimates from column (2) of Table 6 in Nunn and Wantchekon (2011) alongside the comparable OLS estimate. The RESET test again overwhelmingly rejects the null of rich covariates. In this case, the IV estimate of β_{iv} is almost four times as large as the DDML estimate of β_{rich}, representing one and a half times the difference in magnitude between the IV and OLS estimates. The DDML estimate has a similar standard error to the IV estimate. Based on the DDML estimate, the null hypothesis that the slave trade had no impact on levels of trust would not be rejected at conventional significance levels, contrary to the central finding of Nunn and Wantchekon (2011). Note that unconditional ACR estimates cannot be computed for this application because the instrument is not binary.

Next, we consider a paper by Dube and Harish (2020), who estimate the effect of queen rule on war using panel data on the polities of Europe covering the years 1480 to 1913. The outcome Y is a binary indicator for whether a polity-year observation was at war. The treatment T is a binary indicator for whether a queen ruled in that polity-year. The instrument Z is an indicator for whether the previous monarch had a legitimate firstborn male child. The covariates X in their main results (Dube and Harish, 2020, Table 3, column (3)) are polity and decade identifiers, whether the previous monarchs were corulers unrelated to one another, whether they had any legitimate children (with and without missing birth years), and whether the gender of the previous firstborn child is missing.

Dube and Harish (2020) justify most of their covariates with concerns about exogeneity of the instrument. For example, they argue that controlling for whether the previous monarch had any legitimate children is necessary because the firstborn son instrument is mechanically zero whenever the previous monarch had no children (Dube and Harish, 2020, pp. 2601–2602). In Table SA.2 we show that without polity fixed effects their IV estimates are implausibly large, sometimes exceeding the logical value of 1, albeit with large standard errors. With both polity and decade fixed effects, but without the previous monarch controls, their estimates are close to half as large in magnitude. Covariates apparently matter substantially for their conclusions.

Dube and Harish (2020, pg. 2605) explicitly invoke a LATE interpretation for their estimates:

> *If there are heterogeneous treatment effects, the IV estimate will be the LATE (Imbens and Angrist 1994). It will tell us the effect for the specific group of women who were eligible to rule and induced into ruling because of the presence of a firstborn female or sister among previous monarchs (i.e., the set of women who were compliers).*

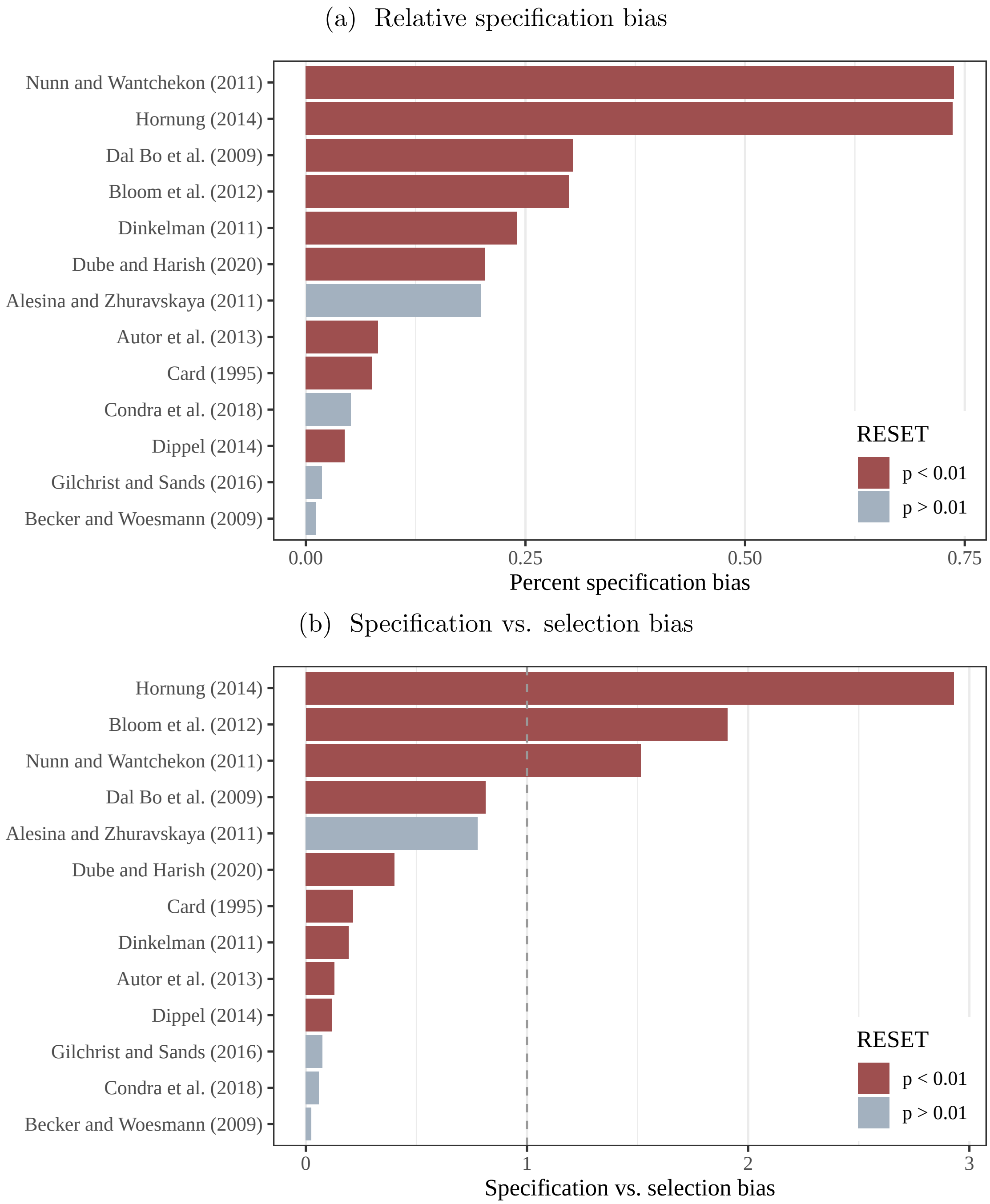
Both the treatment and instrument are binary, so the idea of a single LATE as introduced in Imbens and Angrist (1994) is well-defined. However, even if covariates are rich, linear IV estimates not an unconditional LATE, but β_{rich}, which is a weighted average of different covariate-specific LATEs. If covariates are not rich, then β_{iv} is not even weakly causal, let alone an estimate of the LATE.

Column (3) of Table 3 replicates Table 3, column (3) of Dube and Harish (2020) along with the comparable OLS estimates. The RESET test once again overwhelmingly rejects the null hypothesis that β_{iv} is weakly causal. The DDML estimate of β_{rich} is about 20% smaller than the original estimate, or 40% of the difference between the original IV and OLS estimates. While estimated with similar precision as linear IV, it is no longer significantly different from zero at conventional levels.

In the fifth row of column (3), we report the κ-weighted estimator discussed in Section 5.3, which we can apply here because both the instrument and treatment are binary. The estimates of κ use a logit to estimate $\mathbb{E}[Z|X]$. Proposition 7 showed that this estimator will converge to $\beta_{\text{iv}} = \beta_{\text{rich}}$ if rich covariates hold. In this application, we find that the resulting estimate is implausibly large in magnitude and imprecisely estimated. This may be because the logit estimator is contaminated by incidental parameter bias due to the large number of fixed effects.

Because the instrument is binary, we can also estimate β_{acr} using either DDML or IPSW. The treatment is also binary, so $\beta_{\text{acr}} = \beta_{\text{late}} \equiv \mathbb{E}[Y(1) - Y(0)|D(1) > D(0)]$. The DDML estimate of β_{late} is about half the size of the original IV estimate and about two thirds the size of the DDML estimate of β_{rich}. Although it is estimated more precisely, it is not statistically different from zero at conventional levels. The IPSW estimate of β_{late} is implausibly large and imprecisely estimated. As with the κ-weighted estimator, this may be because of incidental parameters bias.

Figure 4: Relative magnitude across several applications



Notes: These figures present the absolute difference between linear IV and DDML estimates relative to the linear IV estimate (panel A) and relative to the difference between the OLS and DDML estimates (panel B). The definitions of the two relative biases are as in Table 3. Details on the sample size and number of included variables for each specification are provided in Table SA.4.

6.3 Patterns from multiple studies

To conduct a more systematic evaluation, we return to our survey of IV papers from Section 2.4. We consider all papers from the survey for which data was available and for which we were able to replicate the main IV estimates. We limit our attention to papers that fit the framework of Section 4 with a single endogenous variable and a single instrument used as the sole excluded variable. We omit papers that use panel data with two-way fixed effects, as we found that this was often challenging to implement with DDML. Imposing these restrictions leaves us with ten studies. For comparison, we also include Card (1995), Nunn and Wantchekon (2011), and Dube and Harish (2020), bringing the total to thirteen.

Figure 4 summarizes the differences between the IV estimate of β_{iv} and the DDML estimate of β_{rich} for the main specification in each of these thirteen studies. Panel (a) measures the differences relative to the original IV estimates, while panel (b) measures them relative to the difference between the original IV and comparable OLS estimates. The bars are shaded according to whether the RESET test rejects the null that β_{iv} is weakly causal at the 1% level. Table SA.3 provides tabular results for each study and shows that standard errors for the IV and DDML estimates are generally similar.

The RESET test rejects in nine out of the thirteen studies, implying that for most of these studies β_{iv} is not weakly causal. The magnitude of the difference between the IV estimate of β_{iv} and the DDML estimate of β_{rich} varies across studies, but is often large measured either relative to the original estimates or relative to the difference between the OLS and IV estimates. Cases when this difference is large are reliably detected by the RESET test. The one exception, Alesina and Zhuravskaya (2011), has only 97 observations, so is likely underpowered. Conversely, the studies for which the RESET test does not reject also tend to exhibit small differences between the IV and DDML estimates, suggesting that $\beta_{\text{iv}} = \beta_{\text{rich}}$, and that the original linear IV estimate is indeed weakly causal.

7 Conclusion and recommendations for practice

In discussing the LATE interpretation of linear IV estimates, Angrist and Krueger (1999, pg. 1326) conjectured:

> *That is, IV estimates in models with covariates can be thought of as producing a weighted average of covariate-specific Wald estimates as long as the model for covariates is saturated .... In other cases it seems reasonable to assume that some sort of approximate weighted average is being generated, but we are unaware of a precise causal interpretation that fits all cases.*

We have shown that this seemingly-reasonable assumption is false. Unless rich covariates is satisfied, the linear IV estimand cannot be interpreted as "weakly causal," and so cannot be interpreted as a non-negatively weighted average of LATEs. We tested the null hypothesis of rich covariates in several empirical studies and found that it was commonly rejected.

Based on our theoretical results and empirical applications, we recommend that researchers using linear IV estimators take the following steps.

1. Consider the role of covariates in the IV analysis. If covariates are not essential for justifying instrument exogeneity, then report estimates without covariates. Estimates with covariates can still be reported if the covariates are helpful for precision. If covariates play an important role in justifying exogeneity, then think carefully about which covariates ought to be included and why. Using a "kitchen sink" approach to controlling for covariates makes it less likely that rich covariates is satisfied and so more likely that the resulting IV estimate is not weakly causal.

2. Report the Ramsey (1969) RESET test for a regression of the instrument on the covariates. The null hypothesis of this test is equivalent to the null hypothesis that the IV estimand is weakly causal. The RESET test can be implemented in Stata with the command `estat ovtest` and in `R` through the `resettest` function in the `lmtest` package (Zeileis and Hothorn, 2002).[15] If the RESET test rejects, then proceed to the next step. Otherwise, proceed to step four.

3. Estimate β_{rich} with DDML and report the result alongside the linear IV estimate of β_{iv}. A Stata implementation of DDML has been developed by Ahrens et al. (2023). There are at least two `R` packages (Ahrens et al., 2024a; Bach et al., 2024).

4. If the instrument is binary, then estimate the unconditional ACR, β_{acr}, which is equal to the unconditional LATE, β_{late}, if the treatment is also binary. This can be implemented with DDML in either Stata or `R`. It can also be implemented in Stata with IPSW (Słoczyński et al., 2024). We are not aware of an IPSW package for `R`, although it is straightforward to construct point estimates by fitting a binary response model and then constructing four weighted means.

It is important to emphasize that the criterion of "weakly causal" used throughout the analysis is an extremely weak one. Being weakly causal may be necessary for a quantity to represent an interesting causal effect, but it is not sufficient. Even if rich covariates is satisfied, β_{rich} may be hard to interpret. As we showed empirically, it can

[15] Testing whether a TSLS estimand with multiple excluded variables is weakly causal is less straightforward because the rich covariates condition now concerns the conditional mean of the aggregated excluded variables (see Section 4.4). A bootstrapped version of the RESET test may be an adequate solution in this case. See our working paper (Blandhol et al., 2022) for more details and an example.

also be quite different from a more interpretable quantity, such as the unconditional ACR, β_{acr}, or unconditional LATE, β_{late}.

These interpretation difficulties were already reason to explore alternative IV methods designed to estimate quantities, such as an unconditional LATE or the average treatment on the treated, that are not only weakly causal but also have clear counterfactual interpretations. Such methods rely on explicitly stated parametric assumptions (e.g. Heckman, 1976; Imbens and Rubin, 1997; Heckman et al., 2003) or are semiparametric (e.g. Carneiro et al., 2011; Brinch et al., 2017; Mogstad et al., 2018; Słoczyński et al., 2024) or nonparametric (e.g. Heckman and Vytlacil, 1999; Manski and Pepper, 2000; Chernozhukov et al., 2018). By showing that common interpretations of linear IV estimands also rely on either parametric assumptions or nonparametric implementations, our findings provide another reason to pursue such approaches.

A Proofs

Proof of Proposition 1. The expression for β_{iv} is a special case of Proposition 2 with $\epsilon = 1$.

If $\mathbb{E}[T\tilde{Z}] > 0$, then because $\mathbb{E}[Z|X] \in [0,1]$ for binary Z, the sign of $\omega(\text{CP}, X)$ depends on the sign of $1 - \mathbb{L}[Z|X]$, which is negative if and only if $\mathbb{L}[Z|X] > 1$. The sign of $\omega(\text{AT}, X)$ varies with X according to the sign of $\mathbb{E}[\tilde{Z}|X]$. Because X contains a constant, $\mathbb{E}[\mathbb{E}[\tilde{Z}|X]] = \mathbb{E}[\tilde{Z}] = 0$, so $\mathbb{E}[\tilde{Z}|X]$ is either zero with probability 1, or else it has positive probability of taking both positive and negative values. In the latter case, the sign of $\omega(\text{AT}, X)$ is negative for some values of X regardless of whether $\mathbb{E}[T\tilde{Z}]$ is positive or negative. *Q.E.D.*

Proof of Proposition 2. The numerator of β_{iv} can be written as

$$\mathbb{E}[Y\tilde{Z}] = \mathbb{E}\left[\mathbb{E}\left[Y\tilde{Z}|X\right]\right] = \mathbb{E}\left[\mathbb{C}[Y, Z|X]\right] + \mathbb{E}\left[\mathbb{E}[Y|X]\,\mathbb{E}[\tilde{Z}|X]\right], \tag{20}$$

where $\mathbb{C}$ denotes covariance. The same argument as in Imbens and Angrist (1994) applied conditional-on-covariates yields

$$\mathbb{C}[Y, Z|X] = \Delta(\text{CP}, X)\,\mathbb{C}[T, Z|X] = \Delta(\text{CP}, X)\,\mathbb{P}[G = \text{CP}|X]\,\mathbb{E}[Z|X](1 - \mathbb{E}[Z|X]). \tag{21}$$

As for the second term of (20),

$$\begin{aligned}\mathbb{E}\left[Y|X\right] = \mathbb{E}\left[Y|G=\textsc{at},X\right]\mathbb{P}\left[G=\textsc{at}|X\right] + \mathbb{E}\left[Y|G=\textsc{nt},X\right]\mathbb{P}\left[G=\textsc{nt}|X\right] \\ + \mathbb{E}\left[Y|G=\textsc{cp},X\right]\mathbb{P}\left[G=\textsc{cp}|X\right] \\ = \mathbb{E}\left[Y(1)|G=\textsc{at},X\right]\mathbb{P}\left[G=\textsc{at}|X\right] + \mathbb{E}\left[Y(0)|G=\textsc{nt},X\right]\mathbb{P}\left[G=\textsc{nt}|X\right] \\ + \mathbb{E}\left[(1-Z)Y(0)+ZY(1)|G=\textsc{cp},X\right]\mathbb{P}\left[G=\textsc{cp}|X\right].\end{aligned} \tag{22}$$

Adding and subtracting $\mathbb{E}[Y(0)|G=\textsc{at},X]\,\mathbb{P}[G=\textsc{at}|X]$ gives

$$\mathbb{E}\left[Y|X\right] = \Delta(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X]\,\mathbb{E}[Z|X] + \Delta(\textsc{at},X)\,\mathbb{P}[G=\textsc{at}|X] + \eta_0'X \tag{23}$$

due to both the exogeneity of Z and the linearity assumption on $\mathbb{E}[Y(0)|X=x]$. Alternatively, adding and subtracting $\mathbb{E}[Y(1)|G=\textsc{nt},X]\,\mathbb{P}[G=\textsc{nt}|X]$ to (22) gives

$$\mathbb{E}\left[Y|X\right] = \Delta(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X](\mathbb{E}[Z|X]-1) - \Delta(\textsc{nt},X)\,\mathbb{P}[G=\textsc{nt}|X] + \eta_1'X. \tag{24}$$

So multiplying (23) by ϵ and summing it with (24) multiplied by $1-\epsilon$ gives

$$\begin{aligned}\mathbb{E}[Y|X] = \Delta(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X]\,(\mathbb{E}[Z|X]+\epsilon-1) + \Delta(\textsc{at},X)\epsilon\,\mathbb{P}[G=\textsc{at}|X] \\ + \Delta(\textsc{nt},X)(\epsilon-1)\,\mathbb{P}[G=\textsc{nt}|X] + \epsilon\eta_0'X + (1-\epsilon)\eta_1'X.\end{aligned}$$

Because X and $\tilde{Z}$ are orthogonal,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[Y|X]\,\mathbb{E}[\tilde{Z}|X]\right] = \mathbb{E}\Big[\Delta(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X]\,(\mathbb{E}[Z|X]+\epsilon-1)\,\mathbb{E}[\tilde{Z}|X] \\ + \Delta(\textsc{at},X)\epsilon\,\mathbb{P}[G=\textsc{at}|X]\,\mathbb{E}[\tilde{Z}|X] \\ + \Delta(\textsc{nt},X)(\epsilon-1)\,\mathbb{P}[G=\textsc{nt}|X]\,\mathbb{E}[\tilde{Z}|X]\Big].\end{aligned} \tag{25}$$

Summing (21) and (25), and noting that

$$\begin{aligned}\mathbb{E}[Z|X](1-\mathbb{E}[Z|X]) + (\mathbb{E}[Z|X]+\epsilon-1)\,\mathbb{E}[\tilde{Z}|X] \\ = \left(\mathbb{E}[Z|X]-\mathbb{E}[\tilde{Z}|X]\right)(1-\mathbb{E}[Z|X]) + \epsilon\,\mathbb{E}[\tilde{Z}|X] \\ = \mathbb{L}[Z|X](1-\mathbb{E}[Z|X]) + \epsilon\,\mathbb{E}[\tilde{Z}|X]\end{aligned}$$

yields a weighting expression with weights proportional to the claimed expression but missing a common multiple of $\mathbb{E}[\tilde{Z}T]^{-1}$, which comes from the denominator of β_{iv}. *Q.E.D.*

Proof of Proposition 3. Note that T is only stochastic due to Z after conditioning on X and G, as a direct consequence of the definition of G. Assumption EXO then implies that T and $Y(t)$ are independent conditional on X and G. We use this observation to write

$$\begin{aligned}
\beta &= \sum_{g,x} \mathbb{E}\left[b(T,x,Z)Y \middle| G=g, X=x\right] \mathbb{P}[G=g, X=x] \\
&= \sum_{g,x,j} \mathbb{E}\left[\mathbb{1}[T=t_j] b(t_j,x,Z) Y(t_j) \middle| G=g, X=x\right] \mathbb{P}[G=g, X=x] \\
&= \sum_{g,x,j} \mu_j(g,x)\, \mathbb{E}\left[\mathbb{1}[T=t_j] b(t_j,x,Z) \middle| G=g, X=x\right] \mathbb{P}[G=g, X=x] \\
&\equiv \sum_{g,x,j} \mu_j(g,x)\psi_j(g,x),
\end{aligned}$$

where all summations are taken over $g \in \mathcal{G}, x \in \mathcal{X}, j \in \{0,1,\dots,J\}$, and

$$\psi_j(g,x) \equiv \mathbb{E}\left[\mathbb{1}[T=t_j] b(t_j,x,Z) \middle| G=g, X=x\right] \mathbb{P}[G=g, X=x].$$

Notice that $\omega_j(g,x) = \sum_{k=j}^{J} \psi_k(g,x)$, so that (6) follows from Lemma 1.

Q.E.D.

Lemma 1. For any constants $\{a_j, c_j\}_{j=0}^{J}$,

$$\sum_{j=0}^{J} a_j c_j = a_0 \tilde{c}_0 + \sum_{j=1}^{J} (a_j - a_{j-1}) \tilde{c}_j,$$

where $\tilde{c}_j \equiv \sum_{k=j}^{J} c_k$.

Proof of Lemma 1. Since $c_j = \tilde{c}_j - \tilde{c}_{j+1}$,

$$\begin{aligned}
\sum_{j=0}^{J} a_j c_j &= \sum_{j=0}^{J} a_j \left(\tilde{c}_j - \tilde{c}_{j+1}\right) \\
&\qquad = a_0\tilde{c}_0 + \sum_{j=1}^{J} a_j \tilde{c}_j + \sum_{j=0}^{J-1} a_j \tilde{c}_{j+1} = a_0 \tilde{c}_0 + \sum_{j=1}^{J} (a_j - a_{j-1}) \tilde{c}_j,
\end{aligned}$$

where the final equality used a change of variables in the second summand from j to $j+1$. *Q.E.D.*

Proof of Proposition 4. If $\omega_j(g,x) \geq 0$ and $\omega_0(g,x) = 0$ for all g and x, then it follows immediately from (6) that β satisfies Definition WC.

We will prove the converse by contraposition. That is, we will show that if either the non-negative weights or level independence condition is not satisfied, then there exists a μ such that $\mu_j(g,x) - \mu_{j-1}(g,x)$ has the same sign for every $j \geq 1$, and all g and x, and that this common sign is different from the sign of β. This shows that if the weights do not satisfy both the non-negative and level independence conditions, then β is not weakly causal. Or, by contraposition, if β is weakly causal, then the weights satisfy both conditions.

First, suppose that the level independence condition does not hold, but that the non-negative weights condition does hold. Then there exists a $(g^\star, x^\star)$ such that $\omega_0(g^\star, x^\star) \neq 0$, but $w_j(g,x) \geq 0$ for all $j \geq 1$, g, and x. Set

$$\mu_j(g,x) = \begin{cases} \bar{\mu}, & \text{if } (g,x) \neq (g^\star, x^\star) \\ \mu^\star, & \text{if } (g,x) = (g^\star, x^\star) \text{ and } j < j^\star \\ \mu^\star + \Delta^\star, & \text{if } (g,x) = (g^\star, x^\star) \text{ and } j \geq j^\star \end{cases}, \tag{26}$$

where $\bar{\mu}, \mu^\star$, and $\Delta^\star$ are numbers we will choose, and $j^\star \geq 1$ can be chosen arbitrarily. Then $\mu_j(g,x) - \mu_{j-1}(g,x)$ is zero for all $(g,x) \neq (g^\star, x^\star)$, while for $(g,x) = (g^\star, x^\star)$ it is $\Delta^\star$ when $j = j^\star$ and zero otherwise. In particular, the sign of $\mu_j(g,x) - \mu_{j-1}(g,x)$ is the sign of $\Delta^\star$ for all $j \geq 1$ and all (g,x), regardless of the values of $\bar{\mu}$ and $\mu^\star$. If μ is specified as in (26) with $\bar{\mu} = 0$ for simplicity, then (6) becomes

$$\beta = \omega_0(g^\star, x^\star)\mu^\star + \ \omega_{j^\star}(g^\star, x^\star)\Delta^\star. \tag{27}$$

Fix any $\mu^\star \neq 0$, so that $\omega_0(g^\star, x^\star)\mu^\star \neq 0$. If $\omega_0(g^\star, x^\star)\mu^\star > 0$, then choose a $\Delta^\star < 0$ that is sufficiently small in magnitude so that $\omega_{j^\star}(g^\star, x^\star)\Delta^\star > -\omega_0(g^\star, x^\star)\mu^\star$. Then from (27) we have $\beta = \omega_0(g^\star, x^\star)\mu^\star + \omega_{j^\star}(g^\star, x^\star)\Delta^\star > 0$, so that these choices of $\mu^\star$ and $\Delta^\star$ produce a μ that violates the second condition of Definition WC. Similarly, if $\omega_0(g^\star, x^\star)\mu^\star < 0$, then choose $\Delta^\star > 0$ to be sufficiently small to ensure that $\omega_{j^\star}(g^\star, x^\star)\Delta^\star < -\omega_0(g^\star, x^\star)\mu^\star$, so that $\beta < 0$, contradicting the first condition of Definition WC.

On the other hand, suppose that the non-negative weights condition does not hold, so there exist a $j^\star, g^\star$, and $x^\star$ such that $\omega_{j^\star}(g^\star, x^\star) < 0$. Use the same construction as in (26) with these new values of $j^\star, g^\star$, and $x^\star$, where $j^\star$ is no longer arbitrary. Set $\mu^\star = 0$. Then (27) reduces to

$$\beta = \omega_{j^\star}(g^\star, x^\star)\Delta^\star.$$

Selecting any $\Delta^\star > 0$ produces $\beta < 0$, establishing the existence of a μ that violates the first condition of Definition WC. *Q.E.D.*

Proof of Theorem 1. We evaluate the sufficient and necessary conditions in Proposition 4 using the expressions for $\omega_j(g,x)$ given in (9).

First, consider the level independence condition, which given (9) can be written as

$$\omega_0(g,x) = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}|G=g,X=x]\,\mathbb{P}[G=g,X=x] = 0 \tag{28}$$

for all g and x. Assumption EXO implies that $\tilde{Z} \equiv Z - \mathbb{L}[Z|X]$ is independent of G given X, so

$$\mathbb{E}[\tilde{Z}|G=g,X=x] = \mathbb{E}[\tilde{Z}|X=x] = \mathbb{E}[Z|X=x] - \mathbb{L}[Z|X=x].$$

For every x there exists a $g \in \mathcal{G}$ such that $\mathbb{P}[G=g,X=x] > 0$, because G exhaustively partitions possible choice types. So (28) can hold for every g and x if and only if

$$\mathbb{E}[Z|X=x] = \mathbb{L}[Z|X=x]$$

for every x, that is, if and only if the specification has rich covariates. In particular, if the specification does not have rich covariates, then (28) is non-zero for some g and x, and so by Proposition 4, β_{iv} is not weakly causal.

To establish the sufficient direction, suppose that the specification has rich covariates and consider the non-negative weights condition in Proposition 4. Let $\mathcal{Z}_j(g)$ denote the set of instrument values for which individuals in choice group g would choose a treatment value t_j or larger. Then

$$\begin{aligned}
\omega_j(g,x) &= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\tilde{Z}\mathbb{1}[T \geq t_j]\middle|G=g,X=x\right]\mathbb{P}[G=g,X=x] \\
&= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z \in \mathcal{Z}_j(g)]\middle|G=g,X=x\right]\mathbb{P}[G=g,X=x] \\
&= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{C}\left[Z,\mathbb{1}[Z \in \mathcal{Z}_j(g)]\middle|X=x\right]\mathbb{P}[G=g|X=x]\,\mathbb{P}[X=x],
\end{aligned} \tag{29}$$

where the third equality follows from Assumption EXO and the hypothesis of rich covariates. Given rich covariates,

$$\mathbb{E}[\tilde{Z}T] = \mathbb{E}[(Z - \mathbb{E}[Z|X])\,\mathbb{E}[T|X,Z]] = \mathbb{E}\left[\mathbb{C}[Z,\mathbb{E}[T|X,Z]|X]\right], \tag{30}$$

which is non-negative because Assumptions EXO and MON imply that $\mathbb{E}[T|X,Z]$ is a weakly increasing function of Z (Angrist and Imbens, 1995; Vytlacil, 2002), and the covariance between two weakly increasing functions is non-negative (e.g. Thorisson, 1995, Section 2).

It remains to determine the sign of the covariance term in (29). Suppose that $\mathbb{P}[G=g|X=x] > 0$. Then the function $z \mapsto \mathbb{1}[z \in \mathcal{Z}_j(g)]$ must be weakly increasing. For otherwise, there would exist z, z' with $z < z'$ and $z \in \mathcal{Z}_j(g)$ but $z \notin \mathcal{Z}_j(g)$, meaning that for group g, instrument value z leads to $T(z) \geq t_j$, while instrument value z leads to $T(z) < t_j$. Given that $\mathbb{P}[G = g|X = x] > 0$, this would imply that

$$\mathbb{P}[T(z) \geq t_j > T(z')|X = x] \geq \mathbb{P}[G = g|X = x] > 0, \tag{31}$$

in contradiction with Assumption MON. It follows that the covariance term in (29) is non-negative, again because the covariance of two increasing functions of Z is non-negative. We conclude that $\omega_j(g, x) \geq 0$ for all j, g, and x, which by Proposition 4 shows that β_{iv} is weakly causal. *Q.E.D.*

Proof of Proposition 5. Write (10) as

$$\begin{aligned}
\beta_{\text{iv}} &= \Delta + \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}\,\mathbb{E}[Y(t_0)|Z, G, X]] \\
&= \Delta + \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}\,\mathbb{E}[Y(t_0)|G, X]] \\
&= \Delta + \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\mathbb{E}[\tilde{Z}|X]\,\mathbb{E}[Y(t_0)|G, X]] \equiv \Delta + \sum_{g,x} \omega_0(g, x)\mu_0(g, x),
\end{aligned}$$

where $\omega_0(g, x)$ is as defined as (9), noting that $\mathbb{E}[\tilde{Z}|G, X] = \mathbb{E}[\tilde{Z}|X]$ due to Assumption EXO. If rich covariates holds, then $\omega_0(g, x) = 0$ for all g and x, so that $\beta_{\text{iv}} = \Delta$ is weakly causal by Proposition 4. Conversely, if rich covariates does not hold, then, as shown in the proof of Theorem 1, there exists a (g, x) such that $\omega_0(g, x) \neq 0$, so β_{iv} is not weakly causal, again by Proposition 4. *Q.E.D.*

Proof of Proposition 6. Given Assumption CLE, Assumption LIN also implies that

$$\mathbb{E}[Y(t_0)|X = x] = \mathbb{E}[Y(t_j) - Y(t_0)|X = x] + \mathbb{E}[Y(t_j)|X = x] = \Delta(t_j - t_0) + \eta' x,$$

so that $\mathbb{E}[Y(t_0)|X = x] = \eta_0' x$, where η_0 is the same as η but has $\Delta(t_j - t_0)$ added to the coefficient on the constant component of x. Because $\tilde{Z}$ is orthogonal to X,

$$\mathbb{E}[\tilde{Z}Y(t_0)] = \mathbb{E}[\mathbb{E}[\tilde{Z}|X]\,\mathbb{E}[Y(t_0)|X]] = \mathbb{E}[\mathbb{E}[\tilde{Z}|X]X]'\eta_0 = \mathbb{E}[\tilde{Z}X']\eta_0 = 0.$$

From (10), this implies that $\beta_{\text{iv}} = \Delta$, as claimed. *Q.E.D.*

Proof of Proposition 7. Abadie (2003, Proposition 5.1) showed that if rich covariates is satisfied, then the linear IV estimate is numerically equal to the κ-weighted estimate, implying that $\beta_{\text{abadie}} = \beta_{\text{iv}}$, with $\beta_{\text{iv}} = \beta_{\text{rich}}$ by definition in that case.

For the converse, consider the κ-weighted linear regression, which Abadie showed is the same as an unweighted linear regression of Y on T and X among the subpopulation $G = \textsc{cp}$ of compliers. Assumption MON implies that $\mathbb{P}[T = Z|G = \textsc{cp}] = 1$. Assumption EXO then implies that $(Y(0), Y(1))$ is independent of T conditional on $G = \textsc{cp}$. Corollary 1 therefore implies that the population coefficient on T in the κ-weighted regression is weakly causal if and only if $\mathbb{E}[T|X, G = \text{CP}] = \mathbb{L}[T|X, G = \text{CP}] \equiv \gamma' X$ for some γ. However, Assumption EXO implies that

$$\gamma' X \equiv \mathbb{L}[T|X, G = \text{CP}] = \mathbb{E}[T|X, G = \text{CP}] = \mathbb{E}[Z|X, G = \text{CP}] = \mathbb{E}[Z|X], \qquad (32)$$

which implies that $\mathbb{E}[Z|X]$ is linear in X, and therefore that $\mathbb{E}[Z|X] = \mathbb{L}[Z|X]$, so that rich covariates is satisfied. *Q.E.D.*

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263. 4, 20, 23, 24, 40

Ahrens, A., C. B. Hansen, M. E. Schaffer, and T. Wiemann (2023): "Ddml: Double/Debiased Machine Learning in Stata," . 26, 34

——— (2024a): *Ddml: Double/Debiased Machine Learning*. 34

——— (2024b): "Model Averaging and Double Machine Learning," . 26

Alesina, A. and E. Zhuravskaya (2011): "Segregation and the Quality of Government in a Cross Section of Countries," *American Economic Review*, 101, 1872–1911. 33

Angrist, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66, 249–288. 15, 18, 21, 24

——— (2001): "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors," *Journal of Business & Economic Statistics*, 19, 2–28. 24

Angrist, J. D. and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *The American Economic Review*, 88, 450–477. 24

Angrist, J. D. and G. 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, 431–442. 2, 12, 20, 22, 39

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455. 13, 14

Angrist, J. D. and A. B. Krueger (1999): "Chapter 23 Empirical Strategies in Labor Economics," Elsevier, vol. Volume 3, Part A, 1277–1366. 33

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. 2, 3, 6, 10, 11, 15, 18, 19, 20, 21, 23, 24

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–2168.

Bach, P., M. S. Kurz, V. Chernozhukov, M. Spindler, and S. Klaassen (2024): "DoubleML: An Object-Oriented Implementation of Double Machine Learning in R," *Journal of Statistical Software*, 108, 1–56, arXiv:2103.09603 [stat.ML]. 34

Becker, S. O. and L. Woessmann (2009): "Was Weber Wrong? A Human Capital Theory of Protestant Economic History," *Quarterly Journal of Economics*, 124, 531–596.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When Is TSLS Actually LATE?" Tech. Rep. w29709, National Bureau of Economic Research, Cambridge, MA. 16, 17, 19, 34

Bloom, N., R. Sadun, and J. Van Reenen (2012): "The Organization of Firms Across Countries*," *The Quarterly Journal of Economics*, 127, 1663–1705.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017): "Beyond LATE with a Discrete Instrument," *Journal of Political Economy*, 125, 985–1039. 35

CARD, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. N. Christofides, K. E. Grant, and R. Swidinsky, Toronto: University of Toronto Press, 201–222. 24, 27, 28, 33, 46

CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2015): "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 83, 2453–2483. 15

CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101, 2754–81. 35

CHAMBERLAIN, G. AND G. IMBENS (2004): "Random Effects Estimators with many Instrumental Variables," *Econometrica*, 72, 295–306. 11

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2018): "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21, C1–C68. 4, 21, 23, 26, 28, 35

CONDRA, L. N., J. D. LONG, A. C. SHAVER, AND A. L. WRIGHT (2018): "The Logic of Insurgent Electoral Violence," *American Economic Review*, 108, 3199–3231.

DAL BÓ, E., P. DAL BÓ, AND J. SNYDER (2009): "Political Dynasties," *Review of Economic Studies*, 76, 115–142.

DINKELMAN, T. (2011): "The Effects of Rural Electrification on Employment: New Evidence from South Africa," *American Economic Review*, 101, 3078–3108.

DIPPEL, C. (2014): "Forced Coexistence and Economic Development: Evidence from Native American Reservations," *Econometrica*, 82, 2131–2165.

DONALD, S. G., Y.-C. HSU, AND R. P. LIELI (2014): "Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT," *Journal of Business & Economic Statistics*, 32, 395–415. 23

DUBE, O. AND S. P. HARISH (2020): "Queens," *Journal of Political Economy*, 128, 2579–2652. 27, 30, 31, 33, 51

EVDOKIMOV, K. S. AND M. KOLESÁR (2019): "Inference in Instrumental Variables Analysis with Heterogeneous Treatment Effects," *Working paper*. 16

FIRPO, S., M. N. FOGUEL, AND H. JALES (2020): "Balancing Tests in Stratified Randomized Controlled Trials: A Cautionary Note," *Economics Letters*, 186, 108771. 8

FRÖLICH, M. (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139, 35–75. 23

GILCHRIST, D. S. AND E. G. SANDS (2016): "Something to Talk About: Social Spillovers in Movie Consumption," *Journal of Political Economy*, 124, 1339–1382.

GOLDSMITH-PINKHAM, P., P. HULL, AND M. KOLESÁR (2024): "Contamination Bias in Linear Regressions," *American Economic Review*, 114, 4015–4051. 15

GOODMAN-BACON, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277. 3, 15

HECKMAN, J., J. L. TOBIAS, AND E. VYTLACIL (2003): "Simple Estimators for Treatment Parameters in a Latent-Variable Framework," *Review of Economics and Statistics*, 85, 748–755. 35

HECKMAN, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*. 35

HECKMAN, J. J. AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press. 20

HECKMAN, J. J. AND E. J. VYTLACIL (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences of the United States of America*, 96, 4730–4734. 35

HEILER, P. (2022): "Efficient Covariate Balancing for the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 40, 1569–1582. 23, 29

HORNUNG, E. (2014): "Immigration and the Diffusion of Technology: The Huguenot Diaspora

in Prussia," *American Economic Review*, 104, 84–122.
Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 2, 5, 7, 20, 22, 31, 35
Imbens, G. W. and D. B. Rubin (1997): "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance," *The Annals of Statistics*, 25, 305–327. 35
Kolesár, M. (2013): "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity," *Working paper*. 3, 16, 17, 19
Lee, D. S. (2008): "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697. 15
MaCurdy, T., X. Chen, and H. Hong (2011): "Flexible Estimation of Treatment Effect Parameters," *American Economic Review*, 101, 544–551. 23
Manski, C. F. and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. 35
Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters," *Econometrica*, 86, 1589–1619. 35, 46
Nunn, N. and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101, 3221–3252. 27, 29, 30, 33
Ogburn, E. L., A. Rotnitzky, and J. M. Robins (2015): "Doubly Robust Estimation of the Local Average Treatment Effect Curve," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 373–396. 23
Ramsey, J. B. (1969): "Tests for Specification Errors in Classical Linear Least- Squares Regression Analysis," *Journal of the Royal Statistical Society: Series B (Methodological)*, 31, 350–371. 4, 12, 28, 34
Singh, R. and L. Sun (2024): "Double Robustness for Complier Parameters and a Semi-Parametric Test for Complier Characteristics," *The Econometrics Journal*, 27, 1–20. 23
Słoczyński, T. (2020): "When Should We (Not) Interpret Linear IV Estimands as LATE ?" . 4, 5, 12, 18, 19, 22, 23
——— (2024): "When Should We (Not) Interpret Linear IV Estimands as LATE?" . 4, 5, 12, 18, 19, 22, 23, 29
Słoczyński, T., S. D. Uysal, and J. M. Wooldridge (2024): "Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 1–14. 4, 23, 29, 34, 35
Sun, B. and Z. Tan (2022): "High-Dimensional Model-Assisted Inference for Local Average Treatment Effects With Instrumental Variables," *Journal of Business & Economic Statistics*, 40, 1732–1744. 23
Sun, L. and S. Abraham (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225, 175–199. 3, 15
Tan, Z. (2006): "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618. 4, 23
Thorisson, H. (1995): "Coupling Methods in Probability Theory," *Scandinavian Journal of Statistics*, 22, 159–182. 39
Uysal, S. D. (2011): "Three Essays on Doubly Robust Estimation Methods," . 4, 23, 29
Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. 39, 45
Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, MIT press. 12, 20
Zeileis, A. and T. Hothorn (2002): "Diagnostic Checking in Regression Relationships," *R News*, 2, 7–10. 34

Supplemental Appendix

SA.1 Rich covariates under conditional random assignment

Suppose that $X = (X_1, X_2)$ has two components and that Z is randomly assigned conditional on X_1. This could happen in a stratified experiment, where X_1 describes the strata. It could also happen in other settings, for example if judges Z are thought to be randomly assigned, but only conditionally on the day of the week, X_1. The rich covariates condition is still the same in this case: $\mathbb{E}[Z|X_1, X_2] = \mathbb{L}[Z|X_1, X_2]$. However, random assignment implies that rich covariates reduces to the requirement that

$$\mathbb{E}[Z|X_1] = \mathbb{L}[Z|X_1, X_2], \tag{33}$$

because Z is independent of X_2, given X_1. In some situations, it will be natural to control for X_1 nonparametrically, for example using strata or day-of-week indicators. If this is done, then (33) reduces further to the requirement that

$$\mathbb{L}[Z|X_1] = \mathbb{L}[Z|X_1, X_2]. \tag{34}$$

Condition (34) can be evaluated by regressing Z onto X_1 and X_2, then testing the null hypothesis that the coefficients on X_2 are zero.

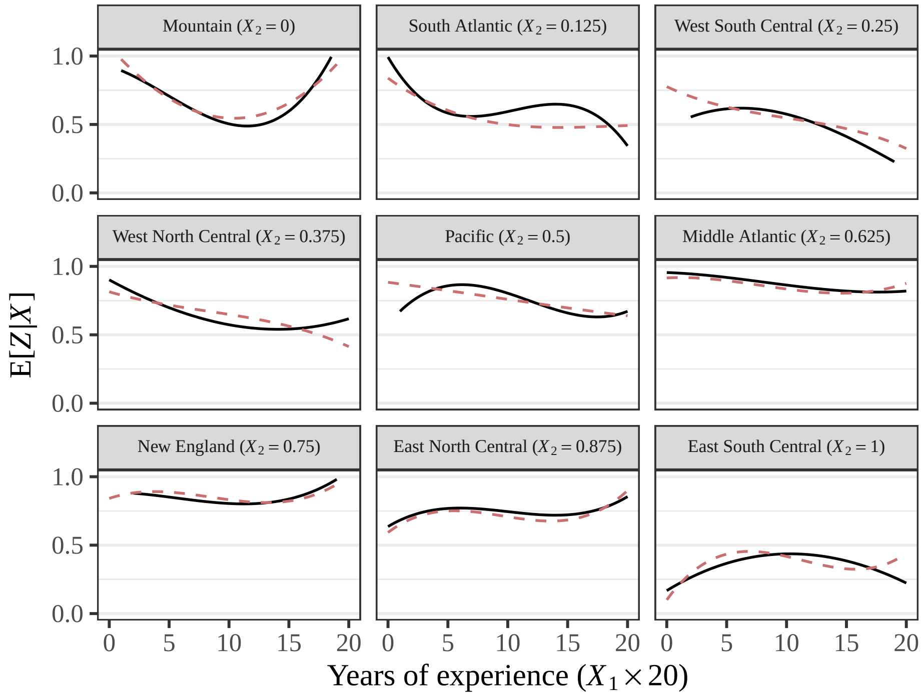
SA.2 Details on the simulation design

In this section, we discuss in detail how we constructed the DGP used in Section 5.4.

We set $X = (X_1, X_2)$ to be a two-dimensional vector of covariates, where X_1 takes many values and X_2 takes nine values. The support of X_1, which we vary in the simulations, is determined by a Halton sequence on $[0, 1]$, while the support of X_2 is $0, 1/8, 2/8, \ldots, 1$. The distribution of both X_1 and X_2 is taken to be uniform, with X_1 and X_2 independent.

We calibrate $\mathbb{E}[Z|X]$ to Card's data by setting Z to be the binary indicator for near four-year college, X_1 to be experience divided by 20, which is roughly the maximum in the data, and X_2 to be one of nine geographic regions. There are $9! = 362{,}880$ possible ways to map region to the numerical support of X_2. For each one, we regress Z onto a fully interacted cubic polynomial between X_1 and X_2, weighting each observation with $X = x$ by the inverse empirical probability that $X = x$. We select the region mapping that yields the regression with the smallest sum of squared residuals. The resulting

Figure SA.1: Relationship between college-presence instrument and covariates



Notes: This figure plots the mean of the college-presence instrument used by Card (1995), conditional on region and years of experience. The solid black line is the line of best fit in the data, obtained by regressing $\mathbb{E}[Z|X]$ on a set of region-specific cubic polynomials in years of experience. The dashed red line is $\mathbb{E}[Z|X = x]$ for the DGP in our simulations.

specification of $\mathbb{E}[Z|X]$ can be written as

$$\mathbb{E}[Z|X = (x_1, x_2)] = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \end{bmatrix} \begin{pmatrix} 1.07 & -2.71 & 7.61 & -5.87 \\ -1.96 & 6.69 & -10.64 & 8.32 \\ 1.72 & -1.91 & -1.65 & -3.19 \\ 0.21 & -8.30 & 20.76 & -9.71 \end{pmatrix} \begin{bmatrix} 1 \\ x_2 \\ x_2^2 \\ x_2^3 \end{bmatrix}, \quad (35)$$

which is linear in 16 terms. Figure SA.1 plots (35) for $20 \times x_1$ against region-specific cubic regressions of the four-year college indicator onto years of experience.

We generate the binary treatment, T, by the threshold-crossing equation

$$T = \mathbb{1}[U \leq p(Z)], \quad (36)$$

where U is distributed uniformly over $[0, 1]$, independently of Z and X. Assumption MON is satisfied under (36) (Vytlacil, 2002). We take $p(0) = .42$ and $p(1) = .54$, which matches the propensity score in Card's data when T is defined as 13 years or more of completed schooling (some college). The group indicator is determined directly from

(36) as

$$G = \begin{cases} \text{AT} & \text{if } U \leq p(0) \\ \text{CP} & \text{if } U \in (p(0), p(1)] \\ \text{NT} & \text{if } U > p(1) \end{cases} .$$

To generate potential outcomes $Y(t)$, we let

$$\mathbb{E}[Y(t)|G = g, X = x] = \theta_0' h(t|g, x)', \tag{37}$$

where $h(t|g, x)$ are basis functions that contain cubic terms in $x = (x_1, x_2)$ that vary freely with t and g. The coefficients on these basis functions, θ_0, are found as solutions to the optimization problem described ahead. The dimension of h (and θ_0) is $96 = 2\times3\times16$ for two treatment arms, three groups, and sixteen cubic polynomial coefficients for each group. We generate $Y(t)$ by adding a normal error with mean zero and variance .2 to (37). The variance of .2 is roughly equal to the sample variance of log wages in Card's data.

The optimization problem we use to find θ_0 is set up to match some key estimates in Card's data. To implement the problem, we utilize an observation from Mogstad et al. (2018) that many estimands can be written as weighted averages of θ_0. We write the weights in these weighted averages as $w\{\text{estimand}\}$. The form of w can be complicated, so we do not provide explicit expressions here, but they depend on h and the joint distribution of (G, T, X, Z), for which we use the distribution implied by the DGP through the above constructions when X_1 has 24 points of support. Having these linear-in-θ expressions is useful because it allows us to define the optimization problem as a convex quadratic program with linear constraints.

The objective of the optimization problem is to match a weighted average of treated outcomes for always-takers and average untreated outcomes for never-takers. Letting $\bar{Y}_{tz}$ denote the sample average of Y among the subpopulation with $T = t$ and $Z = z$ in Card's data, the objective we minimize is:

$$\Omega(\theta) \equiv \left(\bar{Y}_{10} - \theta' w\{\mathbb{E}[Y|T = 1, Z = 0]\}\right)^2 + \left(\bar{Y}_{01} - \theta' w\{\mathbb{E}[Y|T = 0, Z = 1]\}\right)^2 .$$

The constraints involve the following estimates from Card's data:

- $\bar{Y} \approx 6.26$ is the sample average of log wages.
- $\hat{\beta}_{\text{ols}} \approx .24$ is the OLS estimate of the coefficient on the some college indicator (defined as above) in a regression of log wage on some college, controlling for the covariates used by Card (1995, Table 3A, column (5)), which is the same

specification we consider in Section 6.1.

- $\hat{\beta}_{\text{iv}} \approx .66$ is the corresponding IV estimate where the near college indicator is used to instrument for some college.
- $\hat{\beta}_{\text{rich}} \approx .43$ is the DDML-PLIV estimate, constructed using the same DDML estimator as in the simulations.
- $\hat{\beta}_{\text{late}} \approx .20$ is the DDML estimate of the unconditional LATE, constructed using the same algorithms as the DDML-PLIV estimate.

We use these estimators to impose the following linear constraints on θ:

$$\theta' w\{\mathbb{E}[Y]\} = \bar{Y} \tag{38}$$

$$\theta' w\{\beta_{\text{ols}}\} = \hat{\beta}_{\text{ols}} \tag{39}$$

$$\theta' w\{\beta_{\text{iv}}\} = \hat{\beta}_{\text{iv}} \tag{40}$$

$$\theta' w\{\beta_{\text{rich}}\} = \hat{\beta}_{\text{rich}} \tag{41}$$

$$\theta' w\{\beta_{\text{late}}\} = \hat{\beta}_{\text{late}}. \tag{42}$$

We additionally constrain θ so that the implied values of $\mathbb{E}[Y(t)|X = x]$ are linear (Assumption LIN) to match the special cases discussed in Propositions 1 and 2:

$$\theta' w\{\mathbb{E}[Y(t)|X = x]\} = \vartheta_0(t) + \vartheta_1(t)x_1 + \vartheta_2(t)x_2 \quad \text{for all } t \text{ and } x, \tag{43}$$

where $\vartheta_1(t), \vartheta_2(t), \vartheta_3(t)$ are additional variables of optimization.[16] We also impose three additional constraints that restrict treatment effects:

$$\theta' w\{\mathbb{E}[Y(1) - Y(0)|G = g, X = x]\} \in [-2, 2] \quad \text{for all } g \text{ and } x. \tag{44}$$

$$\theta' w\{\mathbb{E}[Y(1) - Y(0)|G = \text{NT}]\} = 0 \tag{45}$$

$$\theta' w\{\mathbb{E}[Y(1) - Y(0)|G = \text{AT}, X = x]\} \geq 0 \quad \text{for all } x. \tag{46}$$

The overall optimization problem that we solve is then:

$$\theta_0 = \arg\min_{\theta,\vartheta} \ \Omega(\theta) \quad \text{s.t.} \quad (38)\text{–}(46). \tag{47}$$

The problem is a linearly-constrained convex quadratic program.

[16]In practice, we do this by restricting the nonlinear terms of $\theta' w\{\mathbb{E}[Y(t)|X = x]\}$ to be zero, so that the number of imposed constraints does not depend on the number of support points that X has.

SA.3 Additional figures and tables

Figure SA.2: Alternative weights for β_{iv} in the simulation DGP

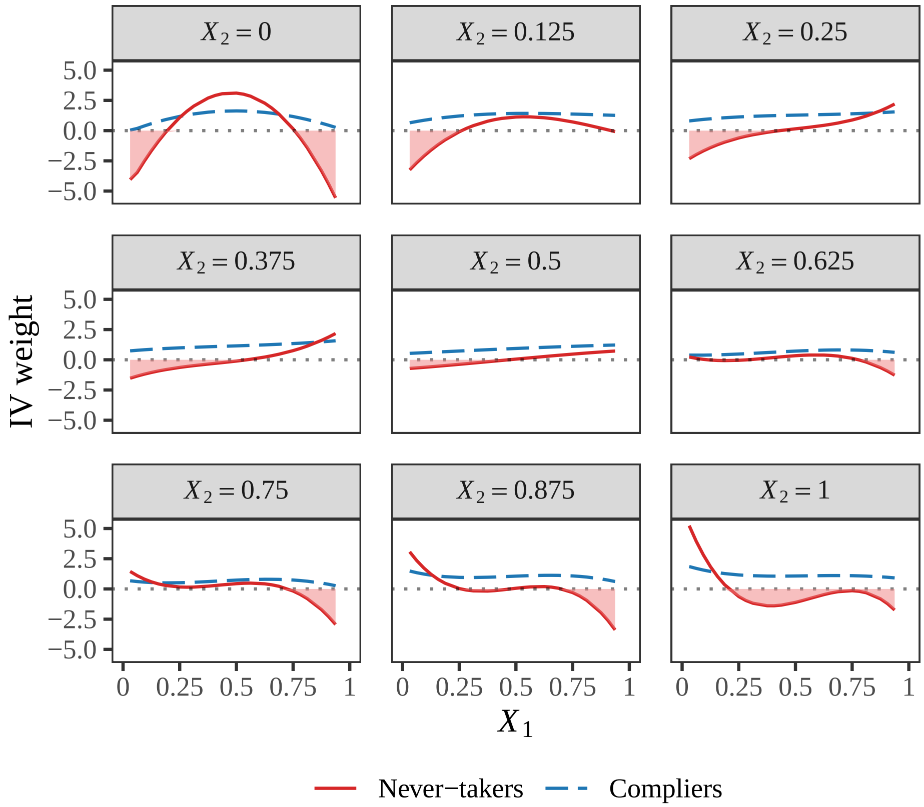


Figure SA.3: Population values of the estimands in the simulation DGP

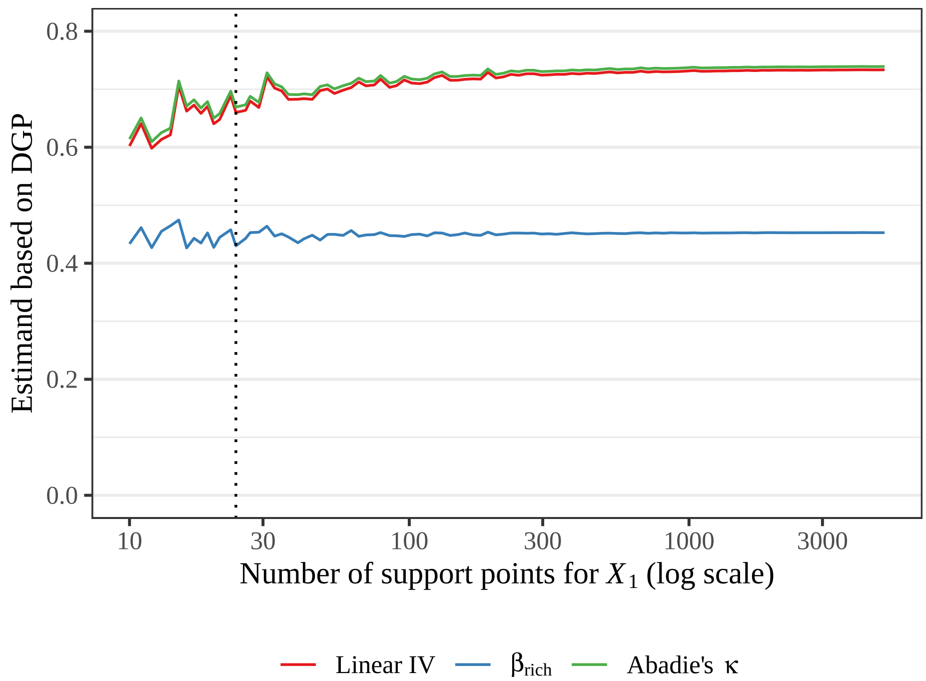


Table SA.1: Detailed simulation results

Estimator	Estimand	Mean (SD)	RMSE	10%	25%	Median	75%	90%
	$N = 500$, $\lvert\mathcal{X}_1\rvert = 24$							
Linear IV	0.660	0.720 (0.784)	0.836	–0.238	0.168	0.611	1.183	1.946
Correctly specified	0.430	0.483 (0.785)	0.786	–0.531	–0.055	0.407	0.955	1.625
Saturated	0.430	0.449 (1.251)	1.252	–1.077	–0.251	0.356	1.010	1.845
PLIV (DDML)	0.430	0.587 (0.774)	0.790	–0.447	0.051	0.507	1.056	1.700
Abadie's κ	0.669	0.729 (0.925)	0.972	–0.382	0.110	0.617	1.230	2.030
	$N = 500$, $\lvert\mathcal{X}_1\rvert = 100$							
Linear IV	0.709	0.838 (0.873)	0.957	–0.180	0.260	0.705	1.262	2.208
Correctly specified	0.445	0.542 (0.802)	0.808	–0.445	0.036	0.437	0.986	1.742
Saturated	0.445	0.546 (2.231)	2.234	–1.476	–0.442	0.424	1.280	2.866
PLIV (DDML)	0.445	0.677 (0.784)	0.818	–0.335	0.168	0.589	1.135	1.859
Abadie's κ	0.716	0.868 (0.936)	1.027	–0.284	0.297	0.730	1.348	2.206
	$N = 3{,}000$, $\lvert\mathcal{X}_1\rvert = 24$							
Linear IV	0.660	0.663 (0.249)	0.341	0.339	0.478	0.641	0.848	1.072
Correctly specified	0.430	0.434 (0.238)	0.238	0.098	0.262	0.409	0.601	0.803
Saturated	0.430	0.428 (0.245)	0.245	0.079	0.247	0.404	0.601	0.801
PLIV (DDML)	0.430	0.521 (0.242)	0.259	0.192	0.344	0.492	0.701	0.919
Abadie's κ	0.669	0.685 (0.290)	0.386	0.319	0.474	0.647	0.873	1.168
	$N = 3{,}000$, $\lvert\mathcal{X}_1\rvert = 1{,}000$							
Linear IV	0.731	0.748 (0.249)	0.386	0.399	0.553	0.749	0.922	1.118
Correctly specified	0.452	0.454 (0.232)	0.232	0.109	0.278	0.457	0.632	0.788
Saturated	0.452	0.451 (0.824)	0.824	–0.694	–0.046	0.381	0.985	1.604
PLIV (DDML)	0.452	0.546 (0.237)	0.255	0.196	0.364	0.546	0.730	0.899
Abadie's κ	0.737	0.768 (0.296)	0.433	0.359	0.554	0.755	0.968	1.234
	$N = 10{,}000$, $\lvert\mathcal{X}_1\rvert = 24$							
Linear IV	0.660	0.659 (0.133)	0.265	0.460	0.557	0.664	0.756	0.862
Correctly specified	0.430	0.427 (0.126)	0.126	0.235	0.330	0.430	0.529	0.620
Saturated	0.430	0.426 (0.127)	0.127	0.236	0.324	0.431	0.532	0.619
PLIV (DDML)	0.430	0.499 (0.128)	0.146	0.303	0.398	0.501	0.602	0.688
Abadie's κ	0.669	0.658 (0.135)	0.265	0.456	0.555	0.665	0.762	0.851
	$N = 10{,}000$, $\lvert\mathcal{X}_1\rvert = 3{,}000$							
Linear IV	0.733	0.737 (0.133)	0.314	0.555	0.621	0.735	0.837	0.928
Correctly specified	0.453	0.453 (0.127)	0.127	0.262	0.355	0.455	0.552	0.641
Saturated	0.453	0.440 (0.325)	0.326	–0.043	0.197	0.431	0.677	0.918
PLIV (DDML)	0.453	0.521 (0.126)	0.144	0.338	0.420	0.524	0.623	0.714
Abadie's κ	0.739	0.730 (0.145)	0.313	0.531	0.607	0.728	0.841	0.939

Notes: Simulations based on 500 replications.

Table SA.2: Sensitivity to covariate specification in Dube and Harish (2020)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV estimate	1.011	0.511	0.681	0.984	1.220	0.262	1.190	0.400
	(0.523)	(0.231)	(0.355)	(0.519)	(0.640)	(0.170)	(0.639)	(0.211)
	[0.011]	[0.005]	[0.029]	[0.015]	[0.013]	[0.142]	[0.014]	[0.039]
Polity fixed effects		✓				✓		✓
Decade fixed effects			✓			✓		✓
Missing gender control				✓			✓	✓
Previous monarch controls					✓		✓	✓

Notes: Clustered standard errors are reported in parentheses. Brackets contain p-values for the clustered wild bootstrap procedure implemented by Dube and Harish (2020) with 1000 replications. Column (8) replicates Table 3, column (3) of Dube and Harish (2020). The sample size is 3,586.

Table SA.3: Detailed results from all applications

Application	(1) $\hat{\beta}_{\text{ols}}$	(2) $\hat{\beta}_{\text{iv}}$, no X	(3) $\hat{\beta}_{\text{iv}}$	(4) $\hat{\beta}_{\text{rich}}$	(5) RESET p-value	(6) Included variables	(7) Sample size
Panel A. Illustrative examples							
Card (1995)	0.075	0.188	0.132	0.122	0.000	14	3,010
	(0.004)	(0.026)	(0.054)	(0.053)			
Nunn and Wantchekon (2011)	-0.203	-0.190	-0.271	-0.071	0.000	99	16,679
	(0.033)	(0.111)	(0.088)	(0.091)			
Dube and Harish (2020)	0.115	1.011	0.400	0.318	0.000	66	3,586
	(0.035)	(0.522)	(0.211)	(0.240)			
Panel B. IV survey							
Alesina and Zhuravskaya (2011)	-1.984	-5.727	-3.646	-2.919	0.182	14	97
	(0.639)	(1.289)	(1.307)	(1.115)			
Autor et al. (2013)	-0.171	-0.666	-0.596	-0.547	0.000	15	1,444
	(0.028)	(0.143)	(0.099)	(0.091)			
Becker and Woesmann (2009)	0.099	0.422	0.189	0.186	0.012	12	452
	(0.010)	(0.071)	(0.027)	(0.031)			
Bloom et al. (2012)	1.669	2.708	3.071	2.152	0.000	160	422
	(0.789)	(1.918)	(1.253)	(1.304)			
Condra et al. (2018)	-0.016	-0.135	-0.092	-0.097	0.923	18	410
	(0.007)	(0.128)	(0.047)	(0.067)			
Dal Bo et al. (2009)	0.027	-0.015	0.083	0.058	0.000	141	5,502
	(0.006)	(0.030)	(0.037)	(0.032)			
Dinkelman (2011)	-0.001	0.025	0.095	0.118	0.004	22	1,816
	(0.005)	(0.045)	(0.055)	(0.118)			
Dippel (2014)	-0.295	-0.676	-0.443	-0.462	0.000	44	182
	(0.048)	(0.326)	(0.103)	(0.235)			
Gilchrist and Sands (2016)	0.619	0.939	0.843	0.828	0.401	213	2,064
	(0.058)	(0.245)	(0.279)	(0.292)			
Hornung (2014)	1.741	5.437	3.380	0.892	0.004	10	150
	(0.287)	(4.180)	(1.137)	(0.773)			

Table SA.4: Details on specifications used in all applications

Study	Specification	Sample size	Num. included variables
Alesina and Zhuravskaya (2011)	Table 7, Panel A, Column (2)	97	14
Autor et al. (2013)	Table 3, Panel I, Column (6)	1,444	15
Becker and Woesmann (2009)	Table 3, Column (2)	452	12
Bloom et al. (2012)	Table 2, Column (7)	422	160
Card (1995)	Table 3, Panel A, Column (5)	3,010	14
Condra et al. (2018)	Table 2, Panel A, Column (3)	410	18
Dal Bo et al. (2009)	Table 5, Panel B, Column (3)	5,501	141
Dinkelman (2011)	Table 4, Column (8)	1,816	22
Dippel (2014)	Table 5, Panel B, Column (6)	182	44
Dube and Harish (2020)	Table 3, Column (3)	3,586	66
Gilchrist and Sands (2016)	Table 4, Column (6)	2,064	213
Hornung (2014)	Table 4, Column (5)	150	10
Nunn and Wantchekon (2011)	Table 6, Column (2)	16,679	99