EMPIRICAL DECOMPOSITION OF THE IV-OLS GAP WITH HETEROGENEOUS AND NONLINEAR EFFECTS

Shoya Ishimaru*

Abstract—This study proposes an econometric framework to interpret and empirically decompose the difference between instrumental variables (IV) and ordinary least squares (OLS) estimates given by a linear regression model when the true causal effects of the treatment are nonlinear in treatment levels and heterogeneous across covariates. I show that the IV-OLS coefficient gap consists of three estimable components: the difference in weights on the covariates, the difference in weights on the treatment levels, and the difference in identified marginal effects that arises from endogeneity bias. Applications of this framework to return-to-schooling estimates demonstrate the empirical relevance of this distinction in properly interpreting the IV-OLS gap.

I. Introduction

INSTRUMENTAL variables (IV) regression is the most common approach for estimating the causal effect of a potentially endogenous regressor. A standard empirical approach specifies the following linear model:

\[ Y = \beta X + W' \gamma + \epsilon, \]  

(1)

where \( Y \) is the outcome of interest, \( X \) is a scalar (multi-valued) treatment, and \( W \) is a vector of covariates. A standard econometric textbook takes the linear regression equation (1) as the true causal relationship and interprets the gap between the IV and ordinary least squares (OLS) coefficient estimates of \( X \) as a consequence of endogeneity bias associated with omitted variables, selection, or measurement error.

If these interpretations fail to provide a plausible explanation for the IV-OLS coefficient gap, empirical researchers often consider the possibility that it instead arises because of how the IV and OLS coefficients place different weights on different treatment margins or groups of individuals. This interpretation treats the regression equation (1) as a linear projection model rather than a causal model, allowing for the treatment effects to be heterogeneous and nonlinear in the true causal relationship. For example, Card (1995a, 1999, 2001) suggests that the positive IV-OLS gaps in many return-to-schooling studies could be explained by higher returns among credit-constrained individuals, who are more likely to be affected by cost-related instruments. In addition, researchers sometimes perform an OLS regression restricted to a sample of likely complier groups or complying treatment margins to examine the robustness of the IV-OLS coefficient gap to the weight difference. As a prominent example, Angrist and Krueger (1991) compute an OLS estimate of the return to schooling by restricting their sample to individuals with 9–12 years of schooling, who are expected to be most influenced by quarter-of-birth instruments. However, despite the widely recognized importance of the weight difference interpretation, relatively few studies attempt to formally quantify how much it actually matters for the IV-OLS coefficient gap.\(^1\)

In this study, I propose an econometric framework to quantify the sources of the IV-OLS coefficient gap. I begin my analysis by considering the following causal model:

\[ Y = g(X, W) + U, \]  

(2)

with a valid instrument \( Z \) that is uncorrelated with an unobservable \( U \) conditional on covariates \( W \). With no restriction on the structural function \( g \), the equation allows for the treatment effects to be heterogeneous across covariates \( W \) and nonlinear in treatment levels \( X \) in any manner.\(^2\) I demonstrate that the OLS and IV estimates based on the linear regression equation (1), when the true model is equation (2), represent different weighted averages of the marginal effects of the treatment they identify.

Using the weighted-average representations, I then decompose the IV-OLS coefficient gap into three estimable components. These are the following: (i) the covariate weight difference, being the difference in how the IV and OLS coefficients place weights on the covariates \( W \); (ii) the treatment-level weight difference, as the difference in how they place weights on treatment levels \( X \); and (iii) the endogeneity bias (or marginal effect difference), being the difference between the IV- and OLS-identified marginal effects originating from the correlation between the treatment \( X \) and the unobservable \( U \). For instance, in the return-to-schooling context, (i) arises from heterogeneous returns across observed personal backgrounds with different responses to the instrument, which motivates the conjecture of Card (1995a, 1999, 2001), (ii) follows from nonlinear returns across schooling levels with different sensitivity to the instrument, which motivates the robustness check of Angrist and Krueger (1991), and (iii) corresponds to endogeneity bias associated with omitted unobserved ability.

To implement the decomposition, I propose a two-step approach to estimate the “IV-weighted OLS” coefficients that serve as the intermediate points between the IV and OLS

\(^1\) Notable exceptions include Kling (2001); Lochner and Moretti (2001); Mogstad and Wiswall (2010); Løken et al. (2012); and Lochner and Moretti (2015).

\(^2\) The separability restriction on the equation, which is relaxed in section IIIB, rules out unobserved heterogeneity in the treatment effects.
coefficients. The first step predicts the conditional mean of the outcome $Y$ given $(X, W)$, and the second step uses the prediction to construct a dependent variable and performs a quasi-IV regression. Depending on the functional form restriction in the first step, the estimated OLS coefficients in the second step have the IV weights on the covariates or on both the covariates and treatment levels. Comparing the estimated IV-weighted OLS coefficients with the IV and OLS coefficients reveals the contributions of the weight difference components (i) and (ii) and the endogeneity bias component (iii). Standard statistical packages can compute the IV-weighted OLS estimates and perform statistical inference based on them.\(^4\)

In an extension, I consider a class of identification strategies that use an instrument $Z$ deterministic in the covariates $W$, as in difference-in-differences (DID) and regression discontinuity (RD) designs.\(^5\) I show that the weighted-average interpretation and the decomposition approach can also be applied to these setups, with only a small modification of the weight function.

It should be noted that my decomposition framework cannot fully isolate endogeneity bias in the presence of unobserved heterogeneity in the treatment effects. In an extension that relaxes the separability restriction on equation (2), I demonstrate that the marginal effect difference component (iii) captures not only endogeneity bias but also the unobservable-driven discrepancy between the IV-identified and the average marginal effects.\(^6\) While the weight difference components (i) and (ii) remain informative about the implications of observed heterogeneity and nonlinearity, it can be misleading to attribute the marginal effect difference component (iii) entirely to endogeneity bias.

Using my framework, I examine the return-to-schooling estimates using several common IV strategies. The first example employs geographic variation in college costs (Cameron & Taber, 2004; Carneiro et al., 2011). The second exploits a discontinuity in the minimum school-leaving age across cohorts (Oreopoulos, 2006). The third and final example uses DID variation in compulsory schooling laws across cohorts and regions (Acemoglu & Angrist, 2000). In these empirical examples, the weight difference components are found to be as important as the endogeneity bias component in explaining the IV-OLS coefficient gap. The direction or extent of endogeneity bias implied by the estimated IV-OLS gap differs entirely by taking into consideration how the two coefficients place weights on the different observed personal backgrounds and schooling margins.

A. Related Literature and Roadmap

This paper advances the literature on the interpretation and decomposition of linear regression coefficients by exploring a general and empirically relevant setting in which the treatment effects are nonlinear in treatment levels and heterogeneous across covariates. The local average treatment effect (LATE) interpretation proposed by Imbens and Angrist (1994) in a binary treatment context originates the idea that the IV coefficient is a weighted average of the marginal causal effects of the treatment. Angrist and Imbens (1995) and Angrist et al. (2000) extend the basic insights of the LATE interpretation to a multivalued treatment case. Yitzhaki (1996) and Angrist and Krueger (1999) suggest the analogous weighted-average interpretation of the OLS coefficient. Much of the focus of the literature has been on a univariate model with no or fixed covariates, which makes it difficult to immediately apply these results to empirical settings. While drawing on these existing results, my framework synthesizes them into an empirically relevant format and provides an estimable decomposition of the IV-OLS coefficient gap.

Motivated by the weighted-average interpretation developed in the literature, Mogstad and Wiswall (2010); Løken et al. (2012); and Lochner & Moretti (2015) propose the empirical decomposition of the IV-OLS coefficient gap into weight difference and endogeneity bias components. They consider a model in which the treatment effects are nonlinear in treatment levels but homogeneous across covariates. My framework generalizes these previous works by allowing the treatment effects to be heterogeneous across covariates. This is an empirically meaningful generalization, as empirical applications demonstrate the relevance of the covariate weight difference in interpreting the IV-OLS coefficient gap. While the linear IV regression is often advocated for its transparency (Angrist & Pischke, 2010), it is also criticized for its lack of a clear connection to an economic parameter of interest (Heckman & Urzua, 2010). A related strand of the literature based on the latter view aims to develop alternatives to the linear IV regression, including the policy-relevant treatment effect proposed by Heckman and Vytlacil (2001) and Carneiro et al. (2010). Nevertheless,

---

3. The concept of an IV-weighted OLS coefficient originates in Mogstad and Wiswall (2010), where they use the model $Y = g(X) + U$ and account for the treatment-level weight difference.

4. The Stata package that implements the decomposition, ivolsdec, is available from the Boston College Statistical Software Components (SSC) archive. Type “ssc install ivolsdec” in the Stata command window to install it.

5. Throughout this paper, the term “DID” refers to an identification strategy that exploits DID variation in the instrument using a two-way fixed effects regression, in which the treatment and the instrument can be non-binary. Examples include Acemoglu and Angrist (2000); Duflo (2001); Black et al. (2005), and a fuzzy DID setup considered in de Chaisemartin and D’Haultfoeuille (2018). The term “RD” refers to a fuzzy RD design, which is usually implemented as an IV regression.

6. This corresponds to a classic impossibility result in the setting with a binary treatment and no covariates, where endogeneity bias cannot generally be separated from the difference between the local and population average treatment effects.

7. The OLS interpretation is also explored by Angrist (1998); Aronow and Samii (2016); and Sloczynski (2022) in a binary treatment case with a focus on observed heterogeneity.

8. For the linear OLS regression, Sloczynski (2022) shows that the OLS coefficient of a binary treatment with covariates generally does not represent the average treatment effect (ATE), the average treatment effect on treated (ATT), or untreated (ATU).
many empirical researchers use the linear regression for its simplicity. My framework aims to provide useful diagnostics for empirical researchers who use the linear regression while recognizing its potential limitations.

The rest of the paper proceeds as follows. Section II presents the interpretation of the linear IV and OLS coefficients and proposes the decomposition of the IV-OLS coefficient gap. Section III extends these econometric results by exploring settings with alternative assumptions. Section IV proposes the estimators for the IV-weighted OLS coefficients for empirically performing the decomposition. Section V presents the results from the empirical applications, and section VI concludes. Online appendix presents the proofs of the theorems, explores additional econometric results, and provides details for the empirical applications.

II. Econometric Framework

A. Setup and Assumptions

I consider a random draw of \((Y, X, W, Z)\), where \(Y\) is a scalar outcome variable, \(X\) is a scalar treatment variable, \(W\) is a vector of covariates, and \(Z\) is a scalar instrument. Multinstrument two-stage least squares can fit into this setup by regarding the projection of the treatment \(X\) onto the instrument vector and covariates \(W\) (i.e., the first-stage predicted value) as a synthetic scalar instrument.9

Throughout this paper, I assume the existence of the first and second moments of any random variable. I use \(L_w(R) = w’E(WW’)^{-1}E(WR)\) to denote a linear projection of a random variable \(R\) onto \(W\) evaluated at \(W = w\) (i.e., a predicted value from a linear regression of \(R\) on \(W\)).10 I define \(\tilde{R} = R - L_w(R)\) to be a residual from the linear projection. Let \(m(x, w) = E[Y|X = x, W = w]\) be the conditional mean function of \(Y\) given \((X, W)\).

To assess the impact of the treatment \(X\) on the outcome \(Y\), a standard approach specifies a linear regression model:

\[
Y = \beta X + W’Y + \epsilon, \quad E(\epsilon W) = 0. \tag{3}
\]

Additional moment conditions \(E(\epsilon Z) = 0\) and \(E(\epsilon X) = 0\), respectively pin down the linear IV coefficient \(\beta_{IV}\) and OLS coefficient \(\beta_{OLS}\) as

\[
\beta_{IV} = E(\tilde{Y}\tilde{Z})/E(\tilde{X}\tilde{Z}), \tag{4}
\]

\[
\beta_{OLS} = E(\bar{Y}\bar{X})/E(\bar{X}^2). \tag{5}
\]

Note that I treat equation (3) merely as a statistical model to characterize \(\beta_{IV}\) and \(\beta_{OLS}\), which does not impose any assumption on the underlying causal relationship.

To consider the causal interpretation of the IV and OLS coefficients, I define \(Y(x)\) to be the potential outcome associated with the treatment level \(x\), which produces the observed outcome as \(Y = Y(X)\). The derivative \(Y'(x)\) is the marginal effect of the treatment in a causal sense. I make the following assumptions.

Assumption S. (Separability) The potential outcome is given by \(Y(x) = g(x, W) + U, E(U|W) = 0\).

Assumption C. (Continuous Treatment) The treatment \(X\) is continuously distributed on support \((x, \tau)\), with \(-\infty < x < \tau \leq \infty\).

Assumption D. (Regularity Conditions on Derivatives) Let \(V^b_a(f, w) = \int_{\min(a,b)}^{\max(a,b)} \frac{\partial}{\partial x} f(x, w) dx\) be the total variation of a function \(f(x, w)\) differentiable in \(x\) between points \(a\) and \(b\).

(i) \(g(x, w)\) is differentiable in \(x\) and \(E[V^X_{x_0}(g, W)^2] < \infty\) for some \(x_0 \in (x, \tau)\);

(ii) \(m(x, w)\) is differentiable in \(x\) and \(E[V^X_{x_0}(m, W)^2] < \infty\) for some \(x_0 \in (x, \tau)\).

Assumption IV. The instrument \(Z\) satisfies the conditions:

(i) (Exogeneity) \(E(U\tilde{Z}) = 0\); (ii) (Relevance) \(E(\tilde{X}\tilde{Z}) \neq 0\).

Assumption OLS. The treatment residual has positive variance, i.e., \(E(\tilde{X}^2) > 0\).

Assumption L. (Linearity of Conditional Means)

(i) \(E(Z|W)\) is linear in \(W\); (ii) \(E(X|W)\) is linear in \(W\).

Given assumption S, the marginal causal effect of the treatment is \(Y’(x) = \frac{\partial}{\partial x} g(x, W)\), which can be nonlinear in treatment levels \(x\) and heterogeneous across covariates \(W\). However, it rules out unobserved heterogeneity in the effect.11 Section IIIIB relaxes this assumption.

I make assumption C to focus on a continuous treatment case, which is merely for expository convenience.12 All econometric results can be applied to a discrete treatment case by extending \(g(x, w)\) to nonsupport points of the treatment.13 Assumption D concerns the derivatives of the structural function \(g\) and the conditional mean function \(m\). Assumption IV is a set of standard IV assumptions that require the instrument to be exogenous and relevant after controlling for covariates. Assumption OLS is a standard OLS assumption.

Assumption L is required for the exact weighted-average interpretation of the regression coefficients. A similar

9If a regression of \(X\) on a vector instrument \((Z_1, \ldots, Z_M)\) controlling for \(W\) yields a first-stage coefficient \((\pi_1, \ldots, \pi_M)\), I regard \(Z = \sum_{m=1}^{M} \pi_m Z_m\) as a synthetic scalar instrument. As my setting allows for negative IV weights, a partial monotonicity condition for each individual instrument \(Z_m\) as in Mogstad et al. (2021) is not required as long as the treatment \(X\) and the synthetic instrument \(Z\) are correlated conditional on \(W\).

10The covariate vector \(W\) includes unity as one of its elements.

11The literature on the nonparametric IV approach typically makes a similar separability assumption. See, for example, Newey and Powell (2003), Blundell et al. (2007), and Horowitz (2011).

12As indicated in the assumption, the support \((\alpha, \tau)\) can be unbounded. Any integral expression of \(x\) in this paper is taken over the support \((\alpha, \tau)\), which is kept implicit to simplify the exposition.

13For example, define \(g(x, w)\) at nonsupport points by a linear interpolation without loss of generality. Then, the derivatives and integrals of \(x\) in all econometric results can be replaced by the differences and summations of \(x\).
B. Weighted-Average Interpretation

I start by presenting a key theorem for interpreting the IV and OLS coefficients. The OLS part of the theorem is shown by Angrist and Krueger (1999) in a discrete treatment setting. I present it for completeness and for reinterpretation in my setting.

**Theorem 1.** The IV and OLS coefficients have a weighted-average interpretation as below:

(i) With assumptions $S$, $C$, $D$-(i), $IV$, and $L$-(i),
\[
\beta_{IV} = \int \int \frac{\partial}{\partial x} g(x, w) \omega_Z(x, w) dF_W(w)dx;
\]

(ii) (Angrist & Krueger, 1999) With assumptions $C$, $D$-(ii), OLS, and $L$-(ii),
\[
\beta_{OLS} = \int \int \frac{\partial}{\partial x} m(x, w) \omega_X(x, w) dF_W(w)dx.
\]

The weight function is given by $\omega_k(x, w) = E[1_{X\geq x}]R[w]/E(XR)$ for $R = Z, X$, which satisfies $\int \int \omega_R(x, w) dF_W(w)dx = 1$.

Theorem 1 implies that the IV and OLS coefficients are expressed as weighted averages of the marginal effects they identify. While the IV coefficient identifies a weighted average of the causal effects $\frac{\partial}{\partial x} g(x, w)$, the OLS coefficient identifies a weighted average of the slopes of the conditional mean function $\frac{\partial}{\partial x} m(x, w)$.

The relationship between the OLS-identified and IV-identified marginal effects, $\frac{\partial}{\partial x} m(x, w)$ and $\frac{\partial}{\partial x} g(x, w)$, can be expressed as
\[
\frac{\partial}{\partial x} m(x, w) = \frac{\partial}{\partial x} g(x, w) + \frac{\partial}{\partial x} E[U|X = x, W = w].
\]

Therefore the difference between the two marginal effects arises from endogeneity bias, that is, the correlation between the treatment $X$ and the unobservable $U$.

While endogeneity bias makes the IV and OLS coefficients differ, an important implication from theorem 1 is that the difference in the weight functions, $\omega_Z$ and $\omega_X$, also gives rise to the IV-OLS coefficient gap. To explore what makes the IV weight $\omega_Z$ and the OLS weight $\omega_X$ differ, let $\omega_R(x, w) dx$ be the marginal weight on covariates $w$ for $R = Z, X$. The marginal IV and OLS weights on $W = w$ are given by
\[
\omega_Z(w) = \text{Cov}(X, Z|W = w)/E(XZ),
\]
\[
\omega_X(w) = \text{Var}(X|W = w)/E(X^2).
\]

The IV weight $\omega_Z(\cdot)$ is proportional to the conditional covariance $\text{Cov}(X, Z|W = w)$, which is the product of the regression coefficient of $X$ on $Z$ given $W = w$ and the conditional variance $\text{Var}(Z|W = w)$. This means that covariates $W$ with greater sensitivity of the treatment to the instrument or larger variation in the instrument are weighted more. In contrast, the OLS weight $\omega_X(\cdot)$ is proportional to the conditional variance $\text{Var}(X|W = w)$. This implies that covariates $W$ with larger variation in the treatment are weighted more. Similarly, let $\omega_R(x, w) dF_W(w)$ be the marginal weight on the treatment level $X = x$ for $R = Z, X$.

The marginal IV and OLS weights on $X = x$ are given by
\[
\omega_Z(x) = E(1_{X\geq x}Z)/E(XZ),
\]
\[
\omega_X(x) = E(1_{X\geq x}X)/E(X^2).
\]

The IV weight $\omega_Z(x)$ is proportional to a regression coefficient of $1_{X\geq x}$ on the instrument $Z$, controlling for $W$. This implies that the more the instrument $Z$ influences the treatment $X$ at $x$, the more weighted the treatment level $x$ is. For example, if a compulsory schooling instrument $Z$ increases years of schooling $X$ through primary and secondary education but does not influence college education, the IV weight is expected to be positive with $x \leq 12$ and zero with $x > 12$.

In general, the IV weight $\omega_Z(x)$ is not guaranteed to be positive because the instrument can have positive effects on the treatment at some margins while having negative effects at others.

Interpreting the OLS weight expression is less straightforward, but appendix C.1 shows
\[
\omega_X(x) \propto \int \int (x_1 - x_2) dF_X(x_1|w) dF_X(x_2|w) dF_W(w).
\]

This implies that the OLS weight is proportional to the sum of differences between the pairs of conditionally independent observations $X_1, X_2 \overset{i.i.d.}{\sim} F_X|W(\cdot|w)$ with $X_1 \geq x > X_2$. Therefore the treatment level $x$ is weighted more if the

\[14\]Appendix B.1 relaxes assumption L to explore what happens when a good linear approximation is not feasible because of data limitations.

\[15\]If $X$ is discrete, the weight on $x$ represents a change in treatment levels between $x - 1$ and $x$. 

treatment \( X \) is densely distributed both above and below \( x \).\(^{16}\) As is evident from equation (11), the OLS weight \( \hat{\sigma}_X(x) \) on each treatment level \( x \) is nonnegative.

Theorem 1 can be considered as a generalization of Lochner and Moretti (2015), allowing for heterogeneity in the marginal effects across covariates. In fact, restricting \( g(x, w) \) to be additively separable in \( x \) and \( w \) yields a weighted-average expression comparable to theirs. If \( g(x, w) \) is additively separable, \( \frac{\partial g(x, w)}{\partial x} \) depends only on \( x \). This yields \( \beta_{IV} = \int \frac{\partial g(x, \cdot)}{\partial x} \hat{\sigma}_Z(x) dx \), which matches the weighted-average expression provided by proposition 1 in Lochner and Moretti (2015).

C. Related Work

Theorem 1 closely relates to many existing results in the literature. To provide further intuition for the weighted-average interpretation, I explore the relationship between my results and those in the literature.

Denote the IV and OLS coefficients conditional on \( W = w \) as

\[
\begin{align*}
\beta_{IV}(w) &= \text{Cov}(Y, Z|W = w)/\text{Cov}(X, Z|W = w), \\
\beta_{OLS}(w) &= \text{Cov}(Y, X|W = w)/\text{Var}(X|W = w).
\end{align*}
\]

The following result from Lochner and Moretti (2015) shows that the IV and OLS coefficients, \( \beta_{IV} \) and \( \beta_{OLS} \), can be viewed as weighted averages of the covariate-specific coefficients, \( b_{IV}(w) \) and \( b_{OLS}(w) \).

**Theorem 2.** (Lochner & Moretti, 2015)

(i) With assumptions IV-(ii) and L-(i),

\[
\beta_{IV} = \int b_{IV}(w) \hat{\sigma}_Z(w) dF_W(w);
\]

(ii) with assumptions OLS and L-(ii),

\[
\beta_{OLS} = \int b_{OLS}(w) \hat{\sigma}_X(w) dF_W(w).
\]

The weighted-average interpretation of the covariate-specific coefficients, \( b_{IV}(w) \) and \( b_{OLS}(w) \), can be derived by applying theorem 1 conditional on \( W = w \). This result originates in Yitzhaki (1996) and Schechtman and Yitzhaki (2004).

**Theorem 3.** Let \( \omega_R(x|w) = \omega_R(x, w)/\overline{\omega_R}(w) \) be the conditional weight on the treatment level \( x \) given \( W = w \).

(i) (Schechtman & Yitzhaki, 2004) With assumptions S, C, D-(i), \( \text{Cov}(U, Z|W = w) = 0 \), and \( \text{Cov}(X, Z|W = w) \neq 0 \),

\[
b_{IV}(w) = \int \frac{\partial}{\partial x} g(x, w) \omega_Z(x|w) dx.
\]

(ii) (Yitzhaki, 1996) With assumptions C, D-(ii), and \( \text{Var}(X|W = w) > 0 \),

\[
b_{OLS}(w) = \int \frac{\partial}{\partial x} m(x, w) \omega_X(x|w) dx.
\]

The weighted-average expression for the covariate-specific IV coefficient \( b_{IV}(w) \) could also be derived as a special case of the results from Angrist and Imbens (1995); Angrist et al. (2000); and Heckman et al. (2006), which consider a more general setting that allows for unobserved heterogeneity in the treatment effects.\(^{17}\)

Synthesizing these existing results, theorem 1 can be divided into two components: the linear IV and OLS coefficients, \( \beta_{IV} \) and \( \beta_{OLS} \), are weighted averages of the covariate-specific coefficients, \( b_{IV}(w) \) and \( b_{OLS}(w) \), with different weights on covariates (theorem 2); and the covariate-specific coefficients are weighted averages of the identified marginal effects, \( \frac{\partial}{\partial x} g(x, w) \) and \( \frac{\partial}{\partial x} m(x, w) \), with different weights on treatment levels (theorem 3).

D. Decomposing the IV-OLS Coefficient Gap

The weighted-average interpretation of the IV and OLS coefficients in theorems 1–3 motivates the decomposition of the IV-OLS coefficient gap \( \beta_{IV} - \beta_{OLS} \). I decompose the gap into the following three components:

\[
\Delta_{CW} = \int b_{OLS}(w) (\overline{\omega}_Z(w) - \overline{\omega}_X(w)) dF_W(w),
\]

\[
\Delta_{TW} = \int \int \frac{\partial}{\partial x} m(x, w) (\omega_Z(x|w) - \omega_X(x|w)) \times \overline{\sigma}_Z(w) dF_W(w) dx,
\]

\[
\Delta_{ME} = \int \int \left( \frac{\partial}{\partial x} g(x, w) - \frac{\partial}{\partial x} m(x, w) \right) \times \omega_Z(x, w) dF_W(w) dx.
\]

\(^{16}\)Yitzhaki (1996) derives the OLS weight function in a simpler case with no covariates and shows that the weight function is \( \overline{\sigma}_X(x) \propto (b-x)(x-a) \) if \( X \) is uniformly distributed over \([a, b]\). The weights are zero at both ends of the support despite flat density because no pair of observations can sandwich the endpoints.

\(^{17}\)Angrist and Imbens (1995) allow for covariates in a special case with a “saturated” first stage, where \( Z = E[X|Z_1, \ldots, Z_M, W] \) is generated from an instrument vector \((Z_1, \ldots, Z_M)\) and a covariate vector \( W \) that both consist of indicators for disjoint groups.
The first component $\Delta_{CW}$ which I call “the covariate weight difference,” corresponds to how differently the IV and OLS coefficients place weights on the covariates. The second component, $\Delta_{TW}$, which I call “the treatment-level weight difference,” captures how differently the IV and OLS coefficients place weights on the treatment levels, conditional on the covariates. The third component, $\Delta_{ME}$, which I refer to as “the endogeneity bias” or “the marginal effect difference,” captures the difference between the IV- and OLS-identified marginal effects, which arises from the endogeneity of the treatment as in equation (6). The sum of the three components is the IV-OLS gap, that is, $\beta_{IV} - \beta_{OLS} = \Delta_{CW} + \Delta_{TW} + \Delta_{ME}$.

This decomposition departs from the OLS coefficient $\beta_{OLS}$ and arrives at the IV coefficient $\beta_{IV}$ by first changing the weights and then the marginal effects. This order follows an idea of the “IV-weighted OLS” approach adopted by Mogstad and Wiswall (2010) and Lochner and Moretti (2015). The advantage of this approach is that the decomposition is always feasible. The decomposition requires knowledge of the following IV-weighted OLS coefficients as intermediate points:

$$
\beta_C = \int b_{OLS}(w) \omega_Z(w) dF_W(w),
$$

(17)

$$
\beta_{CT} = \int \frac{\partial}{\partial x} m(x, w) \omega_Z(x, w) dF_W(w) dx.
$$

(18)

The first coefficient $\beta_C$ is the OLS coefficient with the IV weight on covariates, while the second $\beta_{CT}$ is the OLS coefficient with the IV weight on both covariates and treatment levels. By construction, $\Delta_{CW} = \beta_C - \beta_{OLS}$, $\Delta_{TW} = \beta_{CT} - \beta_C$, and $\Delta_{ME} = \beta_{IV} - \beta_{CT}$. These coefficients are always well-defined because $\omega_Z(x, w) \neq 0$ implies $\omega_X(x, w) > 0$.

This is not a unique order in which the IV-OLS gap can be decomposed, as is the case with Blinder-Oaxaca type decomposition methods. For example, one can instead account for the marginal effect difference first, then the weight differences. However, this alternative order requires knowledge of the “OLS-weighted IV” coefficient, i.e., $\int \int \frac{\partial}{\partial x} g(x, w) \omega_X(x, w) dF_W(x) dx$, as an intermediate point. This coefficient is not always identified because the instrument may have no variation or no impact on the treatment at some $(x, w)$, even with $\omega_X(x, w) > 0$. This makes the OLS-weighted IV approach less practical despite its potential theoretical appeal.\(^{19}\)

### III. Extensions

I extend my econometric framework in several directions. Section IIIA considers identification strategies based on DID and RD designs, in which the instrument $Z$ is deterministic in the covariates $W$. Section IIIB relaxes assumption S and allows for unobserved heterogeneity in the treatment effects. Appendix B explores additional extensions that consider a setting without assumption L, a setting with an invalid instrument, and a setting with DID or RD designs in the presence of unobserved heterogeneity.

A. Identification Based on DID or RD Designs

Assumption L does not hold by construction for two important identification strategies: DID and RD designs. I explore the weighted-average interpretation in these cases with an alternative set of assumptions.

With a DID-based identification strategy, each observation belongs to a particular group $g \in \{1, \ldots, G\}$ and period $t \in \{1, \ldots, T\}$, and the instrument $Z$ is constant within each $(g, t)$. The regression equation (3) can be written as

$$
Y = \beta X + \sum_{g=1}^{G} \gamma g D_g + \sum_{t=1}^{T-1} \delta_t D_t + \epsilon,
$$

where $d_g$ indicates membership to a group $g$ and $D_t$ indicates membership to a period $t$. By construction, the instrument $Z$ is a deterministic and nonlinear function of the covariate vector $W = (d_1, \ldots, d_G, D_1, \ldots, D_{T-1})$. This does not satisfy the requirement by assumption L that $E(Z | W)$ should be linear in $W$. An empirical example from Acemoglu and Angrist (2000) in section VC fits into this setting.

With an RD-based identification strategy using a running variable $C$ with a cutoff $c$, the instrument is $Z = 1_{C \geq c}$. The regression equation (3) can be written as

$$
Y = \beta X + \sum_{k=1}^{K} \gamma_k p_k (C) + \epsilon,
$$

where $(p_1, \ldots, p_K)$ is a set of basis functions with $p_k(c) = 0$.\(^{20}\) The instrument $Z$ is a deterministic and nonlinear function of the covariate vector $W = (p_1(C), \ldots, p_K(C))$. An empirical example from Oreopoulos (2006) in section VB fits into this setting.

The definition of the linear IV and OLS coefficients follows equations (4) and (5). I rule out a “sharp” DID or RD setup with $X = Z$, in which the IV and OLS coefficients are identical by construction. Instead of assumption L, I make the following assumption to represent DID- and RD-based identification strategies.

**Assumption LS.** (Linear structural function) $g(x, w)$ is linear in $w$ for any $x \in (x, F)$.

In a DID, this assumption corresponds to a parallel trend assumption, indicating that the group membership $d_g$ and the time membership $D_t$ additively affect a potential outcome.

\(^{19}\)Loeken et al. (2012) propose a mixture of the IV-weighted OLS and the OLS-weighted IV to make decomposition results independent of whether the marginal effect or the weight difference is accounted for first. This approach has the same identification issue as the OLS-weighted IV approach.

\(^{20}\)Given that an RD with local polynomials can be interpreted as a kernel-weighted version of an RD with global polynomials, my description focuses on a global polynomial case. The IV weight function should be multiplied by a kernel weight in a local polynomial case.
In an RD, this assumption implies that the relationship between a potential outcome and a running variable $C$ is well approximated by a linear combination of the basis functions $(p_1(C), \ldots, p_K(C))$ around $C = c$, which also implies continuity at $C = c$. In these settings, assumption IV-(i) follows from $Var(Z|W) = 0$, as $E(UZ) = E(E(U|W)Z) = 0$.

With assumption L replaced by assumption LS, the weighted-average interpretation of the IV coefficient remains valid with a modified weight function expression.

**Theorem 4.** With assumptions S, C, D-(i), IV, and LS, the IV coefficient $\beta_{IV}$ is given by

$$\beta_{IV} = \int \int \frac{\partial}{\partial x} g(x, w) \omega_{Z}(x, w) dF_{W}(w) dx,$$

where $\omega_{Z}(x, w) = L_w(1_{x \geq Z})/E(\tilde{X}Z)$ satisfies $\int \int \omega_{Z}(x, w) dF_{W}(w) dx = 1$.

Note that there are infinitely many weight functions other than $\omega_{Z}$ that can make this weighted-average expression. In particular, $\omega_{Z}(x, w) + h(x, w) - L_w(h(x, W))$ can also be a weight function, where $h(x, w)$ is any nonlinear function. This property is a mere artifact of assumption LS, which makes the marginal effect $\frac{\partial}{\partial x} g(x, w)$ linear in $w$ and orthogonal to any linear projection residual. Therefore it is most reasonable to use the linearly projected function $\omega_{Z}$ and omit the redundant variation in weights.\(^{21}\)

Given the weighted-average interpretation provided by theorem 4, it is possible to define the weight difference and the marginal effect difference components comparable to equations (14)–(16), using $\omega_{Z}$ instead of $\omega_{Z}$ as the IV weight function.\(^{22}\) Note that the marginal weights on covariate $W = w$ and treatment level $X = x$ are given by

$$\bar{\omega}_{Z}(w) = \int \omega_{Z}(x, w) dx = L_w(\tilde{X}Z)/E(\tilde{X}Z),$$

$$\bar{\omega}_{Z}(x) = \int \omega_{Z}(x, w) dF_{W}(w) = E(1_{x \geq Z})/E(\tilde{X}Z).$$

While the treatment-level weight $\bar{\omega}_{Z}(x)$ is identical to $\bar{\omega}_{Z}(x)$ defined in equation (9), the covariate weight $\bar{\omega}_{Z}(w)$ has a different expression from $\bar{\omega}_{Z}(w)$ defined in equation (7).

**B. Unobserved Heterogeneity**

Assumption S rules out unobserved heterogeneity in the marginal effects of the treatment. This is a common but strong assumption. Because the distinction between observables and unobservables arises merely from data availability, it is natural to consider heterogeneity in both observed and unobserved dimensions in the econometric model. Allowing for unobserved heterogeneity in the marginal effects $Y’(x)$, I define the average marginal effect (AME) as $\tau(x, w) = E[Y’(x)|W = w]$. With assumption S, the AME reduces to $\tau(x, w) = \frac{\partial}{\partial x} g(x, w)$.

Removing assumptions S, D-(i), and IV-(i), I make the following set of assumptions about the potential outcome process $Y(x)$.

**Assumption P.** The potential outcome process $Y(x)$ satisfies the following conditions:

(i) (Conditional Independence) $\text{Cov}(Y(x), Z|W) = 0$ for any $x \in (\bar{x}, \bar{x})$;

(ii) (Differentiability) $Y’(x)$ exists for any $x \in (\bar{x}, \bar{x})$ and the total variation $\text{Var}_{\lambda}(Y(x)) = \int_{ \text{min}|x|}^{ \text{max}|x|} |Y’(x)| dx$ satisfies $\text{Var}_{\lambda}(Y(X)) < \infty$ for some $\lambda \neq 0$.

Assumption P-(i) replaces IV-(i), and this is a standard exogeneity assumption.\(^{23}\) Assumption P-(ii) replaces D-(i). Unobserved heterogeneity in the treatment effects arises when $Var(Y(x)|W = w) > 0$. To rule out negative weights, some studies in the IV literature specify the potential treatment process and assume it to be monotonic in the instrument. On the other hand, I do not impose the monotonicity condition and maintain assumption IV-(ii) (i.e., $E(\tilde{X}Z) \neq 0$), thereby allowing for negative weights.\(^{24}\)

The following theorem extends the weighted-average interpretation of the IV coefficient provided by theorem 1.

**Theorem 5.** With assumptions P, C, IV-(ii), and L-(i), the IV coefficient $\beta_{IV}$ is given by

$$\beta_{IV} = \int \int \tau_{IV}(x, w) \omega_{Z}(x, w) dF_{W}(w) dx,$$

where the IV-identified marginal effect $\tau_{IV}(x, w)$ at each $(x, w)$ is given by

$$\tau_{IV}(x, w) = E[Y’(x), (Y’(x)|x, w) | W = w]$$

with $\lambda(t|x, w) = \text{Cov}(\{X_{\geq Z}, Y’(x)|x = t, w = w\})$ and $E[\lambda(Y’(x)|x, w)|W = w] = 1$.

Theorem 5 implies that the IV coefficient is a weighted average of the causal effects $Y’(x)$. Although treatment levels $x$ and covariates $w$ are weighted exactly in the same manner as in theorem 1, unobserved heterogeneity influences how the effects $Y’(x)$ are weighted. In particular, the difference between the IV-identified marginal effect $\tau_{IV}(x, w)$ and the AME $\tau(x, w)$ can be written as

$$\tau_{IV}(x, w) - \tau(x, w) = \text{Cov}(Y’(x), \lambda(Y’(x)|x, w) | W = w).$$

\(^{23}\)This setting implicitly rules out any direct causal effects of the instrument on the potential outcome. Some studies explicitly specify the potential outcome $Y(x, z)$ to be a function of the potential instrument assignment $z$ and then assume $Y(x, z)’ = Y(x, z)$ for any $z \neq z’$.

\(^{24}\)Appendix C.3 considers the weighted-average interpretation under the monotonicity condition and explores its relationship with the LATE interpretation in Angrist and Imbens (1995) and the marginal treatment effect interpretation in Heckman et al. (2006).
This difference arises from unobservable-driven covariance between the treatment effects $Y'(x)$ and the treatment responses to the instrument. For example, suppose some unobservables positively influence $Y'(x)$ and make the treatment more responsive to the instrument. As the weight function $\lambda(t|x, w)$ represents how strongly the instrument $Z$ induces a transition of the treatment $X$ from below $x$ to above $x$ conditional on $Y'(x) = t$ and $W = w$, $\tau_{IV}(x, w) > \tau(x, w)$ results from a greater emphasis on a higher $Y'(x)$. This difference corresponds to the discrepancy between the LATE and ATE in a binary treatment setting with no covariates, although the difference is specific to each treatment level $x$ and covariate value $w$ in this case.

Given the weighted-average interpretation, the decomposition in equations (14)–(16) remains valid by replacing the marginal effect difference component in equation (16) with

$$
\Delta_{ME} = \int \int \left( \tau_{IV}(x, w) - \frac{\partial}{\partial x} m(x, w) \right) \times \omega_Z(x, w) dF_W(w) dx.
$$

However, this component no longer represents endogeneity bias alone. In particular, equation (19) can be further decomposed as

$$
\Delta_{ME} = \int \int \left( \tau(x, w) - \frac{\partial}{\partial x} m(x, w) \right) \omega_Z(x, w) dF_W(w) dx + \int \int (\tau_{IV}(x, w) - \tau(x, w)) \omega_Z(x, w) dF_W(w) dx.
$$

The first term captures endogeneity bias through the difference between the AME $\tau(x, w)$ and the slope of the conditional mean function $\frac{\partial}{\partial x} m(x, w)$. The second term represents the unobservable-driven weight difference discussed above.

In a nonbinary treatment setting, it is well known that the AME $\tau(x, w)$ or analogous causal objects cannot be identified without any restriction on unobserved heterogeneity or causal structure.\(^{25}\) Thus, endogeneity bias and the unobservable-driven weight difference cannot be identified separately in general. Nevertheless, it remains helpful to isolate the covariate and treatment-level weight difference components using the decomposition, rather than observing only the raw IV-OLS gap.

### IV. Estimation and Inference

#### A. IV-Weighted OLS Estimators

Performing the decomposition proposed in section II requires estimators of the IV-weighted OLS coefficients defined in equations (17) and (18), which serve as intermediate points between the IV and OLS estimates. To define the estimators, let $(Y_i, X_i, Z_i, W_i)_{i=1}^N$ be an i.i.d. random sample that satisfies the set of assumptions in section II. The most natural estimators for the IV-weighted OLS coefficients are their direct data counterparts:

$$
\hat{\beta}_C = \frac{1}{N} \sum_{i=1}^N \int \hat{b}_{OLS}(W_i) \hat{\omega}_Z(x, W_i) dx,
$$

$$
\hat{\beta}_{CT} = \frac{1}{N} \sum_{i=1}^N \int \frac{\partial}{\partial x} \hat{m}(x, W_i) \hat{\omega}_Z(x, W_i) dx,
$$

using some estimators $(\hat{b}_{OLS}, \hat{m}, \hat{\omega}_Z)$ for $(b_{OLS}, m, \omega_Z)$.

While there are various choices for these estimators $(\hat{b}_{OLS}, \hat{m}, \hat{\omega}_Z)$, I focus on the most practical setting. I assume that $\hat{m}$ is a consistent estimator for $m$ estimated by minimizing

$$
\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{m}(X_i, W_i))^2,
$$

and that $\hat{m}(x, w)$ is given in a series form

$$
\hat{m}(x, w) = \sum_{k=1}^{K_N} \hat{m}_k(w) p_k(x), \quad \hat{m}_k(w) = \sum_{\ell=1}^{L_N^{(k)}} \hat{\beta}_{k\ell} q_{k\ell}(w),
$$

where $\{p_k(x), q_{k1}(w), \ldots, q_{kL_N^{(k)}}(w)\}_{k=1}^{K_N}$ is a set of basis functions chosen by the researcher. The numbers of the basis functions $K_N$ and $(L_N^{(k)})_{k=1}^{K_N}$ are constant in a parametric approach, while they may increase with the sample size $N$ with a nonparametric sieve estimation adopted. While $b_{OLS}$ can be any consistent estimator for $b_{OLS}$, for concreteness I assume that $\hat{b}_{OLS}$ is also estimated by the series form equation (22) with $K_N = 2$ and $(p_1(x), p_2(x)) = (1, x)$, in which $\hat{\omega}_Z(w)$ corresponds to $b_{OLS}(w)$. In addition, as the estimator for $\omega_Z(x, W_i)$, I consider its sample analogue $\hat{\omega}_Z(x, W_i) = \frac{\hat{I}_{X_i \geq x}}{\sum_{j=1}^N \hat{I}_{X_j \geq x}} Z_j$.

Then, equations (20) and (21) can be rewritten as

$$
\hat{\beta}_C = \frac{\sum_{i=1}^N \hat{b}_{OLS}(W_i) \hat{X}_i \hat{Z}_i}{\sum_{i=1}^N \hat{X}_i \hat{Z}_i},
$$

$$
\hat{\beta}_{CT} = \frac{\sum_{i=1}^N \left( \sum_{k=1}^{K_N} \hat{m}_k(W_i) \hat{P}_k(x) \right) \hat{Z}_i}{\sum_{i=1}^N \hat{X}_i \hat{Z}_i},
$$

where $P_k = p_k(X_i)$. Therefore the following two-step procedure gives the IV-weighted OLS estimates.

**Step 1:** Estimate $\hat{b}_{OLS}(w)$ and $\{\hat{m}_k(w)\}_{k=1}^{K_N}$ using the least-squares method with the series specification equation (22).

**Step 2:** Regress $Y_{2i} = \sum_{k=1}^{K_N} \hat{m}_k(W_i) \hat{P}_k(x)$ on $(X_i, W_i)$ instrumenting $X_i$ with $Z_i$, which yields $\hat{\beta}_{CT}$ as the coefficient of $X_i$. To estimate $\hat{\beta}_C$, let $Y_{2i} = \hat{b}_{OLS}(W_i) \hat{X}_i$ and perform the same regression.
Note that this two-step approach naturally generalizes the one proposed by Lochner and Moretti (2015) by allowing for a more general functional form of \( \hat{m}(x, u) \). In fact, \( \hat{\beta}_{CT} \) given in Step 2 is identical to the reweighted OLS estimator in Lochner and Moretti (2015) if \( \hat{m}(x, u) \) is specified in Step 1 to be additively separable in \( x \) and \( u \).

**DID- or RD-type instrument case.** The setting considered in section IIIA requires a small modification of the estimators due to a difference in the weight functions. Suppose that the weight function \( \omega^*_Z(x, W_i) \) defined in theorem 4 is estimated by its sample analogue

\[
\hat{\omega}^*_Z(x, W_i) = \frac{\hat{L}_W(1_{x \geq z} \tilde{Z}_i)}{\frac{1}{N} \sum_{j=1}^{N} 1_{x \geq z} \tilde{Z}_j},
\]

where \( \hat{L}_w(R_i) = w' \left( \sum_{j=1}^{N} W_j W_j^{-1} \right)^{-1} \left( \sum_{j=1}^{N} W_j R_j \right) \) is the linear projection estimate. Using \( \omega^*_Z \) instead of \( \omega_Z \) in deriving equations (23) and (24) yields

\[
\hat{\beta}_C = \frac{\sum_{i=1}^{N} \hat{L}_W \left( \hat{b}_{OLS}(W_i) \right) X_i \tilde{Z}_i}{\sum_{i=1}^{N} \hat{x}_i \tilde{Z}_i},
\]

\[
\hat{\beta}_{CT} = \frac{\sum_{i=1}^{N} \left( \sum_{k=1}^{K} \hat{L}_W \left( \hat{a}_k(W_i) \right) P_{ik} \right) \tilde{Z}_i}{\sum_{i=1}^{N} \hat{x}_i \tilde{Z}_i}.
\]

These expressions differ from equations (23) and (24) in two ways. First, \( \hat{b}_{OLS}(W_i) \) and \( \hat{a}_k(W_i) \) are linearly projected onto \( W_i \). In practice, \( \hat{b}_{OLS}(W_i) \) and \( \hat{a}_k(W_i) \) may be specified to be linear at the outset, rather than estimating them from a flexible nonlinear specification first and then linearly predicting them. Second, \( X_i \) and \( P_{ik} \) instead of \( \hat{x}_i \) and \( P_{ik} \) enter the numerators. These differences slightly change Step 2 in estimating the IV-weighted OLS estimators as follows.

**Step 2 (DID- or RD-type Instrument):** Regress \( Y_{2i} = \sum_{k=1}^{K} \hat{L}_W \left( \hat{a}_k(W_i) \right) P_{ik} \) on \( (X_i, W_i) \) instrumenting \( \hat{x}_i \) with \( Z_i \), which yields \( \hat{\beta}_{CT} \) as the coefficient of \( X_i \). To estimate \( \hat{\beta}_C \), let \( Y_{2i} = \hat{L}_W \left( \hat{b}_{OLS}(W_i) \right) X_i \) and perform the same regression.

**B. Asymptotic Properties**

Asymptotic properties of the IV-weighted OLS estimator can be derived using the standard econometric results for a two-step estimator. The following discussion focuses on the case in which the first step is parametric, since Ackerberg et al. (2012) illustrate that a semiparametric two-step estimator that uses a series approximation in the first step can be treated as if it were a parametric estimator for the purpose of standard error computation.

Using the standard formula for a parametric two-step estimator (Newey and McFadden, 1994, Section 6), appendix C.4 derives the asymptotic equivalence:

\[
\sqrt{N} \left( \hat{\beta}_{CT} - \beta_{CT} \right) \to_p \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( v_1 \tilde{Z}_i + v_2 \tilde{Z}_i \right),
\]

where \( v_{ij} = Y_i - \sum_{k=1}^{K} a_k(W_i) P_{ik} \) is a residual from the first step, \( v_{2i} = \tilde{Y}_i - \beta_{CT} \tilde{X}_i \) is a residual from the second step, and \( \tilde{Z}_i \) is a predicted value of \( Z_i \) given by the first step using \( \tilde{X}_i \) instead of \( Y_i \) as the dependent variable.\(^{26}\) Most statistical packages can estimate the standard error of \( \hat{\beta}_{CT} \) by estimating the standard error of the right-hand side of equation (27) under a certain distributional assumption (heteroscedasticity-robust, clustered, etc.). For example, in an i.i.d. heteroscedastic case, the asymptotic variance of \( \sqrt{N} \left( \hat{\beta}_{CT} - \beta_{CT} \right) \) is given by \( V_{\hat{\beta}_{CT}} = E \left( \tilde{X}_i \tilde{Z}_i \right)^{-1} E \left( (v_1 \tilde{Z}_i + v_2 \tilde{Z}_i)^2 \right) \) and the standard error of \( \hat{\beta}_{CT} \) can be estimated using the sample analogue of \( \sqrt{N} V_{\hat{\beta}_{CT}} \). Estimating the standard error of \( \hat{\beta}_C \) can follow the same procedure, as it is a special case with \( K = 2 \).

**C. Testing the Treatment Endogeneity**

Testing the significance of the marginal effect difference component \( \Delta_{ME} = \beta_{IV} - \beta_{CT} \) serves as a generalized Durbin-Wu-Hausman (DWH) test that is robust to the non-linearity and observed heterogeneity of the treatment effects. This test further extends the generalized DWH test proposed in Lochner and Moretti (2015) by allowing for observed heterogeneity of the effects. Since \( \hat{\beta}_{IV} \) and \( \hat{\beta}_{CT} \) have a common denominator \( \frac{1}{N} \sum_{i=1}^{N} \hat{x}_i \tilde{Z}_i \), a relevant test statistic is

\[
\hat{T} = \frac{1}{\hat{S}} \sum_{i=1}^{N} d_i \tilde{Z}_i,
\]

where \( d_i = \tilde{Y}_i - \tilde{Y}_i \) is a difference in (residualized) dependent variables between two regressions that yield \( \hat{\beta}_{IV} \) and \( \hat{\beta}_{CT} \). \( \hat{S} \) is the standard error of the numerator.\(^{26}\) For example, in an i.i.d. heteroscedastic case,

\[
N \hat{S}^2 = \frac{1}{N} \sum_{i=1}^{N} \left( d_i \tilde{Z}_i - \frac{1}{N} \sum_{j=1}^{N} d_j \tilde{Z}_j - v_{1i} \tilde{Z}_i \right)^2.
\]

Under a more general distributional assumption, the right hand side of equation (29) should be replaced by the estimated asymptotic variance of \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (d_i \tilde{Z}_i - \frac{1}{N} \sum_{j=1}^{N} d_j \tilde{Z}_j - v_{1i} \tilde{Z}_i) \). Appendix C.5 shows that \( \hat{T} \) converges in distribution to \( N(0, 1) \) under the null.

\(^{26}\)Another correction term appears in the case with DID- or RD-based instruments if \( \hat{b}_{OLS}(W_i) \) and \( \hat{a}_k(W_i) \) are not specified to be linear, as presented in appendix C.4.
hypothesis $\Delta_{ME} = 0$ and diverges under the alternative hypothesis $\Delta_{ME} \neq 0$.

Three limitations of this test are worth noting. First, it is a valid test of endogeneity only when the setup in section II describes the true model. Most importantly, the marginal effect difference $\Delta_{ME}$ cannot be attributed to endogeneity bias alone in the presence of unobserved heterogeneity, as discussed in section IIIIB. Second, this test cannot detect endogeneity bias if the difference between the IV-identified and OLS-identified marginal effects $\frac{\partial g(x, w)}{\partial x} - \frac{\partial m(x, w)}{\partial x}$ in some regions of $(x, w)$ exactly cancels out the difference in other regions. While the direction of endogeneity bias is expected to be unambiguous in many economic contexts, one might consider the nonparametric test proposed by Blundell and Horowitz (2007) in the contexts in which the direction of endogeneity may differ across $(x, w)$. Finally, this test shares the fundamental limitations with the standard DWH test that the instrument must be valid and that it has little power to detect endogeneity bias if the instrument is not sufficiently strong.

V. Applications to Return-to-Schooling Estimates

This section describes the empirical applications of my framework to return-to-schooling estimates with three different identification strategies. First, I use geographic variation in college costs as an instrument as in Cameron and Taber (2004) and Carneiro et al. (2011) and estimate the returns to schooling in the National Longitudinal Survey of Youth 1979 (NLSY79). Next, I exploit a discontinuity in the minimum school-leaving age in the United Kingdom as in Oreopoulos (2006), using the British General Household Survey (GHS). Finally, I exploit DID variation in compulsory schooling laws across cohorts and states in the United States as in Acemoglu & Angrist (2000) and estimate the returns to schooling in the 1960–1980 U.S. Censuses. Appendix D provides additional details for these analyses.

A. College Cost Instrument with Geographic Variation

This analysis uses the civilian sample from the NLSY79. The original sample consists of 5,579 males and 5,827 females born in 1957–1964. After dropping persons with missing information about their Armed Forces Qualification Test (AFQT) score or county of residence at age 14, persons who have not completed 8th grade by age 22, and persons with no wage information in any year, my sample consists of 4,719 males and 4,986 females. For each person, I use observations between ages 25 and 54. The outcome $Y$ is the log hourly wage at the current or most recent job and the treatment $X$ is years of schooling top-coded at 18 years.

Because minority and economically disadvantaged households are sampled at higher rates, persons are weighted by the sampling weights throughout my analysis. Within each person, I use equal weights for the person-year observations. It is important to weight observations appropriately to recover the population OLS and IV coefficients because my framework does not assume the linear structural equation and admits a weighted-average interpretation of the coefficients.

As instruments, I use measures of direct and opportunity costs of college attendance as in Cameron and Taber (2004) and Carneiro et al. (2011). In particular, I use the presence of a public four-year college and tuition rate of the nearest in-state public four-year college in the county of residence at age 14 to represent the direct cost of attendance. While the tuition rate captures a pecuniary cost of attendance, college proximity captures both pecuniary (due to reduced costs of room and board) and nonpecuniary costs of attendance. I use local earnings and unemployment rate in the county of residence at age 14 in the year in which the person turns age 17 (1974 for the oldest cohort and 1981 for the youngest) to represent the opportunity cost of attendance.

For the vector of covariates $W$, I use the AFQT percentile, age, an indicator for female, black and Hispanic dummies, indicators for parental education, the number of siblings, and cohort dummies. In addition, I include urban status, Census division dummies, and the average local earnings
and unemployment rate during 1974–1981 of the county of residence at age 14 in the covariate vector $W$. Using the average local labor market conditions as control variables ensures that variation in the corresponding instruments is driven by a temporary shock to the local labor market. Otherwise, the instruments may capture a permanent difference in local labor market conditions, which can be directly associated with potential earnings. Controlling for local labor market conditions is also important to ensure the plausibility of the college proximity instrument, as emphasized by Cameron and Taber (2004). Appendix D.1 provides further details of the sample and presents the first-stage regression results.

Table 1 reports the decomposition results. The first three columns of the table report the linear OLS coefficient, the linear IV coefficient, and the IV-OLS coefficient gap. The next three columns report the estimates of the covariate weight difference, the treatment-level weight difference, and the marginal effect difference. Here, I discuss the first row of the table, which reports the results from the NLSY79. The second and third rows are discussed in sections IIB and IIC. In the first row, the point estimates of the OLS and IV coefficients are nearly identical, with the OLS coefficient of 0.065 and the IV coefficient of 0.062. An empirical researcher may be tempted to conclude from this result that there is no evidence of ability bias in these data. However, the decomposition using the IV-weighted OLS coefficients indicates that the IV coefficient would be well below the OLS coefficient if they had the same weights on the covariates and treatment levels. In fact, the covariate weight difference is estimated to be 0.011 and the treatment-level weight difference is estimated to be 0.018. With the weight difference components accounted for, the marginal effect difference is $-0.032$, indicating that the IV-identified returns to schooling are lower than the OLS-identified returns. This result is rather consistent with the ability bias story in terms of point estimates, even though the generalized DWH test fails to reject at the 5% level given the large standard error.

I investigate the mechanisms underlying these results by examining the patterns of the IV and OLS weights. Table 2 presents the total IV and OLS weights on each group of covariates. The first three columns of the table present the population share, the total OLS weight, and the total IV weight on each group. Each set of weights sums to one across the whole sample by construction. The last column reports the OLS schooling coefficient from a linear regression performed separately for each group, to illustrate the difference in OLS-identified schooling effects across groups. While the OLS weights are close to the population shares, the IV weights are concentrated on persons with advantaged backgrounds in terms of AFQT score, parental education, and race/ethnicity. Schooling coefficients from the separately performed OLS regressions indicate that more-advantaged groups tend to have higher schooling effects. These data patterns result in the positive contribution of the covariate weight difference to the IV-OLS coefficient gap.

The pattern of IV weights is consistent with the empirical observation by Cameron and Taber (2004) that persons with advantaged backgrounds tend to be more sensitive to local college availability in the NLSY79. However, one may expect that persons with disadvantaged backgrounds should be weighted more because they are expected to be more sensitive to college cost instruments given their financial constraints. Several factors can explain low IV weights on persons with disadvantaged backgrounds. While Cameron and Taber (2004) suggest that increased funding on federal student aid programs in the 1970s is one such factor, low college attendance and graduation rates among persons from disadvantaged backgrounds can also be important. A low college attendance rate implies that the majority would be never-takers instead of compliers. A low college graduation rate implies that their completed years of schooling would not be strongly affected, even if the instruments affect their college attendance decisions.

Panel a of figure 1 illustrates the total IV and OLS weights on each treatment margin. The IV weights are mostly placed on college education margins, with the weights on high school margins close to zero. This weight pattern is consistent with the expectation that college cost instruments affect years of schooling through college attendance decisions. Higher IV weights on college education margins result in the positive contribution of the weight difference to the IV-OLS coefficient gap because marginal effects of years in college are much higher than marginal effects of years in high school in this sample. In fact, the linear OLS coefficient is 0.031 in the subsample with 12 or fewer years of schooling and 0.071 in the subsample with 12 or more years of schooling.

The overall result from this decomposition exercise that the IV coefficient is inflated due to the weight difference appears to match the discount rate bias argument developed by Lang (1993) and Card (1995a). However, the estimated weight patterns indicate that the underlying mechanism is distinct in two ways. First, the IV coefficient places more

36Card (1995b) and Kling (2001) find that persons from less advantaged backgrounds are more sensitive to the presence of a local college in the sample from the National Longitudinal Survey of Young Men, which is based on older cohorts than the NLSY79.

37The share of persons with one or more and four or more years of college education are 19% and 4%, respectively, among the bottom third of AFQT scores. Among persons in the top third of AFQT scores, 80% have one or more years of college education and 56% have four or more years of college education.

38If the college enrollment decision is explained by a logit or probit model, the share of compliers is the largest among the group with a college attendance rate of 50%.

39Building on the canonical model of Becker (1967), they consider a model in which individuals invest in education as long as the marginal return to an additional year of schooling exceeds its marginal cost. The model predicts that the marginal return at the chosen schooling level would be higher for credit-constrained individuals with higher discount rates. Given this prediction, they argue that the IV coefficient could exceed the population average return if credit-constrained individuals are more sensitive to the instruments and thus weighted more.
weight on advantaged rather than disadvantaged groups in the population. Higher IV weights on advantaged groups give rise to the higher IV coefficient because of the higher marginal returns among advantaged groups. This is the opposite mechanism to the discount rate bias argument, even though it shifts the IV coefficient in the same direction. Second, higher marginal returns to college education give rise to the higher IV coefficient in this sample, given the concentration of IV weights on college years relative to high school years. The discount rate bias argument does not consider this possibility, as it assumes nonincreasing marginal returns to schooling.

The empirical strategy in this analysis closely follows Oreopoulos (2006). I use the sample of persons younger than 65 years old from the British General Household Surveys in 1984–1998 who turned 14 in 1935–1965. I exclude persons with missing data on earnings or education and persons leaving school before age 10. The outcome $Y$ is log annual earnings. The treatment $X$ is years of schooling, which is given by the age when they left full-time education minus five years, with the top-coding at 20 years. The instrument $Z$ is an indicator for 1933 or later birth cohorts, who turned 14 one year after the reform.

### B. RD-Based Compulsory Schooling Instrument

As in Oreopoulos (2006), I now exploit the 1947 compulsory schooling reform in the United Kingdom to construct an RD instrument. The U.K. government raised the minimum school-leaving age in Great Britain from 14 to 15 years in 1947. The share of people leaving school at age 14 or earlier then fell from 56% in the 1932 birth cohort turning 14 one year before the reform to 9% in the 1934 birth cohort turning 14 one year after the reform.

### Table 1.—Decomposition of the IV-OLS Gap in Return-to-Schooling Estimates

| IV Strategy | Data | Coefficients | Decomposition |
|-------------|------|--------------|---------------|
|             |      | OLS | IV | IV-OLS | $\Delta_CW$ | $\Delta_{TW}$ | $\Delta_{ME}$ |
| College Cost Variation | NLSY79$^{1}$ | 0.065 | 0.062 | -0.004 | 0.011 | 0.018 | -0.032 |
| (0.003) | | (0.087) | (0.087) | (0.011) | (0.010) | (0.086) |
| Compulsory Schooling RD | British GHS$^{2}$ | 0.084 | 0.062 | -0.021 | -0.016 | -0.029 | 0.023 |
| (0.002) | | (0.083) | (0.082) | (0.009) | (0.018) | (0.078) |
| Compulsory Schooling DID | U.S. Census$^{3}$ | 0.067 | 0.084 | 0.017 | 0.011 | 0.003 | 0.003 |
| (0.0004) | | (0.022) | (0.022) | (0.004) | (0.003) | (0.021) |

Standard errors are in parentheses. The first three columns report the OLS estimates, the IV estimates, and their gaps. The next three columns report the estimates of the covariate weight difference, the treatment-level weight difference, and the marginal effect difference components. By construction, these three components sum to the IV-OLS gap. Appendix D describes the empirical specification for estimating the decomposition.

1) Standard errors are robust to heteroskedasticity and correlation across observations on persons living in the same county at age 14. The instruments are college cost measures as defined in the main text.

2) Standard errors are robust to heteroskedasticity and correlation across observations on persons living in the same county at age 14.

3) Standard errors are robust to heteroskedasticity and correlation across observations on the same state and year of birth. The instruments are indicators for compulsory schooling requirements implied by child labor laws (7, 8, and 9 or more years).

### Table 2.—The IV and OLS Weights on the Covariate Groups (NLSY79)

| Variable | Group | Group share | OLS weight | IV weight | Subsample OLS coef. |
|----------|-------|-------------|------------|-----------|---------------------|
| AFQT percentile | 0–1/3 | 0.32 | 0.25 (0.01) | -0.08 (0.16) | 0.056 (0.006) |
|           | 1/3–2/3 | 0.34 | 0.35 (0.01) | 0.19 (0.13) | 0.062 (0.005) |
|          | 2/3–1 | 0.34 | 0.40 (0.01) | 0.89 (0.20) | 0.074 (0.005) |
| Highest parental education | Some HS of less | 0.24 | 0.21 (0.01) | -0.12 (0.15) | 0.055 (0.005) |
|          | HS graduate | 0.42 | 0.41 (0.01) | 0.66 (0.17) | 0.069 (0.005) |
|         | Some college | 0.34 | 0.37 (0.01) | 0.46 (0.15) | 0.068 (0.005) |
|          | Black | 0.14 | 0.14 (0.02) | -0.12 (0.11) | 0.069 (0.005) |
| Race/Ethnicity | Hispanic | 0.06 | 0.06 (0.01) | 0.02 (0.03) | 0.064 (0.006) |
|         | Other | 0.80 | 0.80 (0.02) | 1.10 (0.11) | 0.064 (0.004) |
|          | Male | 0.50 | 0.49 (0.01) | 0.47 (0.17) | 0.057 (0.004) |
|          | Female | 0.50 | 0.51 (0.01) | 0.53 (0.17) | 0.074 (0.005) |
| Urban residence at age 14 | Rural | 0.31 | 0.30 (0.03) | 0.53 (0.17) | 0.073 (0.006) |
|          | Urban | 0.69 | 0.70 (0.03) | 0.47 (0.17) | 0.062 (0.004) |
|         | Northeast | 0.22 | 0.22 (0.04) | 0.18 (0.14) | 0.061 (0.007) |
| Region at age 14 | Midwest | 0.31 | 0.30 (0.04) | 0.51 (0.17) | 0.070 (0.007) |
|          | South | 0.32 | 0.32 (0.04) | 0.09 (0.13) | 0.058 (0.005) |
|          | West | 0.15 | 0.16 (0.03) | 0.23 (0.15) | 0.071 (0.005) |

Standard errors are in parentheses and robust to heteroskedasticity and correlation across observations on persons living in the same county at age 14. Each set of weights sums to one across the whole sample. Subsample OLS coefficient of years of schooling is obtained with the same set of control variables as regressions in table 1. Appendix D.4 describes the empirical specification for estimating the weights.

40) My regression results slightly differ from the originally published results in Oreopoulos (2006) due to the data correction (Oreopoulos, 2008) and a top-coding treatment of schooling described below.

41) Approximately 2.5% of persons report having left full-time education after age 25, and their schooling levels are all treated as 20 years. Oreopoulos (2006) does not make this top-coding treatment. See appendix D.2 for the analysis without top-coding, which reaches the same conclusion regarding the relevance of the weight difference components.

42) Gelman and Imbens (2019) recommend against the use of high-order polynomials in RD designs; nevertheless, I follow the original specification in Oreopoulos (2006). Using two separate quadratic polynomials for pre- and post-reform cohorts instead of global fourth-order polynomials slightly...
The second row of table 1 reports the decomposition of the IV-OLS coefficient gap in this empirical application. As the OLS estimate lies above the IV estimate by 0.021, a researcher who presumes the linear causal model may immediately interpret it as the result of ability bias. However, adjusting for the weight difference suggests otherwise. In fact, the decomposition result indicates that the marginal returns identified by the IV coefficient exceed those identified by the OLS coefficient by 0.023, after accounting for the covariate weight difference and the treatment-level weight difference components. The empirical result is no longer consistent with the ability bias story as a point estimate, although both standard and generalized DWH tests fail to reject at the 5% level due to the imprecise IV estimate.
Table 3.—The IV and OLS Weights on the Covariate Groups (British GHS)

| Year at 14 | Group share | OLS weights | IV weights | Subsample OLS coef. |
|------------|-------------|-------------|------------|---------------------|
| 1935–40    | 0.04        | 0.03        | −0.09      | 0.054               |
| 1941–45    | 0.09        | 0.08        | 0.68       | 0.065               |
| 1946–50    | 0.14        | 0.12        | 0.49       | 0.079               |
| 1951–55    | 0.19        | 0.18        | −0.10      | 0.086               |
| 1956–60    | 0.25        | 0.26        | 0.06       | 0.088               |
| 1961–65    | 0.28        | 0.34        | −0.02      | 0.088               |

Standard errors (in parentheses) are robust to heteroskedasticity and correlation across observations on the same birth cohort and survey year. Each set of weights sums to one across the whole sample. Subsample OLS coefficient of years of schooling is obtained with controlling for quartic terms of ages and birth cohorts. Appendix D.4 describes the empirical specification for estimating the weights.

Table 3 presents the IV and OLS weights on the birth cohorts turning age 14 in 1941–1950. The linear regression restricted to these cohorts yields smaller OLS schooling coefficients than those for the younger cohorts, who receive most of the OLS weights. This explains the negative contribution of the covariate weight difference to the IV-OLS coefficient gap. Panel b of figure 1 shows that the IV weights are almost exclusively placed on the 10th year of schooling, which is exactly what the 1947 reform mandated. This weight pattern pushes down the IV coefficient because the 10th year of schooling has a smaller marginal return than the other schooling margins. In fact, the linear regression restricted to individuals with 9 or 10 years of schooling yields an OLS schooling coefficient of 0.037, which is much smaller than the full-sample OLS coefficient.

This analysis demonstrates that researchers should exercise caution in extending the intuition of Imbens and Angrist (1994) to an empirical setting with covariates and a multi-valued treatment. Observing a sharp decline in the dropout rate at age 14 after the reform, Oreopoulos (2006) argues that the IV and OLS coefficients in this setting are expected to identify the treatment effects for the comparable population, providing an analogy to the comparison between the LATE and ATE. In principle, however, the LATE interpretation in Imbens and Angrist (1994) draws on a model with a binary treatment and no covariates. It is still plausible that the LATE and ATE in this setting match conditional on the schooling level and the birth cohort, judging by the extensive response to the reform. Nevertheless, the estimated patterns of the weights suggest that the IV and OLS coefficients identify the effects for entirely different birth cohorts at completely distinct schooling margins.

C. Compulsory Schooling Instrument with DID Variation

This analysis exploits DID variation in compulsory schooling laws across cohorts and regions, following Acemoglu and Angrist (2000). The analysis sample consists of 40–49-year-old white males born in the United States from the 1960–1980 Censuses. I use log weekly earnings as the outcome $Y$, and limit my analysis to persons with positive earnings and working for at least one week in the previous year. I use years of schooling as the treatment $X$. I include in a vector of covariates $W$ indicators for the state of birth and indicators for the year of birth. As instruments, I use the status of compulsory schooling laws (CSL) at age 14 in the state of birth. As in Acemoglu and Angrist (2000), I construct the CSL instruments based on the required years of schooling associated with child labor laws. Appendix D.3 provides further details of the sample, first-stage regression results, and decomposition results using other common compulsory schooling instruments.

The third row of table 1 presents the estimated OLS and IV schooling coefficients and the decomposition of the IV-OLS coefficient gap. While the IV estimate is slightly above the OLS estimate with a coefficient gap of 0.017, the decomposition result indicates that the gap is primarily associated with the weight difference components. Although the IV-OLS gap is not statistically distinguishable from zero even without accounting for the weight difference components, this result reshapes the quantitative implication from the gap. In particular, Acemoglu and Angrist (2000) attribute the small positive IV-OLS gap to modest external returns to schooling. Accounting for the weight difference components, this result indicates even less important externality than their original interpretation.

Table 4 presents the total IV and OLS weights on each group of covariates. The IV weights attached to some birth cohorts or birth states are negative. In fact, the IV weights aggregate to negative values among persons in the 1930–1939 birth cohorts and among persons born in the Midwest or West. Moreover, I find that 54% of covariate-specific IV weights are negative and sum to $-5.93$.

In a usual IV setting, negative weights imply the presence of both compliers and defiers. In a DID setting, however, the weights are not guaranteed to be positive even with the perfect compliance $X = Z$, as pointed out by de Chaisemartin and D’Haultfoeuille (2020) in the binary treatment case. While the IV-weighted OLS coefficient $\beta_C = \int b_{OLS}(w)\sigma_Z(w)dw(w)$ is a weighted average of the covariate-specific OLS coefficients $b_{OLS}(w)$, negative weights can push $\beta_C$ out of the support of $b_{OLS}(w)$. In fact, the estimates of $b_{OLS}(w)$ are no greater than 0.076, 0.037, which is much smaller than the full-sample OLS coefficient.

My specification follows one of their main specifications in estimating private returns to schooling that uses the 1960–1980 Census data with the child labor laws instrument and no state-of-residence controls (Acemoglu and Angrist, 2000, p. 34). While they adjust some variables in the 1960–1980 data to incorporate the 1950 Census data in their alternative specifications, I do not make these adjustments as I focus only on the 1960–1980 Censuses. See appendix D.3 for details.

In fact, the first stage regression indicates the positive relationship between the CSL requirements and schooling levels, even in the subsample of the 1930–1939 birth cohorts or among persons born in the Midwest or West.
Despite $\beta_C$ being estimated to be 0.079. This result suggests that even a small heterogeneity in treatment effects can give rise to a large contribution of the weight difference to the IV-OLS coefficient gap if the IV strategy relies on DID variation.

Panel c of figure 1 reports the total IV and OLS weights for each schooling level. The IV weights mostly capture the schooling margins up to the 12th year of schooling, which is consistent with the context that primary and secondary education is mandated by the CSL. However, the effect of this treatment-level weight difference on the IV-OLS coefficient gap is mostly obscured by the large contribution of the covariate weight difference.

### VI. Conclusion

When OLS and IV estimates differ, empirical researchers typically consider two explanations. The first takes the linear regression equation literally and interprets the coefficient gap as endogeneity bias. The second extends the intuition of the LATE interpretation (Imbens & Angrist, 1994) to a general regression equation and interprets the coefficient gap as the weight difference. My paper enables researchers to proceed a step further and formally quantify the contributions of the weight difference and endogeneity bias components separately.

I show that the IV-OLS coefficient gap is explained by differences in the weights on the covariates, the weights on the treatment levels, and the identified marginal effects. The marginal effect difference component captures endogeneity bias in the absence of the unobservable-driven interaction between the treatment effects and treatment responses to the instrument. I propose a simple two-step regression approach to perform the decomposition empirically, which can be implemented in standard statistical packages.

I demonstrate the practical value of my framework through its empirical applications to return-to-schooling estimates with compulsory schooling and college cost instruments. The IV-OLS coefficient gaps in these empirical applications are substantially influenced by the weight difference components, and accounting for them leads to different conclusions about the direction or extent of endogeneity bias.

### REFERENCES

Acemoglu, D., and J. Angrist, “How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws,” *NBER Macroeconomics Annual* 15 (2000), 9–59. 10.1086/564003

Ackerberg, D., X. Chen, and J. Hahn, “A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators,” this REVIEW 94:2 (2012), 481–498.

Angrist, J. D., “Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants,” *Econometrica*, 66 (1998), 249–288. 10.2307/2988558

Angrist, J. D., K. Graddy, and G. W. Imbens, “The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish,” *The Review of Economic Studies* 67:3 (2000), 499–527. 10.1111/1467-937X.00141

Angrist, J. D., and G. W. Imbens, “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” *Journal of the American Statistical Association* 90:430 (1995), 431–442. 10.1080/01621459.1995.10476535

Angrist, J. D., and A. B. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” *The Quarterly Journal of Economics* 106:4 (1991), 979–1014. 10.2307/2937954

——— “Empirical Strategies in Labor Economics” (pp. 1277–1366), in *Handbook of Labor Economics*, Vol. 3 (Amsterdam: Elsevier, 1999a). 10.1016/S1573-4463(99)03011-4

Angrist, J. D., and J.-S. Pischke, “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics,” *Journal of Economic Perspectives* 24:2 (2010), 3–30. 10.1257/jep.24.2.3

Arnon, P. M., and C. Samii, “Does Regression Produce Representative Estimates of Causal Effects?” *American Journal of Political Science* 60:1 (2016), 250–267. 10.1111/ajps.12185

Becker, G. S., *Human Capital and the Personal Distribution of Income: An Analytical Approach* (Ann Arbor, MI: Institute of Public Administration, 1967).

Black, S. E., P. J. Devereux, and K. G. Salvanes, “Why the Apple Doesn’t Fall Far: Understanding Intergenerational Transmission of Human Capital,” *American Economic Review* 95:1 (2005), 437–449. 10.1257/0002828053828635

Blundell, R., X. Chen, and D. Kristensen, “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves,” *Econometrica* 75:6 (2007), 1613–1669. 10.1111/j.1468-0262.2007.00808.x

Blundell, R., and J. L. Horowitz, “A Non-Parametric Test of Exogeneity,” *The Review of Economic Studies* 74:4 (2007), 1035–1058. 10.1111/j.1467-937X.2007.00458.x

Cameron, S. V., and C. Taber, “Estimation of Educational Borrowing Constraints Using Returns to Schooling,” *Journal of Policy Economics* 112:1 (2004), 132–182. 10.1086/379937

Card, D., *Earnings, Schooling, and Ability Revisited*, Vol. 14 (Greenwich Connecticut: JAI Press, 1995a).

——— Using Geographic Variation in College Proximity to Estimate the Return to Schooling (Toronto: University of Toronto Press, 1995b).

——— “The Causal Effect of Education on Earnings” (pp. 1801–1863), in *Handbook of Labor Economics*, Vol. 3 (Amsterdam: Elsevier, 1999). 10.1016/S1573-4463(99)03011-4

——— “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,” *Econometrica* 69:5 (2001), 1127–1160. 10.1111/1468-0262.00237

Carneiro, P., J. J. Heckman, and E. Vytlacil, “Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin,” *Econometrica* 78:1 (2010), 377–394. 10.3982/ECTA7089

Carneiro, P., J. J. Heckman, and E. J. Vytlacil, “Estimating Marginal Returns to Education,” *American Economic Review* 101:6 (2011), 2754–2781. 10.1257/aer.101.6.2754
Chernozhukov, V., G. W. Imbens, and W. K. Newey, “Instrumental Variable Estimation of Nonseparable Models,” *Journal of Econometrics* 139:1 (2007), 4–14. 10.1016/j.jeconom.2006.06.002

de Chaisemartin, C., and X. D’Haultfoeuille, “Fuzzy Differences-in-Differences,” *The Review of Economic Studies* 85:2 (2018), 999–1028. 10.1093/restud/rdx049

——— “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review* 110:9 (2020), 2964–2996. 10.1257/aer.20181169

Donald, S. G., Y.-C. Hsu, and R. P. Lieli, “Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT,” *Journal of Business & Economic Statistics* 32:3 (2014), 395–415.

Duflo, E., “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review* 91:4 (2001), 795–813. 10.1257/ aer.91.4.795

Durbin, J., “Errors in Variables,” *Review of the International Statistical Institute* 22 (1954), 23–32. 10.2307/1401917

Gelman, A., and G. Imbens, “Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs,” *Journal of Business & Economic Statistics* 37:3 (2019), 447–456.

Hausman, J. A., “Specification Tests in Econometrics,” *Econometrica* 46 (1978), 1251–1271. 10.2307/1913827

Heckman, J. J., and S. Urzua, “Comparing IV with Structural Models: What Simple IV Can and Cannot Identify,” *Journal of Econometrics* 156:1 (2010), 27–37. 10.1016/j.jeconom.2009.09.006

Heckman, J. J., S. Urzua, and E. Vyltacil, “Understanding Instrumental Variables in Models with Essential Heterogeneity,” this *Review* 88:3 (2006), 389–432.

Heckman, J. J., and E. Vyltacil, “Policy-Relevant Treatment Effects,” *American Economic Review* 91:2 (2001), 107–111. 10.1257/ aer.91.2.107

Horowitz, J. L., “Applied Nonparametric Instrumental Variables Estimation,” *Econometrica* 79:2 (2011), 347–394. 10.3982/ECTA8662

Imbens, G. W., and J. D. Angrist, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica* 62:2 (1994), 467–475. 10.2307/2951620

Imbens, G. W., and W. K. Newey, “Identification and Estimation of Triangular Simultaneous Equations Models without Additivity,” *Econometrica* 77:5 (2009), 1481–1512. 10.3982/ECTA7108

Kane, T. J., and C. E. Rouse, “Labor-Market Returns to Two-and Four-Year College,” *The American Economic Review* 85:3 (1995), 600–614.

Kling, J. R., “Interpreting Instrumental Variables Estimates of the Returns to Schooling,” *Journal of Business & Economic Statistics* 19:3 (2001), 358–364.

Lang, K., “Ability Bias, Discount Rate Bias and the Return to Education,” MPRA paper 24651 (1993).

Lochner, L., and E. Moretti, “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports,” NBER working paper 8605 (2001).

——— “Estimating and Testing Models with Many Treatment Levels and Limited Instruments,” this *Review* 97:2 (2015), 387–397.

Loken, K. V., M. Mogstad, and M. Wiswall, “What Linear Estimators Miss: The Effects of Family Income on Child Outcomes,” *American Economic Journal: Applied Economics* 4:2 (2012), 1–35.

Masten, M. A., and A. Torgovitsky, “Identification of Instrumental Variable Correlated Random Coefficients Models,” this *Review* 98:5 (2016), 1001–1005.

Mogstad, M., A. Santos, and A. Torgovitsky, “Using instrumental variables for inference about policy relevant treatment parameters,” *Econometrica* 86:5 (2018), 1589–1619. 10.3982/ECTA15463

Mogstad, M., A. Torgovitsky, and C. R. Walters, “The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables,” *American Economic Review* 111:11 (November 2021), 3663–3698. 10.1257/aer.20190251

Mogstad, M., and M. Wiswall, “Linearity in Instrumental Variables Estimation: Problems and Solutions,” IZA discussion paper 5216 (2010).

Newey, W. K., and D. McFadden, “Large Sample Estimation and Hypothesis Testing” (pp. 2111–2245), in *Handbook of Econometrics*, Vol. 4 (Amsterdam: Elsevier, 1994). 10.1016/S1573-4412(05)80005-4

Newey, W. K., and J. L. Powell, “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica* 71:5 (2003), 1565–1578. 10.1111/1468-0262.00459

Oreopoulos, P., “Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter,” *American Economic Review* 96:1 (2006), 152–175. 10.1257/00028806776157641

——— “Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter: Corrigendum,” *American Economic Review* (2008), https://www.aeaweb.org/ae/contents/corrregea/ae_jrnl_96.1.152.pdf

Schechtman, E., and S. Yitzhaki, “The Gini Instrumental Variable, or the ‘Double Instrumental Variable’ Estimator,” *Metron* 62 (2004), 287–313.

Słoczyński, T., “When Should We (Not) Interpret Linear IV Estimands as LATE?” CESifo working paper 9064 (2021).

——— Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights,” this *Review* 104:3 (2022), 501–509.

Staiger, D., and J. H. Stock, “Instrumental Variables Regression with Weak Instruments,” *Econometrica* 65:3 (1997), 557–586. 10.2307/2171753

Wu, D.-M., “Alternative Tests of Independence between Stochastic Regressors and Disturbances,” *Econometrica* 41:4 (1973), 733–750. 10.2307/1914093

Yitzhaki, S., “On Using Linear Regressions in Welfare Economics,” *Journal of Business & Economic Statistics* 14:4 (1996), 478–486.