Negative Weights are No Concern in Design-Based Specifications

Kirill Borusyak
Berkeley

Peter Hull
Brown

ASSA 2024

Motivation

A recent literature raises concerns with common OLS & IV specifications:

- They may fail to estimate convex-weighted averages of causal effects, even when they succeed at avoiding omitted variables bias (OVB)
- The "negative weights" can yield *sign reversals*: e.g. negative OLS/IV estimates when all causal effects are positive

Motivation

A recent literature raises concerns with common OLS & IV specifications:

- They may fail to estimate convex-weighted averages of causal effects, even when they succeed at avoiding omitted variables bias (OVB)
- The "negative weights" can yield *sign reversals*: e.g. negative OLS/IV estimates when all causal effects are positive

Much of this literature focuses on specifications that address OVB by modeling potential outcomes given the treatment (e.g. "parallel trends")

- The (possibly negative) weights in the estimand representation are *ex-post*: i.e., functions of the realized treatment and controls
- More flexible specifications can sometimes avoid negative ex-post weights (e.g. Wooldridge 2021, Borusyak et al. 2023)

This Paper

We show that negative ex-post weights also arise—but are no concern—in *design-based* OLS & IV specifications

- I.e., those that leverage assumptions on treatment or instrument assignment, rather than a model for potential outcomes

This Paper

We show that negative ex-post weights also arise—but are no concern—in *design-based* OLS & IV specifications

- I.e., those that leverage assumptions on treatment or instrument assignment, rather than a model for potential outcomes

Design-based estimands have an average-effect representation with *ex-ante* weights: expectations of ex-post weights over the assignment distribution

- These weights are guaranteed to be convex in design-based OLS specifications, so sign reversals cannot occur
- In design-based IV specifications, convexity follows under a general first-stage monotonicity condition

Literature Connections

This analysis connects the recent negative-weight literature with a classic one on convex weighting in OLS & IV (e.g., Imbens and Angrist 1994, 1995; Angrist 1998; Angrist and Krueger 1999; Angrist, Graddy and Imbens 2000...)

- Relative to this literature, we use a weaker mean independence condition that highlights the role of expected treatments/instruments (Borusyak and Hull 2023) for design-based OLS/IV identification
- We also use a weaker montonicity condition (c.f. Small et al. 2017) that allows the IV first stage to be non-causal

Literature Connections

This analysis connects the recent negative-weight literature with a classic one on convex weighting in OLS & IV (e.g., Imbens and Angrist 1994, 1995; Angrist 1998; Angrist and Krueger 1999; Angrist, Graddy and Imbens 2000...)

- Relative to this literature, we use a weaker mean independence condition that highlights the role of expected treatments/instruments (Borusyak and Hull 2023) for design-based OLS/IV identification
- We also use a weaker montonicity condition (c.f. Small et al. 2017) that allows the IV first stage to be non-causal

Both extensions can be useful for "formula" treatment/instruments, which combine exogenous shocks with non-random measures of exposure

- E.g. shift-share instruments (Borusyak et al. 2022), treatments capturing economic/network spillovers (Borusyak and Hull 2023), and simulated instruments for policy eligibility (Borusyak and Hull 2021)

Simple Setup

A researcher estimates by OLS:

$$y_i = \beta x_i + w_i'\gamma + e_i,$$

for some outcome y_i, treatment x_i, and vector of controls w_i

Simple Setup

A researcher estimates by OLS:

$$y_i = \beta x_i + w_i' \gamma + e_i,$$

for some outcome y_i, treatment x_i, and vector of controls w_i

To interpret this specification, we consider a linear-effect causal model:

$$y_i = x_i \beta_i + \varepsilon_i,$$

with heterogeneous effects β_i and untreated potential outcomes ε_i

Simple Setup

A researcher estimates by OLS:

$$y_i = \beta x_i + w_i' \gamma + e_i,$$

for some outcome y_i, treatment x_i, and vector of controls w_i

To interpret this specification, we consider a linear-effect causal model:

$$y_i = x_i \beta_i + \varepsilon_i,$$

with heterogeneous effects β_i and untreated potential outcomes ε_i

Assume appropriate asymptotics for OLS to consistently estimate:

$$\beta = \frac{E[\tilde{x}_i y_i]}{E[\tilde{x}_i^2]} = \frac{E[\tilde{x}_i x_i \beta_i] + E[\tilde{x}_i \varepsilon_i]}{E[\tilde{x}_i^2]},$$

where $\tilde{x}_i$ are residuals from the population projection of x_i on w_i

Two Paths to Avoiding OVB

$E[\tilde{x}_i \varepsilon_i] = 0$ under either one of two assumptions:

Two Paths to Avoiding OVB

$E[\tilde{x}_i \varepsilon_i] = 0$ under either one of two assumptions:

ASSUMPTION 1: $E[\varepsilon_i \mid x_i, w_i] = w_i' \gamma$

- Untreated potential outcomes are linear in controls, given treatment
- E.g. parallel trends, where i indexes unit-period pairs in a panel and w_i includes unit and time dummies

Two Paths to Avoiding OVB

$E[\tilde{x}_i \varepsilon_i] = 0$ under either one of two assumptions:

ASSUMPTION 1: $E[\varepsilon_i \mid x_i, w_i] = w_i' \gamma$

- Untreated potential outcomes are linear in controls, given treatment
- E.g. parallel trends, where i indexes unit-period pairs in a panel and w_i includes unit and time dummies

ASSUMPTION 2: $E[x_i \mid \varepsilon_i, \beta_i, w_i] = w_i' \lambda$

- Treatment is conditionally mean-independent of potential outcomes, with a linear *expected treatment* $E[x_i \mid w_i]$ (e.g. the propensity score)
- E.g. a stratified experiment, where x_i is randomly assigned within strata dummied out in w_i

Two Paths to Avoiding OVB

$E[\tilde{x}_i \varepsilon_i] = 0$ under either one of two assumptions:

ASSUMPTION 1: $E[\varepsilon_i \mid x_i, w_i] = w_i' \gamma$

- Untreated potential outcomes are linear in controls, given treatment
- E.g. parallel trends, where i indexes unit-period pairs in a panel and w_i includes unit and time dummies

ASSUMPTION 2: $E[x_i \mid \varepsilon_i, \beta_i, w_i] = w_i' \lambda$

- Treatment is conditionally mean-independent of potential outcomes, with a linear *expected treatment* $E[x_i \mid w_i]$ (e.g. the propensity score)
- E.g. a stratified experiment, where x_i is randomly assigned within strata dummied out in w_i

The second assumption yields a "design-based" OLS specification

Ex-Post Weights

Since $E[\tilde{x}_i \varepsilon_i] = 0$, the estimand has an average-effect representation under either assumption:

$$\beta = \frac{E[\psi_i \beta_i]}{E[\psi_i]}, \qquad \psi_i = \tilde{x}_i x_i$$

Ex-Post Weights

Since $E[\tilde{x}_i \varepsilon_i] = 0$, the estimand has an average-effect representation under either assumption:

$$\beta = \frac{E[\psi_i \beta_i]}{E[\psi_i]}, \qquad \psi_i = \tilde{x}_i x_i$$

But the ex-post weights ψ_i are generally non-convex: $E[\tilde{x}_i] = 0$, so $\tilde{x}_i$ must take on both positive and negative values

Ex-Post Weights

Since $E[\tilde{x}_i \varepsilon_i] = 0$, the estimand has an average-effect representation under either assumption:

$$\beta = \frac{E[\psi_i \beta_i]}{E[\psi_i]}, \qquad \psi_i = \tilde{x}_i x_i$$

But the ex-post weights ψ_i are generally non-convex: $E[\tilde{x}_i] = 0$, so $\tilde{x}_i$ must take on both positive and negative values

- E.g. if $x_i > 0$ then i with low values of x_i (the effective control group) will always receive negative ex-post weight

Ex-Post Weights

Since $E[\tilde{x}_i \varepsilon_i] = 0$, the estimand has an average-effect representation under either assumption:

$$\beta = \frac{E[\psi_i \beta_i]}{E[\psi_i]}, \qquad \psi_i = \tilde{x}_i x_i$$

But the ex-post weights ψ_i are generally non-convex: $E[\tilde{x}_i] = 0$, so $\tilde{x}_i$ must take on both positive and negative values

- E.g. if $x_i > 0$ then i with low values of x_i (the effective control group) will always receive negative ex-post weight
- This can lead to sign reversals: e.g. $\beta < 0$, despite $\beta_i > 0$

Ex-Post Weights

Since $E[\tilde{x}_i \varepsilon_i] = 0$, the estimand has an average-effect representation under either assumption:

$$\beta = \frac{E[\psi_i \beta_i]}{E[\psi_i]}, \qquad \psi_i = \tilde{x}_i x_i$$

But the ex-post weights ψ_i are generally non-convex: $E[\tilde{x}_i] = 0$, so $\tilde{x}_i$ must take on both positive and negative values

- E.g. if $x_i > 0$ then i with low values of x_i (the effective control group) will always receive negative ex-post weight
- This can lead to sign reversals: e.g. $\beta < 0$, despite $\beta_i > 0$

The ex-post weights are the end of the story for β under Assumption 1. But in design-based specifications we can take one more step

- In experiments, who is in the effective control group is *random*...

Ex-Ante Weights

Under Assumption 2 only, the estimand has another representation:

$$\beta = \frac{E[\phi_i \beta_i]}{E[\phi_i]}, \quad \phi_i = E[\psi_i \mid w_i, \beta_i]$$

Ex-Ante Weights

Under Assumption 2 only, the estimand has another representation:

$$\beta = \frac{E[\phi_i \beta_i]}{E[\phi_i]}, \quad \phi_i = E[\psi_i \mid w_i, \beta_i]$$

The ex-ante weights are necessarily convex: $\phi_i = Var(x_i \mid w_i, \beta_i) > 0$

- Sign reversals thus cannot occur in design-based OLS specifications

The Role of the Expected Treatment

Comparing Assumption 2 to alternatives shows that the key to convex weights is the design-based specification of the expected treatment

The Role of the Expected Treatment

Comparing Assumption 2 to alternatives shows that the key to convex weights is the design-based specification of the expected treatment

Stronger models for unobservables need not help: e.g. sign reversal still may occur if we augment Assumption 1 with $E[\beta_i \mid x_i, w_i] = w_i'\delta$

- Though more flexible specifications can avoid negative weights under such models (e.g. Imbens and Wooldridge 2009)

The Role of the Expected Treatment

Comparing Assumption 2 to alternatives shows that the key to convex weights is the design-based specification of the expected treatment

Stronger models for unobservables need not help: e.g. sign reversal still may occur if we augment Assumption 1 with $E[\beta_i \mid x_i, w_i] = w_i'\delta$

- Though more flexible specifications can avoid negative weights under such models (e.g. Imbens and Wooldridge 2009)

Stronger unconfoundedness assumptions, e.g. $x_i \perp\!\!\!\perp (\varepsilon_i, \beta_i) \mid w_i$ turn out to be unnecessary for ensuring no sign reversals

- Though the ex-ante weights are identified under such assumptions: $\phi_i = Var(x_i \mid w_i)$ (e.g. Angrist and Krueger 1999)

General Result

Causal model with potential outcomes $y_i(x)$ and $y_i = y_i(x_i)$. Generalize:

ASSUMPTION 1′: $E[y_i(0) \mid z_i, w_i] = w_i'\gamma$

ASSUMPTION 2′: $E[z_i \mid y_i(\cdot), w_i] = w_i'\lambda$,

where z_i is an instrument (OLS special case: $z_i = x_i$).

General Result

Causal model with potential outcomes $y_i(x)$ and $y_i = y_i(x_i)$. Generalize:

ASSUMPTION 1′: $E[y_i(0) \mid z_i, w_i] = w_i'\gamma$

ASSUMPTION 2′: $E[z_i \mid y_i(\cdot), w_i] = w_i'\lambda$,

where z_i is an instrument (OLS special case: $z_i = x_i$). Further consider:

ASSUMPTION 3: *$Pr(x_i \geq x \mid z_i = z, y_i(\cdot), w_i)$ is non-decreasing in z for all x, almost surely over $(y_i(\cdot), w_i)$*,

and suppose the IV estimator consistently estimates $\beta = E[\tilde{z}_i y_i]/E[\tilde{z}_i x_i]$

General Result

Causal model with potential outcomes $y_i(x)$ and $y_i = y_i(x_i)$. Generalize:

ASSUMPTION 1′: $E[y_i(0) \mid z_i, w_i] = w_i'\gamma$

ASSUMPTION 2′: $E[z_i \mid y_i(\cdot), w_i] = w_i'\lambda$,

where z_i is an instrument (OLS special case: $z_i = x_i$). Further consider:

ASSUMPTION 3: *$Pr(x_i \geq x \mid z_i = z, y_i(\cdot), w_i)$ is non-decreasing in z for all x, almost surely over $(y_i(\cdot), w_i)$,*

and suppose the IV estimator consistently estimates $\beta = E[\tilde{z}_i y_i]/E[\tilde{z}_i x_i]$

PROPOSITION 1: *Let $\beta_i(x) = \frac{d}{dx} y_i(x)$. Under either A1′ or A2′:*

$$\beta = E[\textstyle\int \psi_i(x)\beta_i(x)dx]/E[\textstyle\int \psi_i(x)dx]$$

for non-convex ex-ante weights $\psi_i(x) = \tilde{z}_i \cdot \mathbf{1}[x_i \geq x]$. Under A2′ only:

$$\beta = E[\textstyle\int \phi_i(x)\beta_i(x)dx]/E[\textstyle\int \phi_i(x)dx]$$

for ex-ante weights $\phi_i(x) = E[\psi_i(x) \mid y_i(\cdot), w_i]$ that are convex under A3

Application: Formula Instruments

Proposition 1 applies to treatments/instruments of the form $z_i = f_i(s, g)$ where $g = (g_k)_{k=1}^K$ are exogenous shocks and $f_i(s, \cdot)$ governs exposure

- E.g. shift-share instruments: $z_i = \sum_k s_{ik} g_k$ (Borusyak et al. 2022)

Application: Formula Instruments

Proposition 1 applies to treatments/instruments of the form $z_i = f_i(s, g)$ where $g = (g_k)_{k=1}^K$ are exogenous shocks and $f_i(s, \cdot)$ governs exposure

- E.g. shift-share instruments: $z_i = \sum_k s_{ik} g_k$ (Borusyak et al. 2022)

Ignorability, $z_i \perp\!\!\!\perp y_i(\cdot) \mid w_i$, may be implausible while A2 holds

- E.g. when $E[g_k \mid y_i(\cdot), q_k, s] = q_k' \theta$ and $\sum_k s_{ik} q_k$ is controlled for

Application: Formula Instruments

Proposition 1 applies to treatments/instruments of the form $z_i = f_i(s, g)$ where $g = (g_k)_{k=1}^K$ are exogenous shocks and $f_i(s, \cdot)$ governs exposure

- E.g. shift-share instruments: $z_i = \sum_k s_{ik} g_k$ (Borusyak et al. 2022)

Ignorability, $z_i \perp\!\!\!\perp y_i(\cdot) \mid w_i$, may be implausible while A2 holds

- E.g. when $E[g_k \mid y_i(\cdot), q_k, s] = q_k'\theta$ and $\sum_k s_{ik} q_k$ is controlled for

First-stage monotonicity can hold, despite the first stage not being causal

- E.g. when the shares s_{ik} imperfectly proxy for true shock exposure

Conclusions

Design-based OLS & IV specifications generally avoid the negative ex-post weight concerns of the recent literature

- Given correct specification of the expected treatment/instrument

Conclusions

Design-based OLS & IV specifications generally avoid the negative ex-post weight concerns of the recent literature

- Given correct specification of the expected treatment/instrument

Of course, researchers may have broader goals than avoiding sign reversals

- More flexible specifications, design-based or otherwise, can let them pick other (maybe more policy-relevant) weighting schemes
- Sign reversals may also not arise if effect heterogeneity is limited or uncorrelated with the ex-post weights

Conclusions

Design-based OLS & IV specifications generally avoid the negative ex-post weight concerns of the recent literature

- Given correct specification of the expected treatment/instrument

Of course, researchers may have broader goals than avoiding sign reversals

- More flexible specifications, design-based or otherwise, can let them pick other (maybe more policy-relevant) weighting schemes
- Sign reversals may also not arise if effect heterogeneity is limited or uncorrelated with the ex-post weights

Two other important caveats:

- "Contamination bias" yields negative ex-ante weights in design-based specifications with multiple treatments (Goldsmith-Pinkham et al. 2022)
- High-dimensional controls / FEs can also yield bias (Freedman 2008)