

Assessing Omitted Variable Bias when the Controls are Endogenous*

Paul Diegert[†] Matthew A. Masten[‡] Alexandre Poirier[§]

June 20, 2025

Abstract

Omitted variables are one of the most important threats to the identification of causal effects. Several widely used methods assess the impact of omitted variables on empirical conclusions by comparing measures of selection on observables with measures of selection on unobservables. The recent literature has discussed various limitations of these existing methods, however. This includes a companion paper of ours which explains issues that arise when the omitted variables are endogenous, meaning that they are correlated with the included controls. In the present paper, we develop a new approach to sensitivity analysis that avoids those limitations, while still allowing researchers to calibrate sensitivity parameters by comparing the magnitude of selection on observables with the magnitude of selection on unobservables as in previous methods. We illustrate our results in an empirical study of the effect of historical American frontier life on modern cultural beliefs. Finally, we implement these methods in the companion Stata module `regssensitivity` for easy use in practice.

JEL classification: C18; C21; C51

Keywords: Treatment Effects, Sensitivity Analysis, Unconfoundedness

*This paper is a revised, shorter version of our now-superseded previous working paper titled “Assessing Omitted Variable Bias when the Controls are Endogenous” (Diegert, Masten, and Poirier 2023, arXiv:2206.02303v4), without the design-based framework of Section 3. The design-based framework and associated results can now be found in our companion paper “An Axiomatic Approach to Comparing Sensitivity Parameters” (Diegert, Masten, and Poirier 2025). We thank audiences at Northwestern, Duke, the SEA 2021 conference, Oxford, Brown, Texas A&M, the joint Bonn-Mannheim seminar, Jinau University, the 2022 Interactions Conference at The University of Wisconsin-Madison, University of Virginia, UC Irvine, UC San Diego, UCLA, Ohio State, Yale, Penn, the 2023 AEA winter meeting, Johns Hopkins, George Washington University, Stanford, UC Santa Cruz, Western Ontario, UIUC, Michigan State, Rochester, the Federal Reserve Bank of Cleveland, the 2023 American Causal Inference Conference, and the joint CREST-PSE seminar. We thank audiences at those seminars and conferences, as well as Joe Altonji, Isaiah Andrews, Peter Hull, Evan Rose, and Jon Roth for helpful conversations and comments. We thank Hongchang Guo and Muyang Ren for excellent research assistance. Masten thanks the National Science Foundation for research support under Grant 1943138.

[†]Toulouse School of Economics, paul.diegert@tse-fr.eu

[‡]Department of Economics, Duke University, matt.masten@duke.edu

[§]Department of Economics, Georgetown University, alexandre.poirier@georgetown.edu

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

- Prior papers assume $\text{Corr}(X, W) = 0$ (exogenous controls)

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

- Prior papers assume $\text{Corr}(X, W) = 0$ (exogenous controls)

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

- Prior papers assume $\text{Corr}(X, W) = 0$ (exogenous controls)
- i.e. the X 's (control variables) are themselves exogenous

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

- Prior papers assume $\text{Corr}(X, W) = 0$ (exogenous controls)
- i.e. the X 's (control variables) are themselves exogenous
- But this is rarely true in practice

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

- Prior papers assume $\text{Corr}(X, W) = 0$ (exogenous controls)
- i.e. the X 's (control variables) are themselves exogenous
- But this is rarely true in practice
- Other methods allow d and W to be correlated through $\text{Corr}(d, \varepsilon)$

$$y = \alpha d + X\beta + \underbrace{W\gamma + \nu}_{\varepsilon} \quad (2)$$

- Prior papers assume $\text{Corr}(X, W) = 0$ (exogenous controls)
- i.e. the X 's (control variables) are themselves exogenous
- But this is rarely true in practice
- Other methods allow d and W to be correlated through $\text{Corr}(d, \varepsilon)$
- but do not allow X and W to be correlated

Notational differences:

- d is X in DMP
- X is W_1
- W is W_2
- (α, β, γ) is $(\beta, \gamma_1, \gamma_2)$
- ν is a linear projection (e.g. $Y^{\perp X, W}$)
- y is capitalized in DMP

Terminology

Terminology

- long regression: regress Y on d , X and W

Terminology

- long regression: regress Y on d , X and W
 - this is the regression we *want* to run but can't because W is unobserved

Terminology

- long regression: regress Y on d , X and W
 - this is the regression we *want* to run but can't because W is unobserved
- medium regression: regress Y on d and X

Terminology

- long regression: regress Y on d , X and W
 - this is the regression we *want* to run but can't because W is unobserved
- medium regression: regress Y on d and X
 - this is the regression we *can* run, but which may be of limited use

Terminology

- long regression: regress Y on d , X and W
 - this is the regression we *want* to run but can't because W is unobserved
- medium regression: regress Y on d and X
 - this is the regression we *can* run, but which may be of limited use
- first-stage regression: regress d on X and W

Terminology

- **long regression:** regress Y on d , X and W
 - this is the regression we *want* to run but can't because W is unobserved
- **medium regression:** regress Y on d and X
 - this is the regression we *can* run, but which may be of limited use
- **first-stage regression:** regress d on X and W
 - helpful for conceptualizing what is meant by “selection” (on obs. or unobs.)

Terminology

- **long regression:** regress Y on d , X and W
 - this is the regression we *want* to run but can't because W is unobserved
- **medium regression:** regress Y on d and X
 - this is the regression we *can* run, but which may be of limited use
- **first-stage regression:** regress d on X and W
 - helpful for conceptualizing what is meant by “selection” (on obs. or unobs.)
- α_{long} = true treatment effect; α_{med} = the one we can feasibly estimate

We can use the first-stage regression to quantify selection:

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

then define the ratio of selection on unobservables to selection on observables by

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

then define the ratio of selection on unobservables to selection on observables by

$$r_d = \frac{\sqrt{\mathbb{V}(W\phi)}}{\sqrt{\mathbb{V}(X\psi)}}$$

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

then define the ratio of selection on unobservables to selection on observables by

$$r_d = \frac{\sqrt{\mathbb{V}(W\phi)}}{\sqrt{\mathbb{V}(X\psi)}}$$

Intuition:

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

then define the ratio of selection on unobservables to selection on observables by

$$r_d = \frac{\sqrt{\mathbb{V}(W\phi)}}{\sqrt{\mathbb{V}(X\psi)}}$$

Intuition:

- r_d is analogous to λ in previous video

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

then define the ratio of selection on unobservables to selection on observables by

$$r_d = \frac{\sqrt{\mathbb{V}(W\phi)}}{\sqrt{\mathbb{V}(X\psi)}}$$

Intuition:

- r_d is analogous to λ in previous video
- $\mathbb{V}(X\psi) \approx \text{Corr}(d, X\beta)$

We can use the first-stage regression to quantify selection:

$$d = X\psi + W\phi + \eta$$

then define the ratio of selection on unobservables to selection on observables by

$$r_d = \frac{\sqrt{\mathbb{V}(W\phi)}}{\sqrt{\mathbb{V}(X\psi)}}$$

Intuition:

- r_d is analogous to λ in previous video
- $\mathbb{V}(X\psi) \approx \text{Corr}(d, X\beta)$
- $\mathbb{V}(W\phi) \approx \text{Corr}(d, \varepsilon)$

How should we use this information?

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)
 - present this alongside R^2 , sample size, etc.

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)
 - present this alongside R^2 , sample size, etc.
 - it's a single number that summarizes robustness to selection on unobservables

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)
 - present this alongside R^2 , sample size, etc.
 - it's a single number that summarizes robustness to selection on unobservables
2. present the **identified set** of α_{long} as a function of \bar{r}_d

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)
 - present this alongside R^2 , sample size, etc.
 - it's a single number that summarizes robustness to selection on unobservables
2. present the **identified set** of α_{long} as a function of \bar{r}_d
 - e.g. $\alpha_{\text{long}} \in [\alpha_{\text{med}} - \delta(\bar{r}_d), \alpha_{\text{med}} + \delta(\bar{r}_d)]$ for some adjustment $\delta(\bar{r}_d)$

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)
 - present this alongside R^2 , sample size, etc.
 - it's a single number that summarizes robustness to selection on unobservables
2. present the **identified set** of α_{long} as a function of \bar{r}_d
 - e.g. $\alpha_{\text{long}} \in [\alpha_{\text{med}} - \delta(\bar{r}_d), \alpha_{\text{med}} + \delta(\bar{r}_d)]$ for some adjustment $\delta(\bar{r}_d)$
 - if $\bar{r}_d = 0$ then $\delta(0) = 0$ and this is just α_{med}

How should we use this information?

1. present the **breakdown point** (\bar{r}_d^{bp} , an upper bound on r_d such that $\alpha = 0$)
 - present this alongside R^2 , sample size, etc.
 - it's a single number that summarizes robustness to selection on unobservables
2. present the **identified set** of α_{long} as a function of \bar{r}_d
 - e.g. $\alpha_{\text{long}} \in [\alpha_{\text{med}} - \delta(\bar{r}_d), \alpha_{\text{med}} + \delta(\bar{r}_d)]$ for some adjustment $\delta(\bar{r}_d)$
 - if $\bar{r}_d = 0$ then $\delta(0) = 0$ and this is just α_{med}
 - set widens as \bar{r}_d grows larger

Sensitivity parameters $(\bar{r}_d, \bar{r}_Y, \bar{c})$

Sensitivity parameters $(\bar{r}_d, \bar{r}_Y, \bar{c})$

- \bar{r}_d measures how much the unobservables predict *treatment* compared to how much the observables predict treatment

$$\sqrt{\mathbb{V}(W\phi)} \leq \bar{r}_d \sqrt{\mathbb{V}(X\psi)}$$

Sensitivity parameters $(\bar{r}_d, \bar{r}_Y, \bar{c})$

- \bar{r}_d measures how much the unobservables predict *treatment* compared to how much the observables predict treatment

$$\sqrt{\mathbb{V}(W\phi)} \leq \bar{r}_d \sqrt{\mathbb{V}(X\psi)}$$

- \bar{r}_Y measures how much the unobservables predict the *outcome* compared to how much the observables predict the outcome

$$\sqrt{\mathbb{V}(W\gamma)} \leq \bar{r}_Y \sqrt{\mathbb{V}(X\beta)}$$

Sensitivity parameters $(\bar{r}_d, \bar{r}_Y, \bar{c})$

- \bar{r}_d measures how much the unobservables predict *treatment* compared to how much the observables predict treatment

$$\sqrt{\mathbb{V}(W\phi)} \leq \bar{r}_d \sqrt{\mathbb{V}(X\psi)}$$

- \bar{r}_Y measures how much the unobservables predict the *outcome* compared to how much the observables predict the outcome

$$\sqrt{\mathbb{V}(W\gamma)} \leq \bar{r}_Y \sqrt{\mathbb{V}(X\beta)}$$

- \bar{c} bounds how much X (jointly) can predict W ($\bar{c} = 0$ in all prior literature)

$$R_{W,X} \leq \bar{c}$$

(assumes W is a single omitted variable)

Formula for \bar{r}_d^{bp} is a function of $R_{Y \sim d \cdot X}^2$ and partial R^2 terms:

$$\bar{r}_d^{\text{bp}} = \left(\frac{R_{Y \sim d \cdot X}^2}{\frac{R_{d \sim X}^2}{1 - R_{d \sim X}^2} + R_{Y \sim d \cdot X}^2} \right)^{\frac{1}{2}}$$

Formula for \bar{r}_d^{bp} is a function of $R_{Y \sim d \cdot X}^2$ and partial R^2 terms:

$$\bar{r}_d^{\text{bp}} = \left(\frac{R_{Y \sim d \cdot X}^2}{\frac{R_{d \sim X}^2}{1 - R_{d \sim X}^2} + R_{Y \sim d \cdot X}^2} \right)^{\frac{1}{2}}$$

where $R_{Y \sim d \cdot X}^2$ = “the R^2 from a regression of y on d after partialling out both by X ”

Formula for \bar{r}_d^{bp} is a function of $R_{Y \sim d \cdot X}^2$ and partial R^2 terms:

$$\bar{r}_d^{\text{bp}} = \left(\frac{R_{Y \sim d \cdot X}^2}{\frac{R_{d \sim X}^2}{1 - R_{d \sim X}^2} + R_{Y \sim d \cdot X}^2} \right)^{\frac{1}{2}}$$

where $R_{Y \sim d \cdot X}^2$ = “the R^2 from a regression of y on d after partialling out both by X ”

Depends on just 2 features of the observed data:

Formula for \bar{r}_d^{bp} is a function of $R_{Y \sim d \cdot X}^2$ and partial R^2 terms:

$$\bar{r}_d^{\text{bp}} = \left(\frac{R_{Y \sim d \cdot X}^2}{\frac{R_{d \sim X}^2}{1 - R_{d \sim X}^2} + R_{Y \sim d \cdot X}^2} \right)^{\frac{1}{2}}$$

where $R_{Y \sim d \cdot X}^2$ = “the R^2 from a regression of y on d after partialling out both by X ”

Depends on just 2 features of the observed data:

1. Relationship between y and d (after adjusting for X)

Formula for \bar{r}_d^{bp} is a function of $R_{Y \sim d \cdot X}^2$ and partial R^2 terms:

$$\bar{r}_d^{\text{bp}} = \left(\frac{R_{Y \sim d \cdot X}^2}{\frac{R_{d \sim X}^2}{1 - R_{d \sim X}^2} + R_{Y \sim d \cdot X}^2} \right)^{\frac{1}{2}}$$

where $R_{Y \sim d \cdot X}^2$ = “the R^2 from a regression of y on d after partialling out both by X ”

Depends on just 2 features of the observed data:

1. Relationship between y and d (after adjusting for X)
2. Relationship between d and X

Checklist of what researchers should report:

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables
2. **Plot the identified set** as a function of \bar{r}_d

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables
2. **Plot the identified set** as a function of \bar{r}_d
 - Shows how conclusions change as you allow stronger selection on unobservables

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables
2. **Plot the identified set** as a function of \bar{r}_d
 - Shows how conclusions change as you allow stronger selection on unobservables
3. **Calibration via** ρ_k : Compare \bar{r}_d^{bp} to importance of each of the X 's to d

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables
2. **Plot the identified set** as a function of \bar{r}_d
 - Shows how conclusions change as you allow stronger selection on unobservables
3. **Calibration via** ρ_k : Compare \bar{r}_d^{bp} to importance of each of the X 's to d
 - If $\bar{r}_d^{\text{bp}} > \max_k \rho_k$, unobservables must be more important than any single observed control to overturn results

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables
2. **Plot the identified set** as a function of \bar{r}_d
 - Shows how conclusions change as you allow stronger selection on unobservables
3. **Calibration via** ρ_k : Compare \bar{r}_d^{bp} to importance of each of the X 's to d
 - If $\bar{r}_d^{\text{bp}} > \max_k \rho_k$, unobservables must be more important than any single observed control to overturn results
4. **(Optional) Common maximal impact:** Report \bar{r}^{bp} assuming $\bar{r}_d = \bar{r}_Y$

Checklist of what researchers should report:

1. **Breakdown point** \bar{r}_d^{bp} : Report alongside main regression results
 - Single number summarizing robustness to omitted variables
2. **Plot the identified set** as a function of \bar{r}_d
 - Shows how conclusions change as you allow stronger selection on unobservables
3. **Calibration via** ρ_k : Compare \bar{r}_d^{bp} to importance of each of the X 's to d
 - If $\bar{r}_d^{\text{bp}} > \max_k \rho_k$, unobservables must be more important than any single observed control to overturn results
4. **(Optional) Common maximal impact:** Report \bar{r}^{bp} assuming $\bar{r}_d = \bar{r}_Y$
 - Less conservative: $\bar{r}^{\text{bp}} \geq \bar{r}_d^{\text{bp}}$

Extensions

Extensions

1. Can this approach be used in nonlinear models like logit or probit?

Extensions

1. Can this approach be used in nonlinear models like logit or probit?
 - No, identification becomes much more complicated in nonlinear settings

Extensions

1. Can this approach be used in nonlinear models like logit or probit?
 - No, identification becomes much more complicated in nonlinear settings
2. Can this approach be used in panel data settings?

Extensions

1. Can this approach be used in nonlinear models like logit or probit?
 - No, identification becomes much more complicated in nonlinear settings
2. Can this approach be used in panel data settings?
 - I believe so; I think you can do everything on the within-transformed dataset

Extensions

1. Can this approach be used in nonlinear models like logit or probit?
 - No, identification becomes much more complicated in nonlinear settings
2. Can this approach be used in panel data settings?
 - I believe so; I think you can do everything on the within-transformed dataset
 - Main equation becomes

$$\dot{y} = \alpha \dot{d} + \dot{X}\beta + \underbrace{\dot{W}\gamma + \dot{\nu}}_{\dot{\varepsilon}} \quad (3)$$

where $\dot{y} = y_{it} - \bar{y}_i$, etc. Requires time-varying treatment!