An Omitted Variable Bias Framework for Sensitivity Analysis of Instrumental Variables

Carlos Cinelli* Chad Hazlett†

October 4, 2022

Abstract

We develop an "omitted variable bias" framework for sensitivity analysis of instrumental variable (IV) estimates that naturally handles multiple "side-effects" (violations of the exclusion restriction assumption) and "confounders" (violations of the ignorability of the instrument assumption), exploits expert knowledge to bound sensitivity parameters, and can be easily implemented with standard software. More specifically, we introduce sensitivity statistics for routine reporting, such as (extreme) *robustness values* for IV estimates, describing the minimum strength that omitted variables need to have to change the conclusions of an IV study. Next we provide visual displays that fully characterize the sensitivity of IV point-estimates and confidence intervals to violations of the standard IV assumptions. Finally, we offer formal bounds on the worst possible bias under the assumption that the maximum explanatory power of omitted variables are no stronger than a multiple of the explanatory power of observed variables. Conveniently, we also show that many pivotal conclusions regarding the sensitivity of the IV estimate (e.g. tests against the null hypothesis of zero causal effect) can be reached simply through separate sensitivity analyses of the effect of the instrument on the treatment (the "first stage") and the effect of the instrument on the outcome (the "reduced form"). We apply our methods in a running example that uses instrumental variables to estimate the returns to schooling.

*Assistant Professor, Department of Statistics, University of Washington.
Email: cinelli@uw.edu. URL: carloscinelli.com.

†Associate Professor, Departments of Statistics and Political Science, University of California Los Angeles.
Email: chazlett@ucla.edu. URL: chadhazlett.com.

1 Introduction

Unobserved confounding often complicates efforts to make causal claims from observational data (e.g. Pearl, 2009; Imbens and Rubin, 2015a; Rosenbaum, 2017). Instrumental variable (IV) regression offers a powerful and widely used tool to address unobserved confounding, by exploiting "exogenous" sources of variation of the treatment (e.g. Wright, 1928; Bowden and Turkington, 1990; Angrist et al., 1996; Angrist and Pischke, 2009). IV methods have also become a vital tool in the analysis of randomized experiments with imperfect compliance (Robins, 1989; Balke and Pearl, 1994, 1997; Angrist et al., 1996). These qualities have made IV methods "a central part of the econometrics canon since the first half of the twentieth century" (Imbens, 2014, p.324). Beyond economics, instrumental variables are prominent tools in the arsenal of investigators seeking to make causal claims across the social sciences, epidemiology, medicine, genetics, and other fields (see e.g. Hernán and Robins, 2006; Didelez and Sheehan, 2007; Baiocchi et al., 2014; Burgess and Thompson, 2015).

Yet, IV methods carry their own set of demanding assumptions. Principally, conditionally on certain observed covariates, an instrumental variable must not itself be confounded with the outcome, and it should influence the outcome only by influencing uptake of the treatment. These assumptions can be violated by omitted confounders of the instrument-outcome association, and by omitted "side-effects" of the instrument that influence the outcome via paths not containing the treatment.[1] Although in certain cases the IV assumptions may entail testable implications (Pearl, 1995; Bonet, 2001; Swanson et al., 2018; Gunsilius, 2020; Kédagni and Mourifié, 2020), they are often unverifiable and must be defended by appealing to domain knowledge and theoretical arguments. Whether a given IV study identifies the causal effect of interest, then, turns on debates as to whether these assumptions hold.

Particularly in recent years, economists and other scholars have adopted a more skeptical posture towards IV methods, emphasizing the importance of both defending the credibility of these assumptions as well as assessing the consequences of their failures (see e.g., Deaton, 2009; Heckman and Urzua, 2010). Extensive reviews of many widely-used instrumental variables, such as weather, religion, sibling structure or ethnolinguistic fractionalization, have cataloged several plausible violations of the exclusion restriction for such instruments (Gallen, 2020; Mellon, 2020). More worrisome, if the IV assumptions fail to hold, it is well known that the bias of the IV estimate may be *worse* than the original confounding bias of the simple regression estimate that the IV was supposed to address (Bound et al., 1995). Therefore, researchers are also advised to perform *sensitivity analyses* to assess the degree of violation of the IV assumptions that would be required to alter the conclusions of an IV study. While a number of sensitivity analyses for IV have been proposed (DiPrete and Gangl, 2004; Altonji et al., 2005; Small, 2007; Small and Rosenbaum, 2008; Conley et al., 2012; Wang et al.,

[1] In the recent IV literature, the first assumption is usually called *exogeneity*, *ignorability*, *unconfoundedness* or *independence* of the instrument, whereas the second assumption is called the *exclusion* restriction (Angrist and Pischke, 2009; Pearl, 2009; Imbens and Rubin, 2015a; Swanson et al., 2018). In earlier econometric works, these two assumptions were often combined into one, also labeled the "exclusion restriction" (Imbens, 2014).

2018; Jiang et al., 2018; Cinelli et al., 2019), such sensitivity analyses still remain rare in practice.[2]

We suggest several reasons for this slow uptake. First, the traditional approach for the sensitivity of IV has focused on parameterizing violations of the IV assumptions with a single parameter summarizing the "bias" in the association of the instrument with the outcome. While this parameterization may be well-suited when the bias is only due to the direct effect of the instrument on the outcome (not through the treatment), this parameterization is not as straightforward to use when reasoning about multiple side-effects or confounders of the instrument, in which case that sensitivity parameter is a complicated composite of many source of bias (see Appendix E for a comparison of our proposal with the traditional approach to the sensitivity of IV). Second, while users of IV methods are instructed to routinely report quantities to diagnose certain inferential problems such as "weak instruments" (eg, Stock and Yogo, 2002) we lack sensitivity statistics that can quickly communicate how robust an IV study is to violations in the form of omitted confounders or side-effects of the instrument. Finally, it is often difficult to connect the formal results of a sensitivity analysis to a cogent argument about what types of biases can be effectively ruled out by expert knowledge.

In this paper, we develop an omitted variable bias (OVB) framework for assessing the sensitivity of IV estimates that aims to address these challenges and improve the usability and uptake of sensitivity analysis for IV.[3] Building on recent developments of OVB for ordinary least squares (OLS) (Cinelli and Hazlett, 2020), we develop a suite of sensitivity analysis tools for IV that: (i) has correct test size (or confidence interval coverage) regardless of instrument strength; (ii) naturally handles violations due to multiple "side-effects" and "confounders," possibly acting non-linearly; (iii) is well suited for routine reporting; (iv) exploits expert knowledge to bound sensitivity parameters; and, (v) can be easily implemented with standard software.

More specifically, we introduce two main sensitivity statistics for IV estimates: (i) the *robustness value* (RV) describes the minimum strength of association (in terms of partial R^2) that omitted variables (side-effects or confounders) need to have, both with the instrument and with the untreated potential outcome, such that they are capable of changing the conclusions of the study; and (ii) the *extreme robustness value*, which describes the minimal strength of association that omitted variables need to have with the *instrument alone* (regardless of their association with the untreated potential outcome) in order to be problematic. The routine reporting of these quantities provide a quick and

[2]For instance, in economics, only 1 out of 27 papers using instrumental variables published in the *American Economic Review* in 2020 performed formal sensitivity analysis. In political science, this number was 1 out of 12 papers, considering the top three general interest journals (*American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*) for 2019.

[3]We focus on the "just-identified" case with one treatment and one instrument. One reason for this is that a thorough consideration of the identification assumptions and how they may be violated is already complicated enough with a single instrument (Angrist and Pischke, 2009). Second, and relatedly, in most applied settings, the single-instrument and single-treatment setup is the most common. For example, in a broad review of papers in the *American Economic Review* and 15 other journals of the *American Economic Association*, Young (2022) finds that 80% of IV regressions were of this type. Finally, in many "multiple instrument" studies, it is not uncommon for researchers to also report and give special focus to the analysis of their "best" instrument (Angrist and Pischke, 2009), or to combine multiple instruments into a single instrument, for example, constructing an allele score in Mendelian Randomization (Burgess and Thompson, 2015; Cinelli et al., 2022). Extension of the tools we develop here to the scenario with multiple instruments and treatments is object of future investigations.

simple way to improve the transparency and facilitate the assessment of the credibility of IV studies. Next, we offer intuitive graphical tools for investigators to assess how postulated confounding of any degree would alter the IV hypothesis tests, as well as lower or upper limits of confidence intervals. Finally, these tools can be supplemented with formal bounds on the worst possible bias that side-effects or confounders could cause, under the assumption that the maximum explanatory power of these omitted variables are no stronger than a chosen multiple of the explanatory power of one or more observed variables.

Conveniently, considering that investigators are already well advised to carefully examine their "first stage" (the effect of the instrument on the treatment) and "reduced form" (the effect of the instrument on the outcome) (e.g. Angrist and Krueger, 2001; Angrist and Pischke, 2009) our analysis shows that indeed many pivotal conclusions regarding the sensitivity of the IV estimate can in fact be reached simply through separate sensitivity analyses of these two familiar auxiliary OLS estimates[4]. First, if researchers are interested in the null hypothesis of *zero effect,* all recent OVB tools for OLS (Cinelli and Hazlett, 2020; Cinelli et al., 2020) can simply be directly applied to the reduced-form regression, and confounders or side-effects shown to be problematic there are equally problematic for IV. Second, if interest lies in assessing not just the null of zero, but biases that bring the estimate partway to zero or beyond it, then the robustness of the IV estimate formally reduces to the minimum of the robustness of the reduced-form and the robustness of the first-stage regressions.

A final contribution of this paper is that, while developing OVB tools for IV, we extended the OVB results of Cinelli and Hazlett (2020) providing a new way to perform sensitivity analysis that simply replaces a conventional critical value (e.g. 1.96) with a novel "bias-adjusted" critical value that accounts for a postulated degree of omitted variable bias. Notably, this correction on the critical value does not depend on the data, and can be computed by simply postulating a hypothetical partial R^2 of the omitted variables with the dependent and independent variables of the OLS regression. Researchers, readers, and reviewers can thus quickly and easily perform sensitivity analysis by simply substituting traditional thresholds with bias-adjusted thresholds, when testing a particular null hypothesis, or when constructing confidence intervals. We believe the extreme simplicity of this approach will further aid in the widespread adoption of sensitivity analysis in applied work.

In what follows, Section 2 introduces the running example and provides the essential background on the main IV estimators, all of which depend upon OLS. Next, Section 3 extends the OVB framework of Cinelli and Hazlett (2020), which not only improves the sensitivity tools for OLS, but greatly simplifies the analysis for the IV setting. Section 4 then develops an OVB framework for IV, first showing what can be gleaned from the first-stage and reduced-form regressions alone, then establishing the necessary OVB-type results for a complete sensitivity analysis of the IV estimate. Section 5 returns to our running example to show how these results can can be deployed in practice. Finally, we offer concluding remarks in Section 6. Open-source software for `R`, `Python` and `Stata`

[4]In the context of randomization inference, similar observations can be found in Rosenbaum (1996, 2002); Imbens and Rosenbaum (2005); Small and Rosenbaum (2008); Keele et al. (2017) and Rosenbaum (2017).

implements the methods discussed in this paper.[5]

2 Running example

We begin by introducing the running example and briefly reviewing the required background on instrumental variables.

Ordinary least squares and the OVB problem

Many observational studies have established a positive and large association between educational achievement and earnings using regression analysis (Card, 1999). Here we consider the work of Card (1993), which employed a sample of $n = 3,010$ individuals from the National Longitudinal Survey of Young Men (NLSYM). Considering the following multiple linear regression

$$Y = \hat{\tau}_{\text{OLS,res}} D + \boldsymbol{X}\hat{\beta}_{\text{OLS,res}} + \hat{\varepsilon}_{\text{OLS,res}} \tag{1}$$

where Y denotes *Earnings* and measures the log transformed hourly wages of the individual[6], D denotes *Education* and consists of an integer-valued variable indicating the completed years of education of the individual, and the matrix $\boldsymbol{X}$ comprises race, experience, and a set of regional factors, Card concluded that each additional year of schooling was associated with approximately 7.5% higher wages (i.e, $\hat{\tau}_{\text{OLS,res}} \approx 0.075$; see Table 5 in Appendix F.).

Educational achievement, however, is not randomly assigned; perhaps individuals who obtain more education have higher wages due to other reasons, such as coming from wealthier families, or having higher levels of some unobserved characteristic, such as "ability" or "motivation." If data on these variables were available, then further adjustment for such variables would be able to capture the causal effect of educational attainment on schooling, as in

$$Y = \hat{\tau}_{\text{OLS}} D + \boldsymbol{X}\hat{\beta}_{\text{OLS}} + \boldsymbol{U}\hat{\gamma}_{\text{OLS}} + \hat{\varepsilon}_{\text{OLS}} \tag{2}$$

where $\boldsymbol{U}$ denotes a set of variables that, along with $\boldsymbol{X}$, is sufficient to eliminate confounding concerns[7]. Such detailed information on individuals, however, is not available, and researchers will not even agree upon which variables $\boldsymbol{U}$ are needed. In the absence of such variables, regression estimates that adjust for only a partial list of characteristics (such as $\boldsymbol{X}$) may suffer from "omitted variable

[5] Sensitivity analysis of the reduced form, first stage, and Anderson-Rubin regression for a specific null hypothesis can already be performed using the `R`, `Python` and `Stata` package `sensemakr` (Cinelli et al., 2020; LaPierre et al., 2021). Additional functionality, such as contour plots with lower and upper limits of the Anderson-Rubin confidence interval, is forthcoming.

[6] In this case, regression coefficients can be conveniently interpreted, approximately, as percent changes in earnings.

[7] I.e, the set $\{\boldsymbol{X}, \boldsymbol{U}\}$ is sufficient to render the treatment assignment ignorable. Equivalently, in graphical terms, the set would satisfy the backdoor (or, more generally, the adjustment) criterion (see Pearl, 2009; Angrist and Pischke, 2009; Imbens and Rubin, 2015b; Shpitser et al., 2012; Perkovic et al., 2018; Cinelli et al., 2021). Beyond ignorability, if the treatment effect is heterogeneous, this may affect the causal interpretation of the regression coefficient $\hat{\tau}_{\text{OLS}}$ (see, e.g, Angrist and Pischke, 2009).

bias" (Angrist and Pischke, 2009; Cinelli and Hazlett, 2020) and are likely to overestimate the "true" returns to schooling.

Instrumental variables as a solution to the OVB problem

Instrumental variable methods offer an alternative route to estimate the causal effect of schooling on earnings without having data on the unobserved variables $\boldsymbol{U}$. The key for such methods to work is to find a new variable (the "instrument") that changes the incentives to educational achievement, but is associated with earnings only through its effect on education.

To that end, Card (1993) proposed exploiting the role of geographic differences in college accessibility. In particular, consider the variable *Proximity*, encoding an indicator of whether the individual grew up in an area with a nearby accredited 4-year college, which we denote by Z. Students who grow up far from the nearest college may face higher educational costs, discouraging them from pursuing higher level studies. Next, and most importantly, Card (1993) argues that, conditional on the set of observed variables $\boldsymbol{X}$ (available on the NLSYM), whether one lives near a college is not itself confounded with earnings, nor does proximity to college affect earnings apart from its effect on years of education. If we believe such assumptions hold it is possible to recover a valid estimate of the (weigthed average of local) average treatment effect(s) of *Education* on *Earnings* by simply taking the ratio of two OLS coefficients, one measuring the effect of *Proximity* on *Earnings*, and another measuring the effect of *Proximity* on *Education*.[8]

More precisely, we estimate two OLS models

$$\textbf{First Stage:} \quad Y = \hat{\theta}_{\text{res}} Z + \boldsymbol{X}\hat{\psi}_{\text{res}} + \hat{\varepsilon}_{d,\text{res}} \tag{3}$$

$$\textbf{Reduced Form:} \quad Y = \hat{\lambda}_{\text{res}} Z + \boldsymbol{X}\hat{\beta}_{\text{res}} + \hat{\varepsilon}_{y,\text{res}} \tag{4}$$

Throughout the paper we refer to these equations as the "first stage" (Equation 3) and the "reduced form" (Equation 4), as these are now common usage (Angrist and Pischke, 2009, 2014; Imbens and Rubin, 2015a; Andrews et al., 2019).[9] The coefficient for *Proximity* (Z) on the first-stage regression, $\hat{\theta}_{\text{res}} \approx 0.32$, reveals that those who grew up near a college indeed have higher educational attainment, having completed an additional 0.32 years of education, on average. Likewise, the coefficient for *Proximity* (Z) on the reduced-form regression, $\hat{\lambda}_{\text{res}} \approx 0.042$, suggests that those who grew up near

[8] This identification result requires further functional restrictions on the data-generating process, such as linearity or monotonicity. Conditions that allow a causal interpretation of the "traditional" IV estimand (also known as the "2SLS estimand") are extensively discussed elsewhere and will not be reviewed here, see Angrist et al. (1996); Angrist and Pischke (2009); Imbens (2014); Swanson et al. (2018); Słoczyński (2020) and Blandhol et al. (2022). In particular, Blandhol et al. (2022) provides necessary and sufficient conditions for a "weakly causal" interpretation of the traditional IV estimand. Here we assume the researcher has already performed the required identification analysis, and decided that she is interested in the results of Equations 7, 8 and 9, controlling both for observed covariates $\boldsymbol{X}$ and unobserved covariates $\boldsymbol{W}$. We note the bulk of current applied work using instrumental variables takes this form, and non-parametric estimation is still rare in practice (Blandhol et al., 2022, p.11). It is nevertheless possible to extend these tools to nonparametric settings leveraging recent results in Chernozhukov et al. (2022). We leave this to future work.

[9] Though now well established, these labels abuse the original meaning of the terminology, since both regressions are in their "reduced form." Equation 3 is called the "first stage" due to its operational role on two-stage least squares estimation (see Appendix A). See also Imbens (2014) and Andrews et al. (2019).

a college have 4.2% higher earnings. The IV estimate is then given by the ratio

$$\hat{\tau}_{\text{res}} := \frac{\hat{\lambda}_{\text{res}}}{\hat{\theta}_{\text{res}}} \approx \frac{0.042}{0.319} \approx 0.132 \tag{5}$$

The value of $\hat{\tau}_{\text{res}} \approx 0.132$ suggests that, contrary to the OLS estimate of 7.5%, and perhaps surprisingly, each additional year of schooling instead raises wages by much more—13.2%.

The ratio of Equation 5 is sometimes called the *indirect least squares* (ILS) estimator, the "ratio of coefficients" estimator, or, in the case of a binary instrument, the "Wald estimator" (Wald, 1940). Inference in the ILS framework is usually performed using the delta-method. A closely related approach for instrumental variable estimation is denoted by "two-stage least squares" (2SLS), in which one saves the predictions of the first-stage regression, and then regress the outcome on these fitted values. By the Frisch-Waugh-Lovell (FWL) theorem (Frisch and Waugh, 1933; Lovell, 1963, 2008) one can readily show that 2SLS and ILS are numerically identical (see Appendix A).

Anderson-Rubin regression and Fieller's theorem. The methods of ILS and 2SLS may prove unreliable when the first-stage coefficient is "close" to zero, relative to the sampling variability of its estimator. This is known as the "weak instrument" problem. Two alternative procedures that allow constructing confidence intervals with correct coverage, regardless of the "strength" of the first stage, are the proposals of Anderson and Rubin (1949) and Fieller (1954).[10]

The Anderson-Rubin approach (Anderson and Rubin, 1949) starts by creating the random variable $Y_{\tau_0} := Y - \tau_0 D$ in which we subtract from Y a "putative" causal effect of D, namely, τ_0. If Z is a valid instrument, under the null hypothesis $H_0 : \tau = \tau_0$, we should not see an association between Y_{τ_0} and Z, conditional on $\boldsymbol{X}$. In other words, if we run the OLS model

$$\textbf{Anderson-Rubin:} \quad Y_{\tau_0} = \hat{\phi}_{\tau_0,\text{res}} Z + \boldsymbol{X}\hat{\beta}_{\tau_0,\text{res}} + \hat{\varepsilon}_{\tau_0,\text{res}} \tag{6}$$

we should find that $\hat{\phi}_{\tau_0,\text{res}}$ is equal to zero, but for sampling variation. To test the null hypothesis $H_0 : \phi_{\tau_0,\text{res}} = 0$ in the Anderson-Rubin regression is thus equivalent to test the null hypothesis $H_0 : \tau = \tau_0$. The $1 - \alpha$ confidence interval is constructed by collecting all values τ_0 such that the null hypothesis $H_0 : \phi_{\tau_0,\text{res}} = 0$ is not rejected at the chosen significance level α. This approach is numerically identical to Fieller's theorem (Fieller, 1954). Finally, it is convenient to define the point estimate $\hat{\tau}_{\text{AR,res}}$ as the value τ_0 which makes $\hat{\phi}_{\tau_0,\text{res}}$ exactly equal to zero. By the FWL theorem, we can easily show that this point estimate is numerically identical to that of 2SLS and ILS. Details and derivations of these algebraic identities (and differences) are provided in Appendix A.

[10] See Andrews et al. (2019) for an extensive review of inference with weak instruments. An intuitive visual comparison between the delta-method and Fieller's approach is given by Hirschberg and Lye (2010, 2017).

The IV estimate itself may suffer from OVB

The previous IV estimate relies on the assumption that, conditional on $\boldsymbol{X}$, *Proximity* and *Earnings* are unconfounded, and the effect of *Proximity* on *Earnings* must go entirely through *Education*. As it is often the case, neither assumption is easy to defend in this setting. First, some of the same factors that might confound the relationship between *Education* and *Earnings* could similarly confound the relationship of *Proximity* and *Earnings* (e.g. family wealth or connections). Second, as argued in Card (1993), the presence of a college nearby may be associated with high school quality, which in its turn also affects earnings. Finally, other geographic confounders can make some localities likely to both have colleges nearby and lead to higher earnings. These are only coarsely conditioned on by the observed regional indicators, and residual biases may still remain.

In sum, instead of adjusting only for $\boldsymbol{X}$ as in the previous first-stage and reduced-form regressions, we should have adjusted for both the observed covariates $\boldsymbol{X}$ and unobserved covariates $\boldsymbol{W}$ as in

$$\textbf{First Stage:} \quad Y = \hat{\theta} Z + \boldsymbol{X}\hat{\psi} + \boldsymbol{W}\hat{\delta} + \hat{\varepsilon}_d \tag{7}$$

$$\textbf{Reduced Form:} \quad Y = \hat{\lambda} Z + \boldsymbol{X}\hat{\beta} + \boldsymbol{W}\hat{\gamma} + \hat{\varepsilon}_y \tag{8}$$

Or, in the Anderson-Rubin approach, we should have run instead

$$\textbf{Anderson Rubin:} \quad Y_{\tau_0} = \hat{\phi}_{\tau_0} Z + \boldsymbol{X}\hat{\beta}_{\tau_0} + \boldsymbol{W}\hat{\gamma}_{\tau_0} + \hat{\varepsilon}_{\tau_0} \tag{9}$$

Where $\boldsymbol{W}$ stands for all unobserved factors necessary to make *Proximity* a valid instrument for the effect of *Education* on *Earnings* (e.g, *Family Wealth*, *High School Quality*, *Place of Residence*)[11]. Our task is thus to characterize how the IV point estimates and confidence intervals would have changed due to the inclusion of omitted variables $\boldsymbol{W}$. Since, at their core, IV approaches rely on OLS estimation, we should then be able to leverage all recent developments of OVB tools for OLS (Cinelli and Hazlett, 2020) for examining the sensitivity of IV.

3 Omitted variable bias with the partial R^2 parameterization

In this section, we extend the results of Cinelli and Hazlett (2020) regarding the partial R^2 parameterization of the OVB formula for OLS. In particular, we introduce *bias-adjusted* critical values for OLS, and show how sensitivity analysis can be performed by simply substituting traditional critical values with the adjusted ones. Notably, this adjustment does not depend on the data, and it consists of a simple correction on the critical value based solely on the hypothetical strength of omitted variables (and degrees of freedom). Next we introduce new sensitivity statistics for routine reporting, such as extreme robustness values, characterizing the bare minimum strength that omitted variables

[11] See causal diagrams in Figure 4 of Appendix F for "canonical" models illustrating the traditional assumptions of IV and their violations. Equivalent assumptions can be articulated in the potential outcomes framework (Angrist et al., 1996; Pearl, 2009; Swanson et al., 2018). Here, we assume the researcher has already established that Z is a valid IV for the causal effect of D on Y conditional on the set $\{\boldsymbol{X}, \boldsymbol{W}\}$.

must have to overturn certain conclusions. We formalize such statistics as an answer to an inverse question regarding a set of compatible inferences given bounds on the strength of omitted variables. Finally, we derive a novel bound on the strength of omitted variables on the basis of comparison with observed variables. These results are not only useful on their own, but they greatly simplify the development of a suite sensitivity analysis tools for IV in Section 4.

3.1 Sensitivity in an omitted variable bias framework

For concreteness, suppose we are interested in the coefficient $\hat{\lambda}$ of the regression equation of the outcome Y on the instrument Z, adjusting for a set of observed covariates $\boldsymbol{X}$ and a single *unobserved* covariate W (we generalize to multivariate W below),

$$Y = \hat{\lambda} Z + \boldsymbol{X}\hat{\beta} + \hat{\gamma} W + \hat{\varepsilon}_y \tag{10}$$

where Y, Z and W are $(n \times 1)$ vectors, $\boldsymbol{X}$ is an $(n \times p)$ matrix, with n observations (including a constant), $\hat{\lambda}$, $\hat{\beta}$ and $\hat{\gamma}$ are the OLS estimates of the regression of Y on Z, $\boldsymbol{X}$ and W, and $\hat{\varepsilon}_y$ the corresponding residuals.

However, when W is unobserved the investigator is instead forced to estimate the *restricted* model,

$$Y = \hat{\lambda}_{\text{res}} Z + \boldsymbol{X}\hat{\beta}_{\text{res}} + \hat{\varepsilon}_{y,\text{res}} \tag{11}$$

where $\hat{\lambda}_{\text{res}}$ and $\hat{\beta}_{\text{res}}$ are the coefficients of the restricted OLS adjusting for Z and $\boldsymbol{X}$ alone, and $\hat{\varepsilon}_{y,\text{res}}$ its corresponding residual. The OVB framework seeks to answer the following question: how do the inferences for λ_{res} from the restricted OLS model (omitting W), compare with the inferences for our actual target parameter λ from the full OLS model (adjusting for W)?

Adjusted estimates and standard errors

Let $R^2_{Y\sim W|Z,\boldsymbol{X}}$ denote the partial R^2 of W with Y, after controlling for Z and $\boldsymbol{X}$, and let $R^2_{Z\sim W|\boldsymbol{X}}$ denote the partial R^2 of W with Z after adjusting for $\boldsymbol{X}$. Given the point estimate and (estimated) standard error of the restricted model actually run, $\hat{\lambda}_{\text{res}}$ and $\widehat{\text{se}}(\hat{\lambda}_{\text{res}})$, the values $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$ are sufficient to recover $\hat{\lambda}$ and $\widehat{\text{se}}(\hat{\lambda})$ (Cinelli and Hazlett, 2020). More precisely, define $\widehat{\text{bias}}(\lambda) := \hat{\lambda}_{\text{res}} - \hat{\lambda}$ as the difference between the restricted estimate and the full estimate. Then,

$$|\widehat{\text{bias}}(\lambda)| = \sqrt{\frac{R^2_{Y\sim W|Z,\boldsymbol{X}} R^2_{Z\sim W|\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}\,\text{df}} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) = \text{BF}\sqrt{\text{df}} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) \tag{12}$$

Where here $\text{df} = n - p - 1$ stands for the residual degrees of freedom from the restricted model actually run. For notational convenience, and to aid interpretation, we define the term

$$\text{BF} := \sqrt{\frac{R^2_{Y\sim W|Z,\boldsymbol{X}} R^2_{Z\sim W|\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}} \tag{13}$$

as the "bias factor" of W, which is the part of the bias solely determined by $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$. Likewise, the standard error of the full model can be recovered with

$$\widehat{\text{se}}(\hat{\lambda}) = \sqrt{\frac{1 - R^2_{Y\sim W|Z,\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}} \left(\frac{\text{df}}{\text{df} - 1}\right)} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) = \text{SEF}\sqrt{\text{df}\,/(\text{df} - 1)} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) \tag{14}$$

where again, for convenience, we define

$$\text{SEF} := \sqrt{\frac{1 - R^2_{Y\sim W|Z,\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}} \tag{15}$$

as the "standard error factor" of W, summarizing the factor of the standard error which is solely determined by the sensitivity parameters $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$. Note that SEF consists of the square-root of the product of the familiar "variance inflation factor," $1/\left(1 - R^2_{Z\sim W|\boldsymbol{X}}\right)$ and what could be labeled the "variance reduction factor," $1 - R^2_{Y\sim W|Z,\boldsymbol{X}}$. Cinelli and Hazlett (2020, Sec. 4.2) provide further discussion. Although simple, Equations 12 and 14 form the basis of sensitivity analyses for point estimates, standard errors and t-values in terms of sensitivity parameters $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$.

Multiple unobserved variables. For simplicity of exposition, throughout the text we usually refer to a single omitted variable W. These results, however, can be used for performing sensitivity analyses considering multiple omitted variables $\boldsymbol{W} = [W_1, W_2, \ldots, W_n]$, and thus also non-linearities and functional form misspecification of observed variables. In such cases, barring an adjustment in the degrees of freedom, the equations are conservative, and reveal the maximum bias a multivariate $\boldsymbol{W}$ with such pair of partial R^2 values could cause (Cinelli and Hazlett, 2020, Sec. 4.5).

3.2 Bias-adjusted critical values and set of compatible inferences

We now introduce a novel correction researchers can make to traditional critical values in order to account for omitted variable bias. That is, traditional confidence intervals account for sampling uncertainty, and are constructed by multiplying the standard error of the coefficient by a critical value (for example, in large enough samples, 1.96 for a 95% confidence level). We show that replacing this traditional critical value with a *bias-adjusted critical value*, which we introduce here, accounts for both sampling uncertainty and systematic biases due to omitted variables with a given postulated

strength. Although simple, this perspective will prove useful for OLS in general, but especially for instrumental variables, where we apply it to the test inversion employed in the Anderson-Rubin approach (Section 4).

Specifically, let $t^*_{\alpha,\mathrm{df}-1}$ denote the critical value for a standard t-test with significance level α and df -1 degrees of freedom. Now let $\mathrm{LL}_{1-\alpha}(\lambda)$ be the lower limit and $\mathrm{UL}_{1-\alpha}(\lambda)$ be the upper limit of a $1-\alpha$ confidence interval for λ in the full model, i.e.,

$$\mathrm{LL}_{1-\alpha}(\lambda) := \hat{\lambda} - t^*_{\alpha,\mathrm{df}-1} \times \widehat{\mathrm{se}}(\hat{\lambda}), \quad \mathrm{UL}_{1-\alpha}(\lambda) := \hat{\lambda} + t^*_{\alpha,\mathrm{df}-1} \times \widehat{\mathrm{se}}(\hat{\lambda}), \tag{16}$$

Considering the direction of the bias that further reduces the lower limit, as well as the direction that further increases the upper limit, Equations 12 and 14 imply that both quantities can be written as a function of the restricted estimates and a new multiplier (see Appendix B)

$$\mathrm{LL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\mathrm{res}} - t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}), \quad \mathrm{UL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\mathrm{res}} + t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) \tag{17}$$

where $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ is the *bias-adjusted critical value*

$$t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} := \mathrm{SEF}\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^*_{\alpha,\mathrm{df}-1} + \mathrm{BF}\sqrt{\mathrm{df}}. \tag{18}$$

As the subscript $\boldsymbol{R}^2 = \{R^2_{Y\sim W|Z,\boldsymbol{X}}, R^2_{Z\sim W|\boldsymbol{X}}\}$ conveys, $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ depends on both sensitivity parameters. Notably, this correction does not depend on the data (but for the degrees of freedom). The adjusted critical value $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ uniquely determines the extreme points of the confidence interval for λ after adjusting for an omitted variable W with a given pair of partial R^2. Further, to test the more general null hypothesis of a change of $(100 \times q^*)\%$ of the current estimate $\hat{\lambda}_{\mathrm{res}}$ at the α level, it suffices to rescale the original t-value by q^* and compare this to the adjusted critical threshold $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$. This allows researchers, readers, and reviewers to quickly assess the robustness of reported findings to omitted variables of any postulated strength.

For a numerical example, it is instructive to consider the case in which the omitted variable W has equal strength with Y and Z, i.e, $R^2_{Y\sim W|Z,\boldsymbol{X}} = R^2_{Z\sim W|\boldsymbol{X}} = R^2$. We then have that SEF $= 1$ and BF $= R^2/\sqrt{1-R^2}$ resulting in a very simple correction formula,

$$t^{\dagger}_{\alpha,\mathrm{df}-1,R^2,R^2} \approx t^*_{\alpha,\mathrm{df}-1} + \frac{R^2}{\sqrt{1-R^2}}\sqrt{\mathrm{df}}, \tag{19}$$

where here we also approximate $\sqrt{\mathrm{df}/(\mathrm{df}-1)} \approx 1$. Table 1 shows the adjusted critical values (at the 5% significance level) for this case, considering different strengths of the omitted variable W, ranging from $R^2 = 0$ to $R^2 = .1$, and various sample sizes, ranging from df $= 100$ to df $= 1,000,000$.

The first row of Table 1 starts with the ideal case of zero residual biases. Here the traditional critical threshold is approximately 1.96 (1.98 when df $= 100$) regardless of sample size. Moving to the second row forward, we now perform the omitted variable bias correction of Equation 19. Tests using these new critical values thus account both for sampling uncertainty and residual biases with

the postulated strength, as given by $R^2_{Y\sim W|Z,\boldsymbol{X}} = R^2_{Z\sim W|\boldsymbol{X}} = R^2$. Note how $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ *increases* the larger the sample size. For example, consider the second row, with an adjusted critical value that is robust to omitted variables that explain 1% of the residual variation of both Z and Y, i.e, $R^2_{Y\sim W|Z,\boldsymbol{X}} = R^2_{Z\sim W|\boldsymbol{X}} = .01$. When $\mathrm{df} = 100$, this leads to an adjusted critical value of $\approx$ 2.1, whereas if $\mathrm{df} = 1,000,000$, this leads to the much higher threshold of $\approx$ 12.

R^2	Degrees of Freedom (sample size)				
	100	1,000	10,000	100,000	1,000,000
0.00	1.98	1.96	1.96	1.96	1.96
0.01	2.08	2.28	2.97	5.14	12.01
0.02	2.19	2.60	3.98	8.35	22.16
0.03	2.29	2.92	5.01	11.59	32.42
0.04	2.39	3.25	6.04	14.87	42.78
0.05	2.50	3.58	7.09	18.18	53.26
0.06	2.60	3.92	8.15	21.53	63.85
0.07	2.71	4.26	9.22	24.91	74.55
0.08	2.82	4.60	10.30	28.34	85.37
0.09	2.93	4.94	11.39	31.79	96.31
0.10	3.04	5.29	12.50	35.29	107.37

Table 1: Bias-adjusted critical values, $t^{\dagger}_{\alpha,\mathrm{df}-1,R^2,R^2}$, for different strengths of the omitted variable W (with $R^2_{Y\sim W|Z,\boldsymbol{X}} = R^2_{Z\sim W|\boldsymbol{X}} = R^2$) and various sample sizes. Significance level $\alpha = 5\%$.

This behaviour is simply a consequence of the well-known, but often ovelooked fact that larger samples will eventually detect any signal, even if such signal is spurious. Thus, as the sample size grows, a higher threshold is needed in order to protect inferences against systematic biases.

Compatible inferences given bounds on partial R^2

Given hypothetical values for $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$, the previous results allow us to determine the exact changes in inference regarding a parameter of interest due to the inclusion of W with such strength. Often, however, the analyst does not know the exact strength of omitted variables, and wishes to investigate the *worst* possible inferences that could be induced by a W with bounded strength, for instance, $R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}}$. That is, we wish to find the maximum adjusted critical value due to an omitted variable W with *at most* such strength. Writing $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ as a function of the sensitivity parameters $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$, we solve the maximization problem (see appendix)

$$\max_{R^2_{Y\sim W|Z,\boldsymbol{X}},R^2_{Z\sim W|\boldsymbol{X}}} t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \quad \text{s.t.} \quad R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}, \quad R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \tag{20}$$

Note that, although this maximum is often reached at the extrema of both coordinates, this is not always the case. Due to the variance reduction factor, increasing $R^2_{Y\sim W|Z,\boldsymbol{X}}$ may reduce the standard error more than enough to compensate for the increase in bias, resulting in tighter confidence

intervals. Denoting the solution to the optimization problem in expression (20) as $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$, the *most extreme possible* lower and upper limits after adjusting for W are given by

$$\text{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) = \hat{\lambda}_{\text{res}} - t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}), \quad \text{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2} = \hat{\lambda}_{\text{res}} + t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) \tag{21}$$

The interval composed of such limits,

$$\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) = \Big[\text{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda), \quad \text{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)\Big] \tag{22}$$

retrieves all inferences for λ which are compatible with an omitted variable with such strengths. In other words, $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$ is the union of all confidence intervals that could have been obtained by including an omitted variable with that strength or less in the regression equation. Moreover, if the true (sample) partial R^2 of W lies within the posited bounds, and the confidence interval adjusting for W has nominal coverage, it then immediately follows that $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$ is also a confidence interval with at least $1-\alpha$ coverage.[12]

3.3 Sensitivity statistics for routine reporting

Widespread adoption of sensitivity analysis benefits from simple and interpretable statistics that quickly convey the overall robustness of an estimate. To that end, Cinelli and Hazlett (2020) proposed two sensitivity statistics for routine reporting: (i) the partial R^2 of Z with Y, $R^2_{Y\sim Z|\boldsymbol{X}}$; and, (ii) the *robustness value* (RV). Here we generalize the notion of a partial R^2 as a measure of robustness to extreme scenarios, by introducing the *extreme robustness value* (XRV), for which the partial R^2 is a special case. We also recast these sensitivity statistics as a solution to an "inverse" question regarding the set of compatible inferences, $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$. That is, given a threshold for λ deemed to be of scientific importance (say, zero), what is the *minimum* strength of the sensitivity parameters $\boldsymbol{R}^2$ that could lead $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$ to include that threshold? This framework facilitates extending these metrics to other contexts, in particular to the IV setting, as we show in Section 4.2.3.

3.3.1 The extreme robustness value

One benefit of the partial R^2 parameterization is that the parameter $R^2_{Y\sim W|Z,\boldsymbol{X}}$ can be left completely unconstrained; i.e, in the optimization problem of expression 20, one can set the bound for $R^2_{Y\sim W|Z,\boldsymbol{X}}$ to its trivial bound of 1, and this still results in non-trivial bounds on the set of compatible inferences. This leads to our first inverse question: what is the *bare minimum* strength of association of the omitted variable W with Z that could bring its estimated coefficient to a region where it is no longer statistically different than zero (or another threshold of interest)?

[12] Note that here we are considering sample estimates, and thus this is different from the traditional analysis of confidence intervals for partially identified quantities, as in Imbens and Manski (2004) and Chernozhukov et al. (2022). That is, here we have an exact algebraic result that recovers the union of all confidence intervals that could have been obtained had we adjusted for an omitted variable W with (sample) R^2s bounded by the postulated strength. If the bias analysis is made in terms of population quantities (instead of sample quantities), valid (asymptotic) confidence intervals for the partially identified λ can be constructed as in Chernozhukov et al. (2022, Theorem 4).

To answer this question, we can see $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$ as a function of the bound $R^{2\max}_{Z\sim W|\boldsymbol{X}}$ alone, obtained from maximizing the adjusted critical value in expression 20 where: (i) the parameter $R^2_{Y\sim W|Z,\boldsymbol{X}}$ is left completely unconstrained (i.e, $R^2_{Y\sim W|Z,\boldsymbol{X}} \leq 1$); and, (ii) the parameter $R^2_{Z\sim W|\boldsymbol{X}}$ is bounded by XRV (i.e, $R^{2\max}_{Z\sim W|\boldsymbol{X}} \leq \text{XRV}$). The *Extreme Robustness Value* $\text{XRV}_{q^*,\alpha}(\lambda)$ is defined as the greatest lower bound XRV such that the null hypothesis that a change of $(100 \times q^*)\%$ of the original estimate, $H_0 : \lambda = (1 - q^*)\hat{\lambda}_{\text{res}}$, is not rejected at the α level,

$$\text{XRV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{XRV};\ (1-q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\lambda)\right\} \tag{23}$$

The solution to this problem gives,

$$\text{XRV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if } \ f_{q^*}(\lambda) \leq f^*_{\alpha,\text{df}-1} \\ \dfrac{f^2_{q^*}(\lambda) - f^{*2}_{\alpha,\text{df}-1}}{1 + f^2_{q^*}(\lambda)}, & \text{otherwise.} \end{cases} \tag{24}$$

Where $f_{q^*}(\lambda) := q^*|f_{Y\sim Z|\boldsymbol{X}}|$ (here $f_{Y\sim Z|\boldsymbol{X}}$ stands for the partial Cohen's f and we define the critical threshold $f^*_{\alpha,\text{df}\,-1} := t^*_{\alpha,\text{df}\,-1}/\sqrt{\text{df}\,-1}$).[13] Note $\text{XRV}_{q^*,\alpha}(\lambda)$ can be interpreted as an "adjusted partial R^2" of Z with Y. To see why, let us first consider the case of the minimal strength to bring the point estimate ($\alpha = 1$) to exactly zero ($q^* = 1$). We then have that $f^*_{\alpha=1,\text{df}\,-1} = 0$ and $f^2_{q^*=1}(\lambda) = f^2_{Y\sim Z|\boldsymbol{X}}$, resulting in

$$\text{XRV}_{q^*=1,\alpha=1}(\lambda) = \frac{f^2_{Y\sim Z|\boldsymbol{X}}}{1 + f^2_{Y\sim Z|\boldsymbol{X}}} = R^2_{Y\sim Z|\boldsymbol{X}} \tag{25}$$

This recovers the result of Cinelli and Hazlett (2020), and shows that, for an omitted variable W to bring down the estimated coefficient to zero, it needs to explain at least as much residual variation of Z as Z explains of Y. For the general case, we simply perform two adjustments that dampens the "raw" partial R^2 of Z with Y. First we adjust it by the proportion of reduction deemed to be problematic q^* through $f_{q^*} = q^*|f_{Y\sim Z|\boldsymbol{X}}|$; next, we subtract the threshold for which statistical significance is lost at the α level (via $f^{*2}_{\alpha,\text{df}-1}$).

The extreme robustness value establishes thus the equivalent of a "Cornfield condition" (Cornfield et al., 1959) for OLS estimates, meaning it gives the bare minimum strength of omitted variables necessary to overturn a certain conclusion—if W cannot explain at least $\text{XRV}_{q^*,\alpha}(\lambda)$ of the residual variation of Z, then such variable *is not* strong enough to bring about a change of $(100 \times q^*)\%$ on the original estimate, at the significance level of α, regardless of its association with Y.

3.3.2 The robustness value

Placing no constraints on the association of the omitted variable W with Y may be too conservative an exercise. An alternative measure of robustness of the OLS estimate is to consider the minimal

[13] Cohen's f^2 can be written as $f^2 = R^2/(1 - R^2)$.

strength of association that the omitted variable needs to have, *both* with Z and Y, so that a $1-\alpha$ confidence interval for λ will include a change of $(100 \times q^*)\%$ of the current restricted estimate.

Write $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$ as a function of both bounds varying simultaneously, that is, construct $\text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\lambda)$ by maximizing the adjusted critical value with bounds given by $R^2_{Y\sim W|Z,\boldsymbol{X}} \leq \text{RV}$ and $R^2_{Z\sim W|\boldsymbol{X}} \leq \text{RV}$. The *Robustness Value* $\text{RV}_{q^*,\alpha}(\lambda)$ for not rejecting the null hypothesis that $H_0 : \lambda = (1-q^*)\hat{\lambda}_{\text{res}}$, at the significance level α, is defined as

$$\text{RV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{RV};\ (1-q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\lambda)\right\} \tag{26}$$

We then have that,

$$\text{RV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if}\ \ f_{q^*}(\lambda) \leq f^*_{\alpha,\text{df}-1} \\ \frac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right), & \text{if}\ \ f^*_{\alpha,\text{df}-1} < f_{q^*}(\lambda) < f^{*-1}_{\alpha,\text{df}-1} \\ \text{XRV}_{q^*,\alpha}(\lambda), & \text{otherwise.} \end{cases} \tag{27}$$

Where $f_{q^*,\alpha}(\lambda) := q^*|f_{Y\sim Z|\boldsymbol{X}}| - f^*_{\alpha,\text{df}\,-1}$. In the appendix we show the conditions of Equation 27 are equivalent to those first derived in Cinelli and Hazlett (2020), with the advantage of being simpler to verify. The first case occurs when the confidence interval already includes $(1-q^*)\hat{\lambda}_{\text{res}}$ or the mere change of one degree of freedom achieves this. The second case occurs when both associations of W reach the bound. Finally, in the last case the solution is an interior point—this happens when the bound is large enough such that the constraint on the association with the outcome is not binding; in this case the RV reduces to the XRV.

The robustness value offers a simple interpretable measure that summarizes the strength of omitted variables necessary to change the estimate in problematic ways. If W explains $\text{RV}_{q^*,\alpha}(\lambda)$ of the residual variance of both Z and Y, then such variable is sufficiently strong to bring about a $(100 \times q)\%$ change in the estimate at the significance level of α, while any omitted variable that does not explain $\text{RV}_{q^*,\alpha}(\lambda)$ of the residual variance, neither of Z nor of Y, is not sufficiently strong to do so.

A visual depiction of the RV and XRV

Visually depicting the RV and the XRV in a sensitivity contour plot may be helpful. Consider Figure 1. The horizontal axis describes $R^2_{Z\sim W|\boldsymbol{X}}$ and the vertical axis describes $R^2_{Y\sim W|Z,\boldsymbol{X}}$. The contour lines show the adjusted t-value for testing the null hypothesis of zero effect for the reduced form regression (of Table 5), had we adjusted for W with such hypothetical strength (considering that adjustment reduces the t-value). The red dashed line shows a critical contour of interest, such as statistical significance at the $\alpha = 0.05$ level. The RV (when both values reach their bounds) summarizes the point of equal values on both axis of the critical contour, whereas the XRV summarizes the vertical line tangent to the critical contour, which will never be crossed.

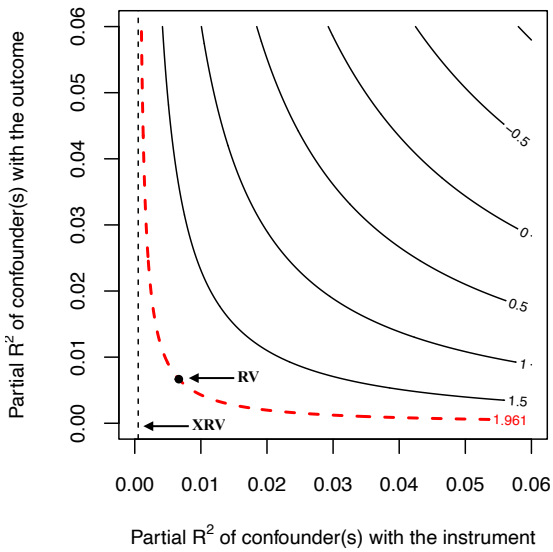


Figure 1: Sensitivity contours of the reduced form of Card (1993) depicting the RV and the XRV.

3.4 Bounding the strength of the omitted variable using observed covariates

One further result is required before turning to the sensitivity of IV estimates. Let X_j be a specific covariate of the set $\boldsymbol{X}$, and define

$$k_Z := \frac{R^2_{Z\sim W|\boldsymbol{X}_{-j}}}{R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{R^2_{Y\sim X_j|Z\boldsymbol{X}_{-j}}}. \tag{28}$$

where $\boldsymbol{X}_{-j}$ represents the vector of covariates $\boldsymbol{X}$ excluding X_j. These new parameters, k_Z and k_Y, stand for how much "stronger" W is relatively to the observed covariate X_j in terms of residual variation explained of Z and Y. Our goal in this section is to re-express (or bound) the sensitivity parameters $R^2_{Z\sim W|\boldsymbol{X}}$ and $R^2_{Y\sim W|Z,\boldsymbol{X}}$ in terms of the relative strength parameters k_Z and k_Y.

We start by restating the bounds derived in Cinelli and Hazlett (2020, Sec. 4.4). These are particularly useful when contemplating X_j and W both *confounders* of Z (violations of the ignorability of the instrument). Let $R^2_{W\sim X_j|\boldsymbol{X}_{-j}} = 0$ (or, equivalently, consider the part of W not linearly explained by $\boldsymbol{X}$). Then the previous sensitivity parameters can be written as

$$R^2_{Z\sim W|\boldsymbol{X}} = k_Z f^2_{Z\sim X_j|\boldsymbol{X}_{-j}}, \qquad R^2_{Y\sim W|Z,\boldsymbol{X}} \leq \eta^2 f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} \tag{29}$$

where η is a function of both parameters k_Y, k_Z and $R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}$.

In the instrumental variable setting, however, W and X_j may be *side-effects* of Z, instead of causes of Z (violations of the exclusion restriction). In such cases, reasoning about the orthogonality of $\boldsymbol{X}$ and W may not be natural, as the instrument itself is a source of dependence between these variables. Therefore, here we additionally provide bounds under the alternative condition

$R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0$. We then have that

$$R^2_{Z\sim W|\boldsymbol{X}} \leq \eta' f^2_{Z\sim X_j|\boldsymbol{X}_{-j}}, \qquad R^2_{Y\sim W|Z,\boldsymbol{X}} = k_Y f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} \tag{30}$$

where η' is a function of k_Z and $R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}$ (see Appendix D for details).

These results allow investigators to leverage knowledge of *relative importance* of variables (Kruskal and Majors, 1989) when making plausibility judgments regarding sensitivity parameters. For instance, if researchers have domain knowledge to argue that a certain observed covariate X_j is supposed to be a strong determinant of the instrument and the outcome variation, and that the omitted variable W is not likely to explain as much residual variance of Z and Y as that observed covariate, such results can be used to set plausible bounds on the maximum bias due to the omission of W.

4 An omitted variable bias framework for the sensitivity of IV

We are now ready to develop a suite of sensitivity analysis tools for instrumental variable regression. In this section, we first show how separate sensitivity analysis of the reduced form and first stage is sufficient to draw many valuable conclusions regarding the sensitivity of IV. We then construct a complete OVB framework for sensitivity analysis of IV within the Anderson-Rubin approach, allowing one to investigate the sensitivity of tests to any specific hypothesis, the sensitivity of lower and upper limits of confidence intervals, to define and compute sensitivity statistics for routine reporting for IV, such as (extreme) robustness values, as well as providing bounds on the sensitivity parameters, on the basis of comparison to observed covariates.

4.1 Sensitivity analysis of the reduced form and first stage

The recent literature on instrumental variables places strong emphasis on the first-stage and the reduced-form estimates. Not only are the first stage and reduced form often substantively meaningful on their own, but their critical examination plays an important role for motivating the causal story behind a particular instrumental variable. For example, in the "local average treatment effect" interpretation of the IV estimand, *both* the first stage and the reduced form must be unconfounded so that the resulting estimate can be interpreted as the average causal effect among compliers (Angrist et al., 1996). Therefore, beyond a means to the final IV estimate, researchers are advised to report and to interpret the first stage and the reduced form by, for example, assessing whether those results are in accordance to the postulated mechanisms that justify the choice of instrument (Angrist and Krueger, 2001; Angrist and Pischke, 2009; Imbens, 2014; Angrist and Pischke, 2014; Imbens and Rubin, 2015a). While investigating these separate regressions, researchers can deploy all sensitivity analysis results discussed in the previous section.

Fortunately, such sensitivity analyses also provide answers to many pivotal sensitivity questions regarding the IV estimate itself. In particular, if the investigator is interested in assessing the strength of confounders or side-effects needed to bring the IV point estimate to zero, or to not reject

the null hypothesis of zero effect, the results of the sensitivity analysis of the reduced form is all that is needed.[14] If interest lies in also determining whether the IV estimate could be arbitrarily large in either direction, then the sensitivity of the first stage must also be assessed, as omitted variables capable of changing the direction of the first stage can lead to unbounded IV estimates. We now give a more precise meaning to these claims.

4.1.1 What the reduced form and first stage reveal about the IV point estimate

Recall that all IV estimators under consideration are algebraically equivalent, equal to the ratio of the reduced-form and the first-stage regression coefficients,

$$\hat{\tau} := \hat{\tau}_{\text{ILS}} = \hat{\tau}_{\text{2SLS}} = \hat{\tau}_{\text{AR}} = \frac{\hat{\lambda}}{\hat{\theta}} \tag{31}$$

This simple algebraic fact allows us to draw two important conclusions regarding the sensitivity of $\hat{\tau}$ from the sensitivity of $\hat{\lambda}$ and $\hat{\theta}$ alone.

First, residual biases can bring the IV point estimate to zero *if, and only if,* they can bring the reduced-form point estimate to zero. Therefore, if sensitivity analysis of the reduced form reveals that omitted variables are not strong enough to explain away $\hat{\lambda}$, then they also cannot explain away the IV point estimate $\hat{\tau}$. Or, more worrisome, if analysis reveals that it takes weak confounding or side-effects to explain away $\hat{\lambda}$, the same holds for the IV estimate $\hat{\tau}$. In sum, for all IV estimators considered here, to assess the strength of biases needed to bring the IV point estimate to zero, one needs only to perform a sensitivity analysis on the reduced-form regression coefficient.

Second, if we cannot rule out confounders or side-effects that are sufficiently strong to *change the sign* of the first-stage point estimate $\hat{\theta}$, then we also cannot rule out that the IV point estimate $\hat{\tau}$ could be *arbitrarily large* in either direction, even if not exactly equal to zero. This can be immediately seen by letting $\hat{\theta}$ approach zero on either side of the limit. Thus, whenever we are interested in biases as large *or larger* than a certain amount, the robustness of the first stage to the zero null puts an upper bound on the robustness of the IV point estimate.

4.1.2 What the reduced form and first stage reveal about IV hypothesis tests

Contrary to the point estimate, the different approaches presented here may lead to different conclusions regarding how omitted variables would have changed inferences. Let us start by examining the Anderson-Rubin/Fieller approach, as it has nominal coverage regardless of instrument strength, and its conclusions match the intuition of current guidelines when assessing the first-stage and reduced-form estimates (Angrist and Krueger, 2001; Angrist and Pischke, 2009, 2014).

[14] The value of null hypothesis significance testing has been the subject of considerable debate (see eg. Ziliak and McCloskey, 2008; Cinelli, 2012; Benjamin et al., 2018; Amrhein and Greenland, 2018). We are not advocating for researchers to focus on the null $H_0 : \tau = 0$. Rather, we show that whenever researchers are interested in such a null—as remains very common in practice—then the sensitivity of the reduced form is all that is needed. Further, while this special case is a convenient one, our methods are not restricted to it. Section 4.2 develops a complete suite of sensitivity analysis tools of IV, allowing the construction of confidence intervals or tests of any null.

Consider again the IV estimand

$$\tau = \frac{\lambda}{\theta}$$

Note that the same arguments we used before for the estimator hold for the estimand. Logically, provided the ratio is well defined ($\theta \neq 0$), we have that $\tau = 0 \iff \lambda = 0$. Therefore, a test of the null hypothesis $H_0 : \lambda = 0$ in the reduced-form regression is logically equivalent to a test of the null hypothesis $H_0 : \tau = 0$ for the IV estimand. Similarly, for a fixed λ, if we cannot rule out that θ is arbitrarily close to zero in either direction, then, logically, we also cannot rule out that τ is arbitrarily large in either direction—a test for the null hypothesis $H_0 : \theta = 0$ is thus logically equivalent to testing whether arbitrarily large sizes for τ can be ruled out.

The Anderson-Rubin/Fieller approach is coherent with respect to these logical implications. Recall the Anderson-Rubin test for the null hypothesis $H_0 : \tau = \tau_0$ is based on the test of $H_0 : \phi_{\tau_0} = 0$. By the FWL theorem, the point estimate and (estimated) standard error for $\hat{\phi}_{\tau_0}$ can be expressed in terms of the first-stage and reduced-form estimates (see Appendix A)

$$\hat{\phi}_{\tau_0} = \hat{\lambda} - \tau_0\hat{\theta}, \qquad \widehat{\text{se}}(\hat{\phi}_{\tau_0}) = \sqrt{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2\widehat{\text{var}}(\hat{\theta}) - 2\tau_0\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})} \tag{32}$$

Testing $H_0 : \phi_{\tau_0} = 0$ requires comparing the t-value for $\hat{\phi}_{\tau_0}$ with a critical threshold $t^*_{\alpha,\text{df}-1}$, and the null hypothesis is not rejected if $|t_{\hat{\phi}_{\tau_0}}| \leq t^*_{\alpha,\text{df}-1}$. Squaring and rearranging terms we obtain the quadratic inequality which must hold for non-rejection:

$$\underbrace{\left(\hat{\theta}^2 - \widehat{\text{var}}(\hat{\theta}) \times t^{*2}_{\alpha,\text{df}-1}\right)}_{a}\tau_0^2 + \underbrace{2\left(\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) \times t^{*2}_{\alpha,\text{df}-1} - \hat{\lambda}\hat{\theta}\right)}_{b}\tau_0 + \underbrace{\left(\hat{\lambda}^2 - \widehat{\text{var}}(\hat{\lambda}) \times t^{*2}_{\alpha,\text{df}-1}\right)}_{c} \leq 0 \tag{33}$$

When considering the null hypothesis $H_0 : \tau_0 = 0$, only the term c remains, and c is less or equal to zero if and only if one cannot reject the null hypothesis $H_0 : \lambda = 0$ in the reduced-form regression. The Anderson-Rubin approach thus comports with the recommendation of Angrist and Krueger (2001) that "if you can't see the causal relation of interest in the reduced form, it's probably not there." Also note that arbitrarily large values for τ_0 will satisfy the inequality in Equation 33 if, and only if, $a < 0$, meaning that we cannot reject the null hypothesis $H_0 : \theta = 0$ in the first-stage regression. This supports the recommendation that, if one is unsure about the direction of the first stage, it is likely that very little can be said about the magnitude of the IV estimate.

Within the Anderson-Rubin framework, we thus reach analogous conclusions regarding hypothesis testing as those regarding the point estimate: (i) when interest lies in the zero null hypothesis, the sensitivity of the reduced form is exactly the sensitivity of the IV—no other analyses are needed. Confounders or side-effects sufficiently strong to bring the reduced form to a region where it is not statistically different than zero can also bring the IV estimate to a region where it is not statistically different than zero, and only omitted variables with such strength are capable of doing so; and, (ii) if one is interested in biases of a certain amount, or larger, then the sensitivity of the first stage to the zero null hypothesis needs also to be assessed. Specifically, for any null hypothesis of interest

$H_0 : \tau = \tau_0$, omitted variables that are strong enough to make the first stage not statistically different from zero may also lead us to not reject values arbitrarily "worse" than τ_0.

As is well known, it is not uncommon for frequentist statistical tests to lead to logically incoherent decisions (Gabriel, 1969; Schervish, 1996; Patriota, 2013; Fossaluza et al., 2017). While inferences made in the Anderson-Rubin approach have the expected behavior in this setting, inferences using ILS or 2SLS may not. Cases can be found for ILS and 2SLS where, for instance, one fails to reject the null hypothesis $H_0 : \lambda = 0$, yet still rejects the null hypothesis $H_0 : \tau = 0$ (and vice-versa). Such claims do not conform to current guidelines for interpreting the first-stage and reduced-form regressions (Angrist and Pischke, 2009).

4.2 Sensitivity analysis of the IV in the Anderson-Rubin approach

We now develop a complete set of sensitivity analysis tools for IV. We focus on the Anderson-Rubin approach for this task because: (i) it allows performing sensitivity analysis of the IV with only two interpretable sensitivity parameters; (ii) it has correct test size regardless of "instrument strength"; and, (iii) its conclusions conform to current recommendations regarding the interpretation of the first-stage and reduced-form regressions.

4.2.1 Sensitivity for testing a specific null hypothesis

We begin by examining the sensitivity of the t-value for testing a specific null hypothesis $H_0 : \tau = \tau_0$, as this is a straightforward application of the tools of Section 3. Recall that, in the Anderson-Rubin approach, a test for the null hypothesis $H_0 : \tau = \tau_0$ is a test for the null hypothesis $H_0 : \phi_{\tau_0} = 0$ in the regression of Y_{τ_0} on the instrument Z and covariates $\boldsymbol{X}$ and W. Therefore, standard OLS sensitivity analysis for testing the null hypothesis $H_0 : \phi_{\tau_0} = 0$ on the Anderson-Rubin regression gives the desired results for $H_0 : \tau = \tau_0$.

In detail, a sensitivity analysis for the null hypothesis that the IV estimate τ equals some τ_0 can be performed as follows:

1. Construct $Y_{\tau_0} = Y - \tau_0 D$ under the null value $H_0 : \tau = \tau_0$;
2. Run the OLS model $Y_{\tau_0} = \hat{\phi}_{\text{res},\tau_0} Z + \boldsymbol{X}\hat{\beta}_{\text{res},\tau_0} + \hat{\varepsilon}_{\tau_0,\text{res}}$;
3. Perform regular OLS sensitivity analysis for the null $H_0 : \phi_{\tau_0} = 0$.

This procedure can both tell us how omitted variables no worse than $\mathbf{R}^2 = \{R^2_{Z\sim W|\boldsymbol{X}}, R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}\}$ would alter inferences regarding the null $H_0 : \tau = \tau_0$, or what is the minimal strength of $\mathbf{R}^2$ that is required to not reject the null $H_0 : \tau = \tau_0$, as given by the RV or XRV.

Making sense of the sensitivity parameters. While separate analyses of the first stage and reduced form regressions may suggest the need of three sensitivity parameters for the sensitivity of IV (e.g, $R^2_{Z\sim W|\boldsymbol{X}}$, $R^2_{D\sim W|Z,\boldsymbol{X}}$ and $R^2_{Y\sim W|Z,\boldsymbol{X}}$), note how within the Anderson-Rubin approach one is able to perform sensitivity with only two parameters ($R^2_{Z\sim W|\boldsymbol{X}}, R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}$). The meaning of the

parameter related with the instrument ($R^2_{Z\sim W|\boldsymbol{X}}$) is unchanged and straightforward, ie., the share of residual variation of the instrument explained by the omitted variable W. The main difference concerns the parameter $R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}$, which stands for the share of residual variance of Y_{τ_0} explained by W. The substantive interpretation of Y_{τ_0} depends on the causal assumptions the researcher is willing to defend. For instance, under $H_0 : \tau = \tau_0$ and a constant treatment effects model, we have that $Y_{\tau_0} = Y - \tau_0 D$ equals the *untreated potential outcome* Y_0 and thus $R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}$ could be interpreted as the share of residual variance of Y_0 explained by W. For simplicity of exposition, we adopt this interpretation throughout the text.

4.2.2 Compatible inferences given bounds on partial R^2

Instead of assessing the sensitivity of the test statistic for specific a null hypothesis, investigators may be interested in recovering the whole set of inferences compatible with plausibility judgments on the maximum strength of W. As discussed in Section 2, for a critical threshold $t^*_{\alpha,\text{df}-1}$, the confidence interval for τ in the Anderson-Rubin framework is given by

$$\text{CI}_{1-\alpha}(\tau) = \{\tau_0;\ t^2_{\phi_{\tau_0}} \leq t^{*2}_{\alpha,\text{df}-1}\} \tag{34}$$

Now consider bounds on sensitivity parameters $R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y_0\sim W|Z,\boldsymbol{X}}$ (which should be judged to hold *regardless* of the value of τ_0) and $R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}}$. Let $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ denote the maximum bias-adjusted critical value under the posited bounds on the strength of W. The set of compatible inferences for τ, $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ is then simply given by

$$\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) = \left\{\tau_0;\ t^2_{\hat{\phi}_{\text{res},\tau_0}} \leq \left(t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right\} \tag{35}$$

This interval can be found analytically using the same inequality as in Equation 33, now with the parameters of the restricted regression actually run, and the traditional critical value replaced by the bias-adjusted critical value $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$

$$\underbrace{\left(\hat{\theta}^2_{\text{res}} - \widehat{\text{var}}(\hat{\theta}_{\text{res}}) \times \left(t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right)}_{a}\tau_0^2 + \underbrace{2\left(\widehat{\text{cov}}(\hat{\lambda}_{\text{res}},\hat{\theta}_{\text{res}}) \times \left(t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2 - \hat{\lambda}_{\text{res}}\hat{\theta}_{\text{res}}\right)}_{b}\tau_0$$
$$+ \underbrace{\left(\hat{\lambda}^2_{\text{res}} - \widehat{\text{var}}(\hat{\lambda}_{\text{res}}) \times \left(t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right)}_{c} \leq 0 \tag{36}$$

Note that users can easily obtain $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ with any software that computes Anderson-Rubin or Fieller's confidence intervals by simply providing the modified critical threshold $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$.

It is now useful to discuss the possible shapes of $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}$ as this will help understanding the robustness values for IV we derive next. Let $\mathbf{r} = \{r_{\min}, r_{\max}\}$ denote the roots of the quadratic equation, which can be written as $\mathbf{r} = -b \pm \sqrt{\Delta}/2a$, with $\Delta = b^2 - 4ac$. If $a > 0$ (i.e, we have a statistically

significant first stage), the quadratic equation will be convex, and thus only the values between the roots will be non-positive. This leads to the connected confidence interval $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2} = [r_{\min}, r_{\max}]$. When $a < 0$ (i.e, the null hypothesis of zero for the first stage is not rejected), the curve is concave and this leads to unbounded confidence intervals. Here we have two sub-cases: (i) when $\Delta < 0$, the quadratic curve never touches zero, and thus the confidence interval is simply the whole real line $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2} = (-\infty, +\infty)$; and, (ii) when $\Delta > 0$ the confidence interval will be union of two disjoint intervals $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2} = (-\infty, r_{\min}] \cup [r_{\max}, +\infty)$.[15]

4.2.3 Sensitivity statistics for routine reporting

Armed with the notion of a set of compatible inferences for IV, $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$, we are now able to formally define and derive (extreme) robustness values for instrumental variable estimates.

Extreme robustness values for IV. The extreme robustness value $\text{XRV}_{q^*,\alpha}(\tau)$ for the IV estimate is defined as the minimum strength of association of omitted variables with the instrument so that we cannot reject a reduction of $(100 \times q^*)\%$ of the original IV estimate; that is,

$$\text{XRV}_{q^*,\alpha}(\tau) := \inf \left\{ \text{XRV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\tau) \right\} \tag{37}$$

It then follows immediately from Equation 35 that

$$\text{XRV}_{q^*,\alpha}(\tau) = \text{XRV}_{1,\alpha}(\phi_{\tau^*}) \tag{38}$$

where $\tau^* = (1-q^*)\hat{\tau}_{\text{res}}$. As in the general case, the extreme robustness value can be interpreted as a "dampened" partial R^2 of the instrument Z with the "putative" untreated potential outcome Y_{τ_0}. Also of interest is the special case of the minimum strength to bring the IV estimate to a region where it is no longer statistically different than zero ($q^* = 1$), in which we obtain $\text{XRV}_{1,\alpha}(\tau) = \text{XRV}_{1,\alpha}(\lambda)$. That is, for the null hypothesis of $H_0 : \tau = 0$, the extreme robustness value of the IV estimate equals the extreme robustness value of the reduced-form estimate, as we discussed in the last section.

The $\text{XRV}_{q^*,\alpha}(\tau)$ computes the minimal strength of W required to not reject a particular null hypothesis of interest. We might be interested, instead, in asking about the minimal strength of omitted variables to not reject a specific value *or worse*. When confidence intervals are connected, such as the case of standard OLS, the two notions coincide. But in the Anderson-Rubin case, as we have seen, confidence intervals for the IV estimate can sometimes consist of disjoint intervals. Therefore, let the upper and lower limits of $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ be $\text{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ and $\text{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ respectively. The extreme robustness value $\text{XRV}_{\geq q^*,\alpha}(\tau)$ for the IV estimate is defined as the minimum strength of association that confounders or side-effects need to have with the instrument so that we

[15] See Mehlum (2020) for an intuitive graphical characterization of Fieller's solutions using polar coordinates.

cannot reject a change of $(100 \times q^*)\%$ *or worse* of the original IV estimate;

$$\text{XRV}_{\geq q^*,\alpha}(\tau) := \inf\left\{\text{XRV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \left[\text{LL}^{\max}_{1-\alpha,1,\text{XRV}}(\tau),\quad \text{UL}^{\max}_{1-\alpha,1,\text{XRV}}(\tau)\right]\right\} \tag{39}$$

Now note that, whenever $\text{CI}^{\max}_{1-\alpha,\text{df}-1}(\tau)$ is connected, we must have that $\text{XRV}_{\geq q^*,\alpha}(\tau) = \text{XRV}_{q^*,\alpha}(\tau)$. On the other hand, recall that $\text{CI}^{\max}_{1-\alpha,\text{df}-1}(\tau)$ will be disjoint only if $t^2_{\hat{\theta}_{\text{res}}} \leq (t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2})^2$, which is precisely the condition for the extreme robustness value of the first stage. Therefore,

$$\text{XRV}_{\geq q^*,\alpha}(\tau) = \min\{\text{XRV}_{1,\alpha}(\phi_{\tau^*}),\quad \text{XRV}_{1,\alpha}(\theta)\} \tag{40}$$

This corroborates our previous conclusion that, when we are interested in biases as large or larger than a certain amount, the robustness of the IV estimate is bounded by the robustness of the first stage assessed at the zero null.

Robustness values for IV. The definitions of the robustness value for IV follow the same logic discussed above, but now considering both bounds on $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}$ varying simultaneously. That is,

$$\text{RV}_{q^*,\alpha}(\tau) := \inf\left\{\text{RV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\tau)\right\} \tag{41}$$

Again from Equation 35 we have that

$$\text{RV}_{q^*,\alpha}(\tau) = \text{RV}_{1,\alpha}(\phi_{\tau^*}) \tag{42}$$

Which for the special case of $q^* = 1$ simplifies to $\text{RV}_{1,\alpha}(\tau) = \text{RV}_{1,\alpha}(\lambda)$, as before. We can also define robustness values for not rejecting the null hypothesis of a reduction of $(100 \times q^*)\%$ *or worse*

$$\text{RV}_{\geq q^*,\alpha}(\tau) := \inf\left\{\text{RV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \left[\text{LL}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\tau),\quad \text{UL}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\tau)\right]\right\} \tag{43}$$

By the same arguments articulated above, $\text{RV}_{\geq q^*,\alpha}(\tau)$ must be the minimum of the robustness value of the Anderson-Rubin regression evaluated at $\tau^* = (1-q^*)\hat{\tau}_{\text{res}}$ and the robustness value of the first-stage regression evaluted at the zero null

$$\text{RV}_{\geq q^*,\alpha}(\tau) = \min\{\text{RV}_{1,\alpha}(\phi_{\tau^*}),\quad \text{RV}_{1,\alpha}(\theta)\} \tag{44}$$

For the special case of $q^* = 1$ (zero null hypothesis), $\text{RV}_{\geq q^*,\alpha}(\tau)$ simplifies to the minimum of the robustness value of the first stage and of the reduced form, $\text{RV}_{\geq q^*=1,\alpha}(\tau) = \min\{\text{RV}_{1,\alpha}(\lambda),\quad \text{RV}_{1,\alpha}(\theta)\}$.

4.2.4 Bounds on the strength of omitted variables

The bounds discussed in Section 3.4 work without any major modifications in the Anderson-Rubin setting. When testing a specific null hypothesis $H_0 : \tau = \tau_0$ in the AR regression, we have k_Z as

before, and instead of k_Y we now have $k_{Y_{\tau_0}}$

$$k_{Y_{\tau_0}} := \frac{R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}_{-j}}}{R^2_{Y_{\tau_0}\sim X_j|Z\boldsymbol{X}_{-j}}}. \tag{45}$$

The plausibility judgment one is making here is that of how strong unobserved confounders or side-effects are, relative to observed covariates, in explaining the residual variance of the untreated potential outcome and of the instrument, under the null hypothesis $H_0 : \tau = \tau_0$.

Since the judgment is made under a specific null, the bounds will be different when testing different hypotheses. Therefore, it may be useful to compute bounds under a slightly more *conservative* assumption. More precisely, consider

$$k^{\max}_{Y_{\tau_0}} := \frac{\max_{\tau_0} R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}_{-j}}}{\max_{\tau_0} R^2_{Y_{\tau_0}\sim X_j|Z\boldsymbol{X}_{-j}}}. \tag{46}$$

That is, we can posit that the omitted variables are no stronger than (a multiple of) the *maximum* explanatory power of an observed covariate, regardless of the value of τ_0. This has the useful property of providing a unique valid bound for any value of the null hypothesis, and can be used to place bounds on sensitivity contours of the lower and upper limit of the AR confidence intervals, as we show next.

5 Using the OVB framework for the sensitivity analysis of IV

In this section we return to our running example of estimating the returns to schooling using proximity to college as an instrumental variable, and show how these tools can be deployed to assess the robustness of those findings to violations of the IV assumptions. We begin the sensitivity analysis by examining the robustness of the first-stage and reduced-form estimates. Not only are these analyses usually important on their own right, but in many cases—including this one—this exercise will be sufficient to establish that the instrumental variable estimate is not very informative regarding the causal effect of interest. We then turn to the sensitivity of the IV estimate itself, and further show how sensitivity contour plots of the adjusted lower and upper limits of the AR confidence interval, supplemented with benchmark bounds, give a complete, yet succinct picture of the whole range of sensitivity of the IV analysis. Throughout, we focus the discussion on violations of the ignorability of the instrument due to confounders, as this is the main threat of the study under investigation. Readers should keep in mind, however, that mathematically all analyses performed here can be equally interpreted as assessing violations of the exclusion restriction (or both).

5.1 Minimal reporting and sensitivity contours of the reduced form

Table 2 shows our proposal for a minimal sensitivity reporting of the reduced-form estimate (here, the effect of *Proximity* on *Earnings*). Beyond the usual statistics such as the point estimate, standard-

Outcome: *Earnings* (log)

Instrument	Estimate	Std. Error	t-value	$R^2_{Y\sim Z\mid \boldsymbol{X}}$	$\text{XRV}_{q^*,\alpha}$	$\text{RV}_{q^*,\alpha}$
Proximity	0.042	0.018	2.33	0.18%	0.05%	0.67%

Bound (1x SMSA): $R^2_{Y\sim W\mid Z,\boldsymbol{X}} = 2\%$, $R^2_{W\sim Z\mid \boldsymbol{X}} = 0.6\%$, $t^{\dagger\,\max}_{\alpha,\text{df}\,-1,\boldsymbol{R}^2} = 2.55$

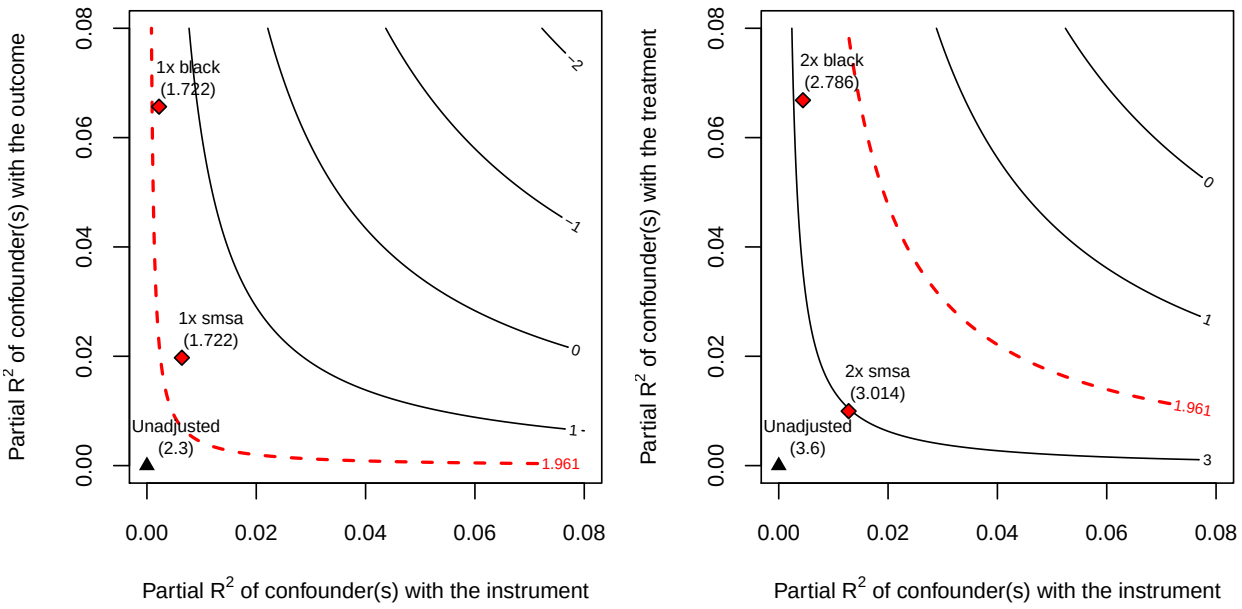
Note: df $= 2994$, $q^* = 1$, $\alpha = 0.05$

Table 2: Minimal sensitivity reporting of the reduced-form regression.

error and t-value, we recommend that researchers also report the: (i) partial R^2 of the instrument with the outcome ($R^2_{Y\sim Z\mid \boldsymbol{X}} = 0.18\%$), as well as (ii) the robustness value ($\text{RV}_{q^*,\alpha} = 0.67\%$), and (iii) the extreme robustness value ($\text{XRV}_{q^*,\alpha} = 0.05\%$), both for where the confidence interval would cross zero ($q^* = 1$), at a chosen significance level (here, $\alpha = 0.05$).

In our running example, the robustness value reveals that confounders that explain 0.67% of the residual variation both of *proximity* and of (log) *Earnings* are sufficiently strong to make the reduced-form estimate statistically insignificant, whereas confounders that explain less than 0.67% of the residual variation of both the instrument and of the outcome are not strong enough to do so. The extreme robustness value and the partial R^2 show that, if we are not willing to impose constraints on the strength of confounders with the outcome, then they would need to explain less than 0.05% or 0.18% of the instrument to escape concerns of eliminating statistical significance or fully eliminating the point estimate, respectively. To aid users in making plausibility judgments, the note of Table 2 provides the maximum strength of unobserved confounding if it were as strong as *SMSA* (an indicator variable for whether the individual lived in a metropolitan region) along with the bias-adjusted critical value for a confounder with such strength, $t^{\dagger\,\max}_{\alpha,\text{df}\,-1,\boldsymbol{R}^2} = 2.55$. Since the observed t-value (2.33) is less than the adjusted critical threshold of 2.55, the table immediately reveals that confounding as strong as *SMSA* (for example, in the form of residual geographic confounding) is sufficiently strong to be problematic.

Beyond the results of Table 2, researchers can also explore sensitivity contour plots of the t-value for testing the null hypothesis of zero effect, while showing different bounds on strength of confounding, under different assumptions of how they compare to the observed variables. This is shown in Figure 2a. The horizontal axis describes the partial R^2 of the confounder with the instrument whereas the vertical axis describes the partial R^2 of the confounder with the outcome. The contour lines show the t-value one would have obtained, had a confounder with such postulated strength been included in the reduced-form regression. The red dashed line shows the statistical significance threshold, and the red diamonds places bounds on strength of confounding as strong as *Black* (an indicator for race) and, again, *SMSA*. As we can see, confounders as strong as either *Black* or *SMSA* are sufficient to bring the reduced form, and hence also the IV estimate, to a region which is not statistically different from zero. Since it is not very difficult to imagine residual confounders as strong or stronger than those (e.g., parental income, finer grained geographic location, etc), these results for the reduced form already call into question the reliability of the instrumental variable estimate.



(a) Sensitivity contours of the reduced form.

(b) Sensitivity contours of the first stage.

Figure 2: Sensitivity contour plots with benchmark bounds for the t-value of: (a) the reduced form; and, (b) the first stage.

5.2 Minimal reporting and sensitivity contours of the first stage

Treatment: *Education* (years)

Instrument	Estimate	Std. Error	t-value	$R^2_{D\sim Z\|\boldsymbol{X}}$	$\text{XRV}_{q^*,\alpha}$	$\text{RV}_{q^*,\alpha}$
Proximity	0.32	0.088	3.64	0.44%	0.31%	3.02%
Bound (1x SMSA): $R^2_{D\sim W\|Z,\boldsymbol{X}} = 0.5\%$, $R^2_{Z\sim W\|\boldsymbol{X}} = 0.6\%$, $t^{\dagger\,\max}_{\alpha,\text{df}\,-1,\boldsymbol{R}^2} = 2.26$						
Note: df $= 2994$, $q^* = 1$, $\alpha = 0.05$						

Table 3: Minimal sensitivity reporting of the first-stage regression.

Table 3 performs the same sensitivity exercises as before, but now for the regression of *Education* (treatment) on *Proximity* (instrument). As expected, the association of proximity to college with years of education is stronger than its association with earnings, and this is also reflected in the robustness statistics, which are slightly higher ($R^2_{D\sim Z|\boldsymbol{X}} = 0.44\%$, $\text{XRV}_{q^*,\alpha} = 0.31\%$ and $\text{RV}_{q^*,\alpha} = 3.02\%$). As the note of Table 3 shows, confounding as strong as *SMSA* would not be sufficiently strong to bring the first-stage estimate to a region where it is not statistically different than zero. Figure 2b supplements those analysis with the sensitivity contour plot for the t-value of the first-stage regression. Here the horizontal axis still describes the partial R^2 of the confounder with the instrument, but now the vertical axis describes the partial R^2 of the confounder with the treatment. The plot reveals that, contrary to the reduced form, the first stage survives confounding once or twice as strong as *Black* or *SMSA*. The contrast of both sensitivity results suggests that, in our running

example, the most evident risk to the validity of the IV estimate comes from residual confounding on the reduced-form estimate.

5.3 Minimal reporting and sensitivity contours of the IV

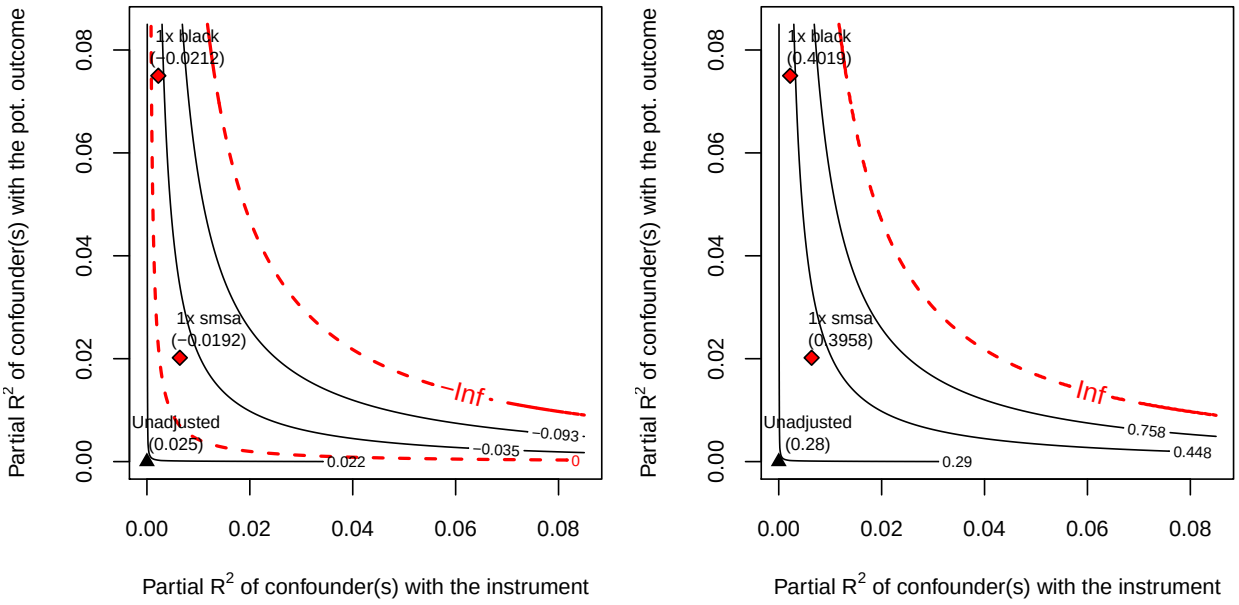
Outcome: *Earnings* (log)

Treatment	Estimate	$\text{LL}_{1-\alpha}$	$\text{UL}_{1-\alpha}$	t-value	$\text{XRV}_{\geq q^*,\alpha}$	$\text{RV}_{\geq q^*,\alpha}$
Education (years)	0.132	0.025	0.285	2.33	0.05%	0.67%
Bound (1x SMSA): $R^2_{Y_0 \sim W \mid Z, \boldsymbol{X}} = 2\%$, $R^2_{W \sim Z \mid \boldsymbol{X}} = 0.6\%$, $t^{\dagger\,\max}_{\alpha, \text{df}-1, \boldsymbol{R}^2} = 2.55$						
Note: df $= 2994$, $q^* = 1$, $\alpha = 0.05$						

Table 4: Minimal sensitivity reporting of IV estimate (Anderson-Rubin).

Finally, we turn our attention to the sensitivity analysis of the IV, and Table 4 shows our proposed minimal sensitivity reporting. We start with the IV point estimate (0.132), as well as the lower limit ($\text{LL}_{1-\alpha} = 0.025$) and the upper limit ($\text{UL}_{1-\alpha} = 0.285$) of the Anderson-Rubin confidence interval. The t-value for testing the null hypothesis of zero effect is also shown (2.33). Next, we propose researchers report the extreme robustness value $\text{XRV}_{\geq q^*,\alpha}$ and the robustness value $\text{RV}_{\geq q^*,\alpha}$ for bringing the lower limit of the confidence interval to or beyond zero (or another meaningful threshold), at the 5% significance level. As derived in Section 4.2.3, the (extreme) robustness value of the IV estimate for bringing the lower limit of the confidence interval to zero or below is the minimum of either the (extreme) robustness value of the reduced form and the (extreme) robustness value of the first stage. Therefore, the sensitivity statistics of Table 4 essentially reproduce the results of Table 2.

After examining the sensitivity of the first stage and reduced form it is thus, more informative to assess the sensitivity of the IV against values *other than zero*. To that end, investigators may wish to examine sensitivity contour plots similar to those of Figure 2, but with contours now showing the adjusted lower and upper limits of the confidence interval. These contours are shown in Figure 3, with the horizontal axis indicating the partial R^2 of the confounder with the instrument, and the vertical axis now indicating the partial R^2 of the confounder with the untreated *potential* outcome. The contour lines show the worst lower (or upper) limit of the set of compatible inferences considering confounders bounded by such strength. Red dashed lines shows a critical contour line of interest (such as zero) as well as the boundary beyond confidence intervals become unbounded. As the plot reveals, even confounding as strong as *SMSA* could lead to an interval of compatible inferences for the causal effect of $\text{CI}^{\max}_{1-\alpha, \boldsymbol{R}^2}(\tau) = [-0.02, 0.40]$, which includes not only the original OLS estimate (7.5%), but also implausibly high values (40%), or even negative values (-2%), and is thus too wide for any meaningful conclusions regarding the "true" returns to schooling. That is, if we are concerned that omitted variables as strong as *SMSA* might exist, then we are unable to reject any estimates in this range, calling into question the strength of evidence provided by this particular IV study.



(a) Sensitivity contours for the lower limit.

(b) Sensitivity contours for the upper limit.

Figure 3: Sensitivity contour plots for the lower (a) and upper (b) limits of the 95% confidence interval for the IV estimate.

6 Conclusion

In this paper we developed a suite of sensitivity analysis tools for IV that naturally handles multiple "side-effects" and confounders of the instrument, does not require assumptions on the functional form of such omitted variables, and allows exploiting expert knowledge to bound sensitivity parameters. In particular, we introduced new sensitivity statistics for IV estimates that are suited for routine reporting, such as (extreme) robustness values, describing the minimum strength that omitted variables need to have, both with the instrument, and with the untreated potential outcome, to overturn the conclusions of an IV study. We also introduced a novel "bias-adjusted" critical value that allows researchers to easily perform hypothesis tests or construct confidence intervals that accounts for omitted variable bias of any postulated strength, by simply replacing traditional critical values with the adjusted ones. Finally, we showed how intuitive visual displays can be deployed to fully characterize the sensitivity of IV to violations of its standard assumptions. In this work we have focused on the sensitivity analysis of the "traditional" IV estimand, consisting of the ratio of two OLS regression coefficients. We have chosen to do so because this encompasses the vast majority of current applied work using instrumental variables. Recent work, however, has questioned the causal interpretation of the traditional IV estimand, as it relies on strong parametric assumptions (Słoczyński, 2020; Blandhol et al., 2022). Extension of the sensitivity tools we present here to the nonparametric case is possible by leveraging recent results in Chernozhukov et al. (2022), and it is an interesting direction for future work.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human resources*, 40(4):791–821.

Amrhein, V. and Greenland, S. (2018). Remove, rather than redefine, statistical significance. *Nature Human Behaviour*, 2(1):4–4.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63.

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11:727–753.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4):69–85.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

Baiocchi, M., Cheng, J., and Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in medicine*, 33(13):2297–2340.

Balke, A. and Pearl, J. (1994). Counterfactual probabilities: Computational methods, bounds and applications. In *Uncertainty Proceedings 1994*, pages 46–54. Elsevier.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1):6.

Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2022). When is TSLS actually late? Technical report, National Bureau of Economic Research.

Bonet, B. (2001). Instrumentality tests revisited. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 48–55. Morgan Kaufmann Publishers Inc.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450.

Bowden, R. J. and Turkington, D. A. (1990). *Instrumental variables*, volume 8. Cambridge university press.

Burgess, S. and Thompson, S. G. (2015). *Mendelian randomization: methods for using genetic variants in causal estimation*. CRC Press.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, volume 3, pages 1801–1863. Elsevier.

Chernozhukov, V., Cinelli, C., Newey, W., Sharma, A., and Syrgkanis, V. (2022). Long story short: Omitted variable bias in causal machine learning. Technical report, National Bureau of Economic Research.

Cinelli, C., Ferwerda, J., and Hazlett, C. (2020). sensemakr: Sensitivity analysis tools for OLS in R and Stata. *Working Paper*.

Cinelli, C., Forney, A., and Pearl, J. (2021). A crash course in good and bad controls. *Sociological Methods & Research*, page 00491241221099552.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Cinelli, C., Kumor, D., Chen, B., Pearl, J., and Bareinboim, E. (2019). Sensitivity analysis of linear structural causal models. *International Conference on Machine Learning*.

Cinelli, C., LaPierre, N., Hill, B. L., Sankararaman, S., and Eskin, E. (2022). Robust mendelian randomization in the presence of residual population stratification, batch effects and horizontal pleiotropy. *Nature communications*, 13(1):1–13.

Cinelli, C. L. K. (2012). Inferência estatística e a prática econômica no brasil: os (ab) usos dos testes de significância.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute*, 22(1):173–203.

Deaton, A. S. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. Technical report, National bureau of economic research.

Didelez, V. and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical methods in medical research*, 16(4):309–330.

DiPrete, T. A. and Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological methodology*, 34(1):271–310.

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 16(2):175–185.

Fossaluza, V., Izbicki, R., da Silva, G. M., and Esteves, L. G. (2017). Coherent hypothesis testing. *The American Statistician*, 71(3):242–248.

Frisch, R. and Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401.

Gabriel, K. R. (1969). Simultaneous test procedures–some theory of multiple comparisons. *The Annals of Mathematical Statistics*, pages 224–250.

Gallen, T. (2020). Broken instruments. *Available at SSRN*.

Gunsilius, F. (2020). Non-testability of instrument validity under continuous treatments. *Biometrika*.

Heckman, J. J. and Urzua, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, 156(1):27–37.

Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, pages 360–372.

Hirschberg, J. and Lye, J. (2010). A geometric comparison of the delta and fieller confidence intervals. *The American Statistician*, 64(3):234–241.

Hirschberg, J. and Lye, J. (2017). Inverting the indirect—the ellipse and the boomerang: Visualizing the confidence intervals of the structural coefficient from two-stage least squares. *Journal of Econometrics*, 199(2):173–183.

Imbens, G. (2014). Instrumental variables: An econometrician's perspective. Technical report, National Bureau of Economic Research.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

Imbens, G. W. and Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109–126.

Imbens, G. W. and Rubin, D. B. (2015a). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Imbens, G. W. and Rubin, D. B. (2015b). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jiang, Y., Kang, H., and Small, D. S. (2018). ivmodel: An r package for inference and sensitivity analysis of instrumental variables models with one endogenous variable. *R package vignette*.

Kédagni, D. and Mourifié, I. (2020). Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika*.

Keele, L., Small, D., and Grieve, R. (2017). Randomization-based instrumental variables methods for binary outcomes with an application to the 'improve'trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(2):569–586.

Kruskal, W. and Majors, R. (1989). Concepts of relative importance in recent scientific literature. *The American Statistician*, 43(1):2–6.

LaPierre, N., Zhang, K., Hill, B., and Cinelli, C. (2021). PySensemakr: sensemakr for Python. `https://github.com/nlapier2/PySensemakr`.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010.

Lovell, M. C. (2008). A simple proof of the FWL theorem. *The Journal of Economic Education*, 39(1):88–91.

Mehlum, H. (2020). The polar confidence curve for a ratio. *Econometric Reviews*, 39(3):234–243.

Mellon, J. (2020). Rain, rain, go away: 137 potential exclusion-restriction violations for studies using weather as an instrumental variable. *Available at SSRN*.

Patriota, A. G. (2013). A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems*, 233:74–88.

Pearl, J. (1995). On the testability of causal models with latent and instrumental variables. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 435–443. Morgan Kaufmann Publishers Inc.

Pearl, J. (2009). *Causality*. Cambridge university press.

Perkovic, E., Textor, J., Kalisch, M., and Maathuis, M. H. (2018). Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18.

Robins, J. M. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, pages 113–159.

Rosenbaum, P. R. (1996). Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association*, 91(434):465–468.

Rosenbaum, P. R. (2002). Observational studies. In *Observational studies*, pages 1–17. Springer.

Rosenbaum, P. R. (2017). *Observation and experiment: an introduction to causal inference*. Harvard University Press.

Schervish, M. J. (1996). P values: what they are and what they are not. *The American Statistician*, 50(3):203–206.

Shpitser, I., VanderWeele, T., and Robins, J. M. (2012). On the validity of covariate adjustment for estimating causal effects. *arXiv preprint arXiv:1203.3515*.

Słoczyński, T. (2020). When should we (not) interpret linear iv estimands as late? *arXiv preprint arXiv:2011.06695*.

Small, D. S. (2007). Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association*, 102(479):1049–1058.

Small, D. S. and Rosenbaum, P. R. (2008). War and wages: the strength of instrumental variables and their sensitivity to unobserved biases. *Journal of the American Statistical Association*, 103(483):924–933.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression.

Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M., and Richardson, T. S. (2018). Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *Journal of the American Statistical Association*, 113(522):933–947.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The annals of mathematical statistics*, 11(3):284–300.

Wang, X., Jiang, Y., Zhang, N. R., and Small, D. S. (2018). Sensitivity analysis and power for instrumental variable studies. *Biometrics*.

Wright, P. G. (1928). *Tariff on animal and vegetable oils*. Macmillan Company, New York.

Young, A. (2022). Consistency without inference: Instrumental variables in practical application. *European Economic Review*, page 104112.

Ziliak, S. and McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. University of Michigan Press.

Appendix for "An Omitted Variable Bias Framework for Sensitivity Analysis of Instrumental Variables"

Carlos Cinelli & Chad Hazlett

A The mechanics of IV estimation

For ease of reference, in this section we show in detail some of the algebraic identities (and differences) of the main approaches to IV estimation.

Notation. We denote by Y the $(n \times 1)$ vector of the outcome of interest with n observations; by D the $(n \times 1)$ treatment vector; by Z the $(n \times 1)$ vector of the instrument; by $\boldsymbol{X}$ an $(n \times p)$ matrix of observed covariates (including a constant), and by $\boldsymbol{W}$ an $(n \times l)$ matrix of unobserved covariates. We use $Y^{\perp \boldsymbol{X}}$ to denote the part of Y not linearly explained by $\boldsymbol{X}$, that is, $Y^{\perp X} := Y - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}Y$. Throughout, we assume that the relevant matrices have full rank. Here $\text{df} := n - p - l - 1$.

A.1 Indirect Least Squares (ILS)

ILS is perhaps the most straightforward approach to instrumental variable estimation. We start with two OLS models, one capturing the effect of the instrument on the treatment (first stage) and another the effect of the instrument on the outcome (reduced form),

$$\textbf{First stage:} \quad D = \hat{\theta}Z + \boldsymbol{X}\hat{\psi} + \boldsymbol{W}\hat{\delta} + \hat{\varepsilon}_d \tag{47}$$

$$\textbf{Reduced form:} \quad Y = \hat{\lambda}Z + \boldsymbol{X}\hat{\beta} + \boldsymbol{W}\hat{\gamma} + \hat{\varepsilon}_y \tag{48}$$

Where $\hat{\theta}$, $\hat{\psi}$ and $\hat{\delta}$ are the OLS estimates of the regression of D on Z, $\boldsymbol{X}$ and $\boldsymbol{W}$, and $\hat{\varepsilon}_d$ its corresponding residuals; analogously, $\hat{\lambda}$, $\hat{\beta}$ and $\hat{\gamma}$ are the OLS estimates of the regression of Y on Z, $\boldsymbol{X}$ and $\boldsymbol{W}$, and $\hat{\varepsilon}_y$ its corresponding residuals.

Point Estimate. The estimator for τ is constructed by simply using the plug-in principle and taking the ratio of $\hat{\lambda}$ and $\hat{\theta}$

$$\hat{\tau}_{\text{ILS}} := \frac{\hat{\lambda}}{\hat{\theta}} \tag{49}$$

Inference. Inference in the ILS framework is usually performed using the delta-method, with estimated variance

$$\widehat{\text{var}}(\hat{\tau}_{\text{ILS}}) := \frac{1}{\hat{\theta}^2}\left(\widehat{\text{var}}(\hat{\lambda}) + \hat{\tau}^2_{\text{ILS}}\widehat{\text{var}}(\hat{\theta}) - 2\hat{\tau}_{\text{ILS}}\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})\right) \tag{50}$$

where, using the FWL formulation,

$$\widehat{\text{var}}(\hat{\lambda}) = \frac{\text{var}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1}, \qquad \widehat{\text{var}}(\hat{\theta}) = \frac{\text{var}(D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{51}$$

are the estimated variances of the reduced form and first stage, and

$$\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) = \frac{\text{cov}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}, D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{52}$$

is the estimated covariance of $\hat{\lambda}$ and $\hat{\theta}$. Here $\text{var}(\cdot)$ and $\text{cov}(\cdot)$ denote *sample* variances of covariances.

A.2 Two-Stage Least Squares (2SLS)

A closely related approach for instrumental variable estimation is denoted by "two-stage least squares" (2SLS). As its name suggests, this involves two nested steps of OLS estimation: a first-stage regression given by Equation 47 to produce fitted values for the treatment ($\widehat{D}$), then regressing the outcome on these fitted values,

$$\textbf{Second stage:} \quad Y = \hat{\tau}_{\text{2SLS}} \widehat{D} + \boldsymbol{X}\hat{\beta}_{\text{2SLS}} + \boldsymbol{W}\hat{\gamma}_{\text{2SLS}} + \hat{\varepsilon}_{\text{2SLS}} \tag{53}$$

The 2SLS estimate corresponds to the coefficient $\hat{\tau}_{\text{2SLS}}$ in Equation 53, called the "second-stage" regression.

Point Estimate. By the FWL theorem, the 2SLS point estimate can be written as

$$\hat{\tau}_{\text{2SLS}} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} \tag{54}$$

In the just-identified case, the ILS and 2SLS point estimates are numerically identical. Expanding $\widehat{D}$ and partialling out $\{\boldsymbol{X}, \boldsymbol{W}\}$ we have that

$$\hat{\tau}_{\text{2SLS}} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \tag{55}$$

$$= \frac{\hat{\theta} \times \text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, Z^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\hat{\theta}^2 \times \text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\hat{\lambda}}{\hat{\theta}} \tag{56}$$

Which establishes the equality $\hat{\tau}_{\text{2SLS}} = \hat{\tau}_{\text{ILS}} =: \hat{\tau}$.

Inference. By the FWL theorem, the standard two-stage least squares estimate of the variance of $\hat{\tau}_{\text{2SLS}}$ can be written as

$$\widehat{\text{var}}(\hat{\tau}_{\text{2SLS}}) := \frac{\text{var}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}} - \hat{\tau} D^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{57}$$

As with the point estimate, for the just-identified case, the estimated variance of ILS and 2SLS are numerically identical. To see why, note the denominator of Equation 57 can be expanded to

$$\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}}) = \text{var}(\hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}}) = \hat{\theta}^2 \, \text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}}) \tag{58}$$

Finally, the numerator can be written as,

$$\begin{aligned}
\text{var}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \hat{\tau} D^{\perp \boldsymbol{X},\boldsymbol{W}}) &= \text{var}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \hat{\tau}(\hat{\theta} Z^{\boldsymbol{X},\boldsymbol{W}} + D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})) && (59)\\
&= \text{var}((Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \hat{\lambda} Z^{\boldsymbol{X},\boldsymbol{W}}) - \hat{\tau} D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) && (60)\\
&= \text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}} - \hat{\tau} D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) && (61)\\
&= \text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) + \hat{\tau}^2 \,\text{var}(D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) - 2\hat{\tau}\,\text{cov}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}, D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) && (62)
\end{aligned}$$

Plugging in Equations 62 and 58 back in Equation 57, then using Equations 51 and 52 establishes the desired equality.

A.3 Anderson-Rubin (AR)

The Anderson-Rubin approach (Anderson and Rubin, 1949) starts by creating the random variable $Y_{\tau_0} := Y - \tau_0 D$ in which we subtract from Y a "putative" causal effect of D, namely, τ_0. If Z is a valid instrument, under the null hypothesis $H_0 : \tau = \tau_0$, we should not see an association between Y_{τ_0} and Z, conditional on $\boldsymbol{X}$ and $\boldsymbol{W}$. In other words, if we run the OLS model

$$\textbf{Anderson-Rubin:} \quad Y_{\tau_0} = \hat{\phi}_{\tau_0} Z + \boldsymbol{X}\hat{\beta}_{\tau_0} + \boldsymbol{W}\hat{\gamma}_{\tau_0} + \hat{\varepsilon}_{\tau_0} \tag{63}$$

we should find that $\hat{\phi}_{\tau_0}$ is equal to zero, but for sampling variation. This forms the basis for the point estimate and confidence interval in the AR approach.

Point Estimate. We define the Anderson-Rubin point estimate to be the value of τ_0 that makes $\hat{\phi} = 0$, ie,

$$\hat{\tau}_{\text{AR}} = \{\tau_0;\ \hat{\phi}_{\tau_0} = 0\} \tag{64}$$

Resorting again to the FWL theorem, we can write the regression coefficient of the AR regression, $\hat{\phi}_{\tau_0}$, as a function of the regression coefficients of the first stage and reduced form,

$$\begin{aligned}
\hat{\phi}_{\tau_0} &= \frac{\text{cov}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \tau_0 D^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} && (65)\\
&= \frac{\text{cov}(Y^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} - \tau_0 \frac{\text{cov}(D^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} && (66)\\
&= \hat{\lambda} - \tau_0 \hat{\theta} && (67)
\end{aligned}$$

Thus solving for the condition $\hat{\phi}_{\tau_0} = 0$ gives us

$$\hat{\tau}_{AR} = \frac{\hat{\lambda}}{\hat{\theta}} \tag{68}$$

Which establishes the equality $\hat{\tau}_{AR} = \hat{\tau}_{ILS}$. Therefore, all the point estimates of ILS, 2SLS and AR are numerically identical.

Inference. The AR confidence interval with significance level α is defined as all values of τ_0 such that we cannot reject the null hypothesis $H_0 : \phi_{\tau_0} = 0$ at the chosen significance level

$$\text{CI}_{1-\alpha}(\tau) = \{\tau_0; t^2_{\hat{\phi}_{\tau_0}} \leq t^{*2}_{\alpha,\text{df}}\} \tag{69}$$

This confidence interval can be obtained analytically as functions of the estimates of the first-stage and reduced form regressions. As shown in Equation 67, $\hat{\phi}_{\tau_0}$ can be written as the linear combination

$$\hat{\phi}_{\tau_0} = \hat{\lambda} - \tau_0 \hat{\theta} \tag{70}$$

Likewise, by the FWL theorem, the estimated variance of $\hat{\phi}_{\tau_0}$ is given by

$$\widehat{\text{var}}(\hat{\phi}_{\tau_0}) = \frac{\text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}} - \tau_0 D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} \times \text{df}^{-1} \tag{71}$$

$$= \left(\frac{\text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} + \tau_0^2 \frac{\text{var}(D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} - 2\tau_0 \frac{\text{cov}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}, D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} \right) \times \text{df}^{-1} \tag{72}$$

$$= \widehat{\text{var}}(\hat{\lambda}) + \tau_0^2 \widehat{\text{var}}(\hat{\theta}) - 2\tau_0 \widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) \tag{73}$$

Thus, we obtain that the t-value $t_{\hat{\phi}_{\tau_0}}$ for testing the null hypothesis $H_0 : \phi_{\tau_0} = 0$ equals to

$$t_{\hat{\phi}_{\tau_0}} = \frac{\hat{\lambda} - \tau_0 \hat{\theta}}{\sqrt{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2 \widehat{\text{var}}(\hat{\theta}) - 2\tau_0 \widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})}} \tag{74}$$

And our task is to find all values of τ_0 such that the following inequality holds

$$\frac{(\hat{\lambda} - \tau_0 \hat{\theta})^2}{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2 \widehat{\text{var}}(\hat{\theta}) - 2\tau_0 \widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})} \leq t^{*2}_{\alpha,\text{df}} \tag{75}$$

First, note that the empty set is not possible here. If we pick $\tau_0 = \hat{\tau}_{\text{AR}}$, then the numerator in Equation 75 is zero, and the inequality trivially holds—therefore, the point-estimate is always included in the confidence interval. Now squaring and rearranging terms we obtain

$$\underbrace{\left(\hat{\theta}^2 - \widehat{\text{var}}(\hat{\theta}) \times t^{*2}_{\alpha,\text{df}}\right)}_{a} \tau_0^2 + \underbrace{2\left(\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) \times t^{*2}_{\alpha,\text{df}} - \hat{\lambda}\hat{\theta}\right)}_{b} \tau_0 + \underbrace{\left(\hat{\lambda}^2 - \widehat{\text{var}}(\hat{\lambda}) \times t^{*2}_{\alpha,\text{df}}\right)}_{c} \leq 0 \tag{76}$$

Our task has simplified to find all values of τ_0 that makes the above quadratic equation, with coefficients a, b and c, non-positive. As discussed in Section 4.2.2, this confidence intervals can take three different forms, depending on the instrument strength: (i) finite and connected, (ii) the union two disjoint half lines; or, (iii) the whole real line.

A.4 Fieller's theorem

Fieller's proposal to test the null hypothesis $H_0 : \tau = \tau_0$ is to construct the linear combination $\hat{\phi}_{\tau_0} = \hat{\lambda} - \tau_0 \hat{\theta}$, and to test the null hypothesis $H_0 : \phi_{\tau_0} = 0$. The standard estimated variance for $\hat{\phi}_{\tau_0}$ equals Equation 73, resulting in a test statistic equal to Equation 74, and thus numerically identical to the AR approach.

B Bias-adjusted critical values and set of compatible inferences

B.1 Bias-adjusted critical values

As in the main text, using the reduced form as an example, let $\text{LL}_{1-\alpha}(\lambda) := \hat{\lambda} - t^*_{\alpha,\text{df}-1} \times \widehat{\text{se}}(\hat{\lambda})$ be the lower limit of a $1-\alpha$ level confidence interval of the full reduced form regression, where $t^*_{\alpha,\text{df}-1}$ denotes the critical α-level threshold of the t-distribution with df -1 degrees of freedom. Considering the direction of the bias that reduces the lower limit, Equations 12 and 14 imply

$$\begin{aligned}
\text{LL}_{1-\alpha}(\lambda) &:= \hat{\lambda} - t^*_{\alpha,\text{df}-1} \times \widehat{\text{se}}(\hat{\lambda}) && (77)\\
&= \hat{\lambda}_{\text{res}} - \text{BF}\sqrt{\text{df}} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) - t^*_{\alpha,\text{df}-1} \times \text{SEF}\sqrt{\text{df}/(\text{df}-1)} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) && (78)\\
&= \hat{\lambda}_{\text{res}} - \left(\text{SEF}\sqrt{\text{df}/(\text{df}-1)} \times t^*_{\alpha,\text{df}-1} + \text{BF}\sqrt{\text{df}}\right) \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) && (79)
\end{aligned}$$

Similarly, now let $\text{UL}_{1-\alpha}(\lambda)$ the upper limit of the confidence interval and consider the direction of the bias that increases the upper limit. By the same algebraic manipulations, we obtain

$$\text{UL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\text{res}} + \left(\text{SEF}\sqrt{\text{df}/(\text{df}-1)} \times t^*_{\alpha,\text{df}-1} + \text{BF}\sqrt{\text{df}}\right) \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) \quad (80)$$

Note that, in both Equations 79 and 80, the only part that depends on the omitted variable W is the common multiple of the observed standard error, which defines the new *bias-adjusted critical value*,

$$t^\dagger_{\alpha,\text{df}-1,\boldsymbol{R}^2} := \text{SEF}\sqrt{\text{df}/(\text{df}-1)} \times t^*_{\alpha,\text{df}-1} + \text{BF}\sqrt{\text{df}}. \quad (81)$$

B.2 Compatible inferences given bounds on the partial R^2

Now suppose the analyst wishes to investigate the worst possible lower (or upper) limits of the confidence intervals induced by a confounder with strength no stronger than certain bounds, for instance, $R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}}$. As per the last section, this amounts to finding the largest *bias-adjusted critical value* induced by an omitted variable W with at most such strength. That is, we need to solve the following maximization problem

$$\max_{R^2_{Y\sim W|Z,\boldsymbol{X}}, R^2_{Z\sim W|\boldsymbol{X}}} t^\dagger_{\alpha,\text{df}-1,\boldsymbol{R}^2} \quad \text{s.t.} \quad R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}, \quad R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \quad (82)$$

Dividing $t^\dagger_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ by $\sqrt{\text{df}}$ and letting $f^*_{\alpha,\text{df}-1} := t^*_{\alpha,\text{df}-1}/\sqrt{\text{df}-1}$, we see that the derivative of $t^\dagger_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ with respect to $R^2_{Z\sim W|\boldsymbol{X}}$ is always increasing, since

$$\begin{aligned}
\frac{\partial(t^\dagger_{\alpha,\text{df}-1,\boldsymbol{R}^2}/\sqrt{\text{df}})}{\partial R^2_{Z\sim W|\boldsymbol{X}}} &= \frac{\partial\,\text{BF}}{\partial R^2_{Z\sim W|\boldsymbol{X}}} + f^*_{\alpha,\text{df}-1} \times \frac{\partial\,\text{SEF}}{\partial R^2_{Z\sim W|\boldsymbol{X}}} && (83)\\
&= \frac{(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} + f^*_{\alpha,\text{df}-1}\frac{(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}} && (84)\\
&= \frac{(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2} + f^*_{\alpha,\text{df}-1}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \geq 0 && (85)
\end{aligned}$$

Therefore, the "optimal" $R^{2*}_{Z\sim W|\boldsymbol{X}}$ (the one the minimizes (maximizes) the lower (upper) limit of the confidence interval) always reaches the bound. However, the same is not true for the derivative with respect to $R^2_{Y\sim W|Z,\boldsymbol{X}}$. To see that, write,

$$\frac{\partial(t^{\dagger}_{\alpha,\mathrm{df}\,-1,\boldsymbol{R}^2}/\sqrt{\mathrm{df}})}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} = \frac{\partial\,\mathrm{BF}}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} + f^*_{\alpha,\mathrm{df}\,-1} \times \frac{\partial\,\mathrm{SEF}}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} \tag{86}$$

$$= \frac{(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}} + \frac{-f^*_{\alpha,\mathrm{df}\,-1}}{2(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \tag{87}$$

$$= \frac{(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2} - f^*_{\alpha,\mathrm{df}\,-1}(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \tag{88}$$

That is, due to the variance reduction factor of the omitted variable (VRF in Equation 14), it could be the case that increasing $R^2_{Y\sim W|Z,\boldsymbol{X}}$ reduces the standard error more than enough to compensate for the increase in bias, resulting in tighter confidence intervals.

We have, thus, two cases. First, consider the case in which the optimal point reaches both bounds. In that case, the numerator of Equation 88 must be positive when evaluated at the solution. Rearranging and squaring, we see that this happens when

$$R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \geq f^{*2}_{\alpha,\mathrm{df}\,-1} \times f^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}} \tag{89}$$

Clearly, when considering the sensitivity of the point estimate, we have $f^*_{\alpha,\mathrm{df}\,-1} = 0$, and the condition always holds. If condition of Equation 89 fails, then the optimal $R^{2*}_{Y\sim W|Z,\boldsymbol{X}}$ will be an interior point. This will happen when the numerator of Equation 88 equals zero. Since we know $R^2_{Z\sim W|\boldsymbol{X}}$ reaches its maximum, the optimal $R^{2*}_{Y\sim W|Z,\boldsymbol{X}}$ will be,

$$R^{2*}_{Y\sim W|Z,\boldsymbol{X}} = \frac{R^{2\,\max}_{Z\sim W|\boldsymbol{X}}}{f^{*2}_{\alpha,\mathrm{df}\,-1} + R^{2\,\max}_{Z\sim W|\boldsymbol{X}}} \tag{90}$$

Denoting the solution to the optimization problem as $t^{\dagger\,\max}_{\alpha,\mathrm{df}\,-1,\boldsymbol{R}^2}$, the *most extreme possible* lower and upper limits after adjusting for W are given by

$$\mathrm{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) = \hat{\lambda}_{\mathrm{res}} - t^{\dagger\,\max}_{\alpha,\mathrm{df}\,-1,\boldsymbol{R}^2} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}), \quad \mathrm{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2} = \hat{\lambda}_{\mathrm{res}} + t^{\dagger\,\max}_{\alpha,\mathrm{df}\,-1,\boldsymbol{R}^2} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) \tag{91}$$

And interval composed of such limits

$$\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) = \Big[\mathrm{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda), \quad \mathrm{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)\Big] \tag{92}$$

Defines the set of compatible inferences given the bounds on the partial R^2, $R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}}$.

C (Extreme) Robustness Values

C.1 The Extreme Robustness Value

The *Extreme Robustness Value* $\text{XRV}_{q^*,\alpha}(\lambda)$ is defined as the greatest lower bound XRV on the sensitivity parameter $R^2_{Z\sim W|\boldsymbol{X}}$, while keeping the parameter $R^2_{Y\sim W|Z,\boldsymbol{X}}$ unconstrained, such that the null hypothesis that a change of $(100\times q)\%$ of the original estimate, $H_0 : \lambda = (1-q^*)\hat{\lambda}_{\text{res}}$, is not rejected at the α level:

$$\text{XRV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{XRV};\ (1-q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\lambda)\right\} \tag{93}$$

First, consider the case where $f_{q^*}(\lambda) < f^*_{\alpha,\text{df}-1}$. Note the XRV will be zero, since we already cannot reject the null hypothesis $H_0 : \lambda = (1-q^*)\hat{\lambda}_{\text{res}}$ even assuming zero omitted variable bias. Next, note that, when $f^*_{\alpha,\text{df}-1} > 0$, we can always pick a large enough value for $R^2_{Y\sim W|Z,\boldsymbol{X}}$ until condition 89 fails (since $f^2_{Y\sim W|Z,\boldsymbol{X}}$ is unbounded). Therefore, XRV will be given by an interior point solution. Using Equation 90 to express $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ solely in terms of the optimal $R^2_{Z\sim W|\boldsymbol{X}}$, and solving for the value that gives us $(1-q^*)\hat{\lambda}_{\text{res}}$, we obtain

$$\text{XRV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if}\ \ f_{q^*}(\lambda) \le f^*_{\alpha,\text{df}-1} \\ \dfrac{f^2_{q^*}(\lambda) - f^{*2}_{\alpha,\text{df}-1}}{1+f^2_{q^*}(\lambda)}, & \text{otherwise.} \end{cases} \tag{94}$$

C.2 The Robustness Value

The *Robustness Value* $\text{RV}_{q^*,\alpha}(\lambda)$ for not rejecting the null hypothesis that $H_0 : \lambda = (1-q^*)\hat{\lambda}_{\text{res}}$, at the significance level α, is defined as

$$\text{RV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{RV};\ (1-q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\lambda)\right\} \tag{95}$$

Where now we consider both sensitivity parameters bounded by RV. Again, consider the case where $f_{q^*}(\lambda) < f^*_{\alpha,\text{df}-1}$. The RV then must be zero, since we already cannot reject the null hypothesis $H_0 : \lambda = (1-q^*)\hat{\lambda}_{\text{res}}$ given the current data. Next, let's consider the case when the bound on $R^2_{Y\sim W|Z,\boldsymbol{X}}$ is not biding—here our optimization problem reduces to the XRV case. Finally, we have the solution in which both coordinates achieve the bound, resulting in a quadratic equation as solved in Cinelli and Hazlett (2020). We thus have,

$$\text{RV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if}\ \ f_{q^*}(\lambda) \le f^*_{\alpha,\text{df}-1} \\ \dfrac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right), & \text{if}\ \ f^*_{\alpha,\text{df}-1} < f_{q^*}(\lambda) < f^{*-1}_{\alpha,\text{df}-1} \\ \text{XRV}_{q^*,\alpha}(\lambda), & \text{otherwise.} \end{cases} \tag{96}$$

The condition $f_{q*}(\lambda) < f^{*-1}_{\alpha,\text{df}-1}$, stems from the fact that the XRV solution cannot satisfy Equation 89. We now show that this is equivalent to the condition $\text{RV}_{q^*,\alpha}(\lambda) > 1 - 1/f^2_{q^*}(\lambda)$ that Cinelli and Hazlett (2020) had previously established. If $f_{q^*}(\lambda) < 1/f^*_{\alpha,\text{df}-1}$ then,

$$\mathrm{RV}_{q^*,\alpha}(\lambda) = \frac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right) \tag{97}$$

$$= \frac{1}{2}\left(\sqrt{(f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^4 + 4(f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^2} - (f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^2\right) \tag{98}$$

$$> \frac{1}{2}\left(\sqrt{(f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^4 + 4(f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^2} - (f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^2\right) \tag{99}$$

$$= \frac{1}{2}\left(\sqrt{\left(\frac{f^2_q(\lambda)-1}{f_{q^*}(\lambda)}\right)^4 + 4\left(\frac{f^2_{q^*}(\lambda)-1}{f_{q^*}(\lambda)}\right)^2} - \left(\frac{f^2_{q^*}(\lambda)-1}{f_{q^*}(\lambda)}\right)^2\right) \tag{100}$$

$$= \left(\frac{1}{2}\right)\left(\frac{f^2_{q^*}(\lambda)-1}{f^2_{q^*}(\lambda)}\right)\left(\sqrt{(f^2_q(\lambda)-1)^2 + 4f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) \tag{101}$$

$$= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(\sqrt{f^4_q(\lambda) + 1 - 2f^2_{q^*}(\lambda) + 4f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) \tag{102}$$

$$= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(\sqrt{f^4_q(\lambda) + 1 + 2f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) \tag{103}$$

$$= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(f^2_{q^*}(\lambda) + 1 - f^2_{q^*}(\lambda) + 1\right) \tag{104}$$

$$= 1 - 1/f^2_{q^*}(\lambda) \tag{105}$$

Therefore, $f_{q^*}(\lambda) < 1/f^*_{\alpha,\mathrm{df}-1} \implies \mathrm{RV}_{q^*,\alpha}(\lambda) > 1 - 1/f^2_{q^*}(\lambda)$. By the same argument one can derive $\mathrm{RV}_{q^*,\alpha}(\lambda) > 1 - 1/f^2_{q^*}(\lambda) \implies f_q(\lambda) > 1/f^*_{\alpha,\mathrm{df}-1}$. Hence, both conditions are equivalent. The new condition, however, is much simpler to verify.

D Bounds on the strength of W

Let X_j be a specific covariate of the set $\boldsymbol{X}$. Now define

$$k_Z := \frac{R^2_{Z\sim W|\boldsymbol{X}_{-j}}}{R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{R^2_{Y\sim X_j|Z\boldsymbol{X}_{-j}}}. \tag{106}$$

Where $\boldsymbol{X}_{-j}$ is the set $\boldsymbol{X}$ excluding covariate X_j. Our goal in this section is to re-express (or bound) both sensitivity parameters as a function of the new parameters k_Z and k_Y and the observed data.

Cinelli and Hazlett (2020) showed how to obtains bounds for the strength of W under the assumption that $R^2_{W\sim X_j|\boldsymbol{X}_{-j}} = 0$, or, equivalently, when we consider the part of W not linearly explained by $\boldsymbol{X}$. This result may be particularly useful when considering both $\boldsymbol{X}$ and W as *causes* of Z, as in such cases contemplating the marginal orthogonality of W (or its part not explained by observed covariates) is more natural.

Here we additionally provide bounds under the assumption that $R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0$. This condition may be helpful when contemplating the strength of W against X_j whenever these variables are *side-effects* of Z, instead of causes of Z. In such cases, reasoning about the marginal orthogonality of W with respect to $\boldsymbol{X}$ may not be natural, as Z itself is also a source of dependence between these variables.

We can thus start by re-expressing $R^2_{Y\sim W|Z,\boldsymbol{X}}$ in terms of k_Y, which in this case is straight-

forward. Using the recursive definition of partial correlations, and considering our two conditions $R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0$ and $R^2_{Y\sim W|Z,\boldsymbol{X}_{-j}} = k_Y R^2_{Y\sim X_j|Z\boldsymbol{X}_{-j}}$, we obtain

$$\left|R_{Y\sim W|Z,\boldsymbol{X}}\right| = \left|\frac{R_{Y\sim W|Z,\boldsymbol{X}_{-j}} - R_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} R_{W\sim X_j|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}}}}\right| \tag{107}$$

$$= \left|\frac{R_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}}\right| \tag{108}$$

$$= \left|\frac{\sqrt{k_Y} R_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}}\right| \tag{109}$$

$$= \sqrt{k_Y}\left|f_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}\right| \tag{110}$$

Hence,

$$R^2_{Y\sim W|Z,\boldsymbol{X}} = k_Y \times f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} \tag{111}$$

Moving to bound $R^2_{Z\sim W|\boldsymbol{X}}$, it is useful to first note that the conditions $R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0$ and $R^2_{Z\sim W|\boldsymbol{X}_{-j}} = k_Z R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}$ allow us to re-express $R_{W\sim X_j|\boldsymbol{X}_{-j}}$ as a function of k_Z

$$R_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0 \implies \frac{R_{W\sim X_j|\boldsymbol{X}_{-j}} - R_{W\sim Z|\boldsymbol{X}_{-j}} R_{X_j\sim Z|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{W\sim Z|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{X_j\sim Z|\boldsymbol{X}_{-j}}}} = 0 \tag{112}$$

$$\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} - R_{W\sim Z|\boldsymbol{X}_{-j}} R_{X_j\sim Z|\boldsymbol{X}_{-j}} = 0 \tag{113}$$

$$\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} = R_{W\sim Z|\boldsymbol{X}_{-j}} R_{X_j\sim Z|\boldsymbol{X}_{-j}} \tag{114}$$

$$\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} = R_{Z\sim W|\boldsymbol{X}_{-j}} R_{Z\sim X_j|\boldsymbol{X}_{-j}} \tag{115}$$

$$\implies |R_{W\sim X_j|\boldsymbol{X}_{-j}}| = \sqrt{k_Z} R^2_{Z\sim X_j|\boldsymbol{X}_{-j}} \tag{116}$$

Now we can re-write $R^2_{Z\sim W|\boldsymbol{X}}$ using the recursive definition of partial correlations

$$\left|R_{Z\sim W|\boldsymbol{X}}\right| = \left|\frac{R_{Z\sim W|\boldsymbol{X}_{-j}} - R_{Z\sim X_j|\boldsymbol{X}_{-j}} R_{W\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|\boldsymbol{X}_{-j}}}}\right| \tag{117}$$

$$\leq \frac{\left|R_{Z\sim W|\boldsymbol{X}_{-j}}\right| + \left|R_{Z\sim X_j|\boldsymbol{X}_{-j}} R_{W\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|\boldsymbol{X}_{-j}}}} \tag{118}$$

$$= \frac{\left|\sqrt{k_Z} R_{Z\sim X_j|\boldsymbol{X}_{-j}}\right| + \left|\sqrt{k_Z} R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}} \tag{119}$$

$$= \left(\frac{\sqrt{k_Z} + \left|R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right) \times \left(\frac{\left|R_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right) \tag{120}$$

$$= \eta' |f_{Z\sim X_j|\boldsymbol{X}_{-j}}| \tag{121}$$

Hence we have that

$$R^2_{Z\sim W|\boldsymbol{X}} \leq {\eta'}^2 f^2_{Z\sim X_j|\boldsymbol{X}_{-j}} \tag{122}$$

Where $\eta' = \left(\frac{\sqrt{k_Z}+\left|R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right)$.

E Comparison with traditional approaches

Traditional approaches for the sensitivity of IV have focused on parameterizing the bias of the IV estimate with a single coefficient that summarizes how strongly the instrument relates to the outcome "not through" the treatment. For example, Conley et al. (2012) considers the model (for simplicity, we omit covariates $\boldsymbol{X}$):

$$Y_i = \tau D_i + \eta Z_i + \varepsilon_i \tag{123}$$

Where τ is the parameter of interest, and $\text{cov}(Z_i, \varepsilon_i) = 0$. Here, the coefficient η is a sensitivity parameter that directly summarizes violations of instrument validity. To recover the target parameter τ, it thus suffices to subtract η from the reduced-form regression coefficient λ,

$$\tau = \frac{\lambda - \eta}{\theta}. \tag{124}$$

Inference for the above estimand can be done in numerous ways. At a given choice of η, one could simply subtract the postulated bias from the reduced form estimate; similarly, confidence intervals can be obtained using the delta-method. Another popular, and computationally simpler alternative is to construct an auxiliary outcome $Y_\eta := Y - \eta Z$, and then proceed with any of the estimation methods discussed here (e.g, 2SLS or Anderson-Rubin regression) using the auxiliary variable Y_η instead of Y.

Applying this approach to our running example we reach the correct, but perhaps trivial con-

clusion that, in order to bring the causal effect estimate to zero ($\tau = 0$), all of the reduced-form estimate (4.2%) must be due to the effects of proximity to college on income, *not* through its effect on years of schooling, i.e. $\eta = 4.2\%$. Other approaches, although different in details, can be understood in similar terms. For instance, starting from a potential outcomes framework, Wang et al. (2018) obtains a similar sensitivity model as Equation 123, and derive the distribution of the Anderson-Rubin statistic for a given postulated value of η.

In contexts where researchers can make direct plausibility judgments about the coefficient η, these approaches offer a simple and useful sensitivity analysis. In many cases, however, such as in our running example, violations of instrument validity arise due to many possible confounding variables acting in concert, such as family wealth, high school quality, and regional indicators. How can we reason whether all these variables are strong enough to bring about an $\eta \approx 4.2\%$? The OVB approach we present here change the focus from η to the omitted variables $\boldsymbol{W}$. That is, instead of asking for direct judgments about η, the OVB approach reveals what one must believe about the maximum explanatory power of such omitted variables in order for them to be problematic. Here $\boldsymbol{W}$ consists of the necessary set of variables to block both confounding between the instrument and the outcome, as well as blocking paths from the instrument to the outcome, not through the treatment (e.g, see Figure 4).

Finally, it is worth mentioning that these two approaches are not necessarily mutually exclusive. To illustrate, suppose we have a structural model

$$Y_i = \tau D_i + \eta Z_i + \gamma W + \varepsilon_i \tag{125}$$

with $\text{cov}(Z_i, \varepsilon_i) = 0$. Here suppose η now effectively stands for the direct effect of Z on Y, not through D nor W. If plausibility judgments on the direct effect of Z are available, we can leverage such knowledge by first subtracting this off and then employing all OVB-based tools we have presented in this paper to perform sensitivity analysis with respect to the remaining bias due to W.

F Supplementary Tables and Figures

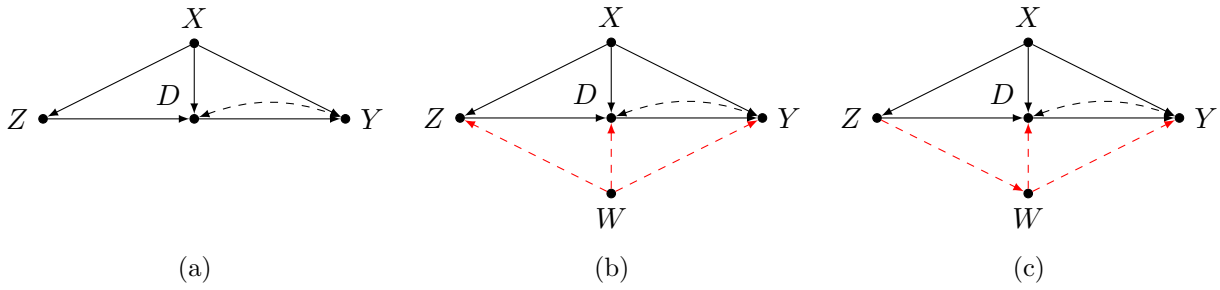


Figure 4: Causal diagrams illustrating traditional IV assumptions. Directed arrows, such as $X \to Y$, denote a possible direct causal effect of X on Y. Bidirected arrows, such as $D \leftrightarrow Y$, stand for latent common causes between D and Y. In Figure 4a, X is sufficient for rendering Z a valid instrumental variable. In Figures 4b and 4c, however, W is also needed to render Z a valid IV, either because it confounds the instrument-outcome relationship (Fig. 4b) or because it is a side-effect of the instrument affecting the outcome other than through its effect of on the treatment (Fig. 4c). In practice, all these violations will be happening simultaneously.

	Dependent variable:			
	Education	Earnings (log)		
	FS	RF	OLS	IV
	(1)	(2)	(3)	(4)
Proximity	0.320***	0.042**		
	(0.088)	(0.018)		
Education			0.075***	0.132**
			(0.003)	(0.055)
Black	−0.936***	−0.270***	−0.199***	−0.147***
	(0.094)	(0.019)	(0.018)	(0.054)
SMSA	0.402***	0.165***	0.136***	0.112***
	(0.105)	(0.022)	(0.020)	(0.032)
Other covariates	yes	yes	yes	yes
Observations	3,010	3,010	3,010	3,010
R^2	0.477	0.195	0.300	0.238
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Results of Card (1993). Columns show estimates and standard errors (in parenthesis) of the First Stage (FS), Reduced Form (RF), Ordinary Least Squares (OLS) and Two-Stage Least Squares (IV). *Black* is an indicator of race; *SMSA* an indicator for whether the individual lived in a metropolitan area. Following Card (1993), other covariates include age, regional indicators, experience and experience squared.