An Omitted Variable Bias Framework for Sensitivity Analysis of Instrumental Variables

BY CARLOS CINELLI
Department of Statistics, University of Washington, Seattle.
cinelli@uw.edu

AND CHAD HAZLETT
Department of Statistics, University of California, Los Angeles.
chazlett@ucla.edu

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) of the instrument, exploits expert knowledge to bound sensitivity parameters, and can be easily implemented with standard software. 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 is no stronger than a multiple of the explanatory power of observed variables. Conveniently, 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.

Some key words: Instrumental Variables; Omitted Variable Bias; Sensitivity Analysis; Robustness Values.

1. INTRODUCTION

Unobserved confounding often complicates efforts to make causal claims from observational data (e.g. Pearl, 2009; Imbens and Rubin, 2015). 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. Angrist et al., 1996; Angrist and Pischke, 2009). IV methods are "a central part of the econometrics canon since the first half of the twentieth century" (Imbens, 2014, p.324), and, beyond economics, are now 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; 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 be confounded with the outcome, and it should influence the outcome only by affecting 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 other than through the treatment.[1] Although in certain cases the IV assumptions may entail testable implications (Pearl, 1995; Gunsilius, 2020; Kédagni and Mourifié, 2020), they are often unverifiable and must be defended by appealing to domain knowledge. 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 (e.g., Deaton, 2009; Heckman and Urzua, 2010). Extensive reviews of many widely-used instrumental variables have cataloged several plausible violations of the exclusion restriction for such instruments (e.g. 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; Small, 2007; Small and Rosenbaum, 2008; Conley et al., 2012; Wang et al., 2018; Masten and Poirier, 2021), 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), it 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 sources of bias (see Supplementary Material 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 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. Building on the Anderson-Rubin approach to IV estimation (Anderson and Rubin, 1949) and on recent developments of OVB for ordinary least squares (OLS) (Cinelli and Hazlett, 2020), we develop a simple suite of sensitivity analysis tools for IV that: (i) has correct test size regardless of instrument strength; (ii) naturally handles violations due to multiple "side-effects" and "confounders," possibly acting non-linearly;

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

[2] In economics, only 1 out of 27 papers using IV 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. In Sociology, it was zero out of 34, in the *American Journal of Sociology* and the *American Sociology Review* from 2004 to 2022 (Felton and Stewart, 2022).

(iii) is well suited for routine reporting; and (iv) exploits expert knowledge to bound sensitivity parameters.[3]

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 outcome, in order to change 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* in order to be problematic. Routine reporting of these quantities provides a quick and 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 is 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 affirms that certain pivotal conclusions regarding the sensitivity of the IV estimate can be reached simply through separate sensitivity analyses of these two familiar auxiliary OLS estimates[4] A final contribution of this paper is the proposal of 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. The extreme simplicity of this approach may further aid in the adoption of sensitivity analysis in applied work. All proofs and details can be found in the Supplementary Materials.

2. Running example

We begin by introducing the running example and reviewing the required background on IV.

2.1. *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. 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 $Y = \hat{\tau}_{\text{OLS,res}} D + \boldsymbol{X}\hat{\beta}_{\text{OLS,res}} + \hat{\varepsilon}_{\text{OLS,res}}$, where Y denotes *Earnings* and measures the log transformed hourly wages of the individual D denotes *Education* and consists of an integer-valued

[3] Here we focus on the one treatment and one instrument ("just-identified") case. We do so for two reasons. First, thoroughly considering how identification assumptions may be violated is complicated enough with one instrument (Angrist and Pischke, 2009). Second, most applied IV work uses this approach. Reviewing papers in the *American Economic Review* and 15 other journals of the *American Economic Association*, Young (2022) finds that 80% of IV regressions used a single instrument. Even in "multiple instrument" studies, it is not uncommon for researchers to also report and give special focus to the analysis of their "best" single instrument (Angrist and Pischke, 2009), or to combine multiple instruments into a single instrument.

[4] In the context of randomization inference, similar observations have been noted by Rosenbaum (1996); Imbens and Rosenbaum (2005); Small and Rosenbaum (2008); Keele et al. (2017) and Rosenbaum (2017)

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.

Educational achievement, however, is not randomly assigned; perhaps individuals who obtain more education have higher wages for other reasons, such as family background, or higher levels of some other unobserved characteristic such as "ability" or "motivation." If data on these variables were available, then further adjustment for such variables would capture the causal effect of educational attainment on schooling, as in $Y = \hat{\tau}_{\text{OLS}} D + \boldsymbol{X}\hat{\beta}_{\text{OLS}} + \boldsymbol{U}\hat{\gamma}_{\text{OLS}} + \hat{\varepsilon}_{\text{OLS}}$, where $\boldsymbol{U}$ is a set of variables that, along with $\boldsymbol{X}$, eliminates confounding concerns.[5] Unfortunately, such detailed information on individuals is not available, and researchers may not agree on which variables $\boldsymbol{U}$ are needed. Regression estimates that adjust for only a partial list of characteristics (such as $\boldsymbol{X}$) may suffer from OVB, likely overestimating the "true" returns to schooling.

2.2. *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[6], one measuring the effect of *Proximity* on *Earnings*, and another measuring the effect of *Proximity* on *Education*, as in the two OLS models

$$\textbf{First Stage:} \quad D = \hat{\theta}_{\text{res}} Z + \boldsymbol{X}\hat{\psi}_{\text{res}} + \hat{\varepsilon}_{d,\text{res}} \tag{1}$$

$$\textbf{Reduced Form:} \quad Y = \hat{\lambda}_{\text{res}} Z + \boldsymbol{X}\hat{\beta}_{\text{res}} + \hat{\varepsilon}_{y,\text{res}} \tag{2}$$

Throughout the paper we refer to these equations as the "first stage" (Equation 1) and the "reduced form" (Equation 2), as these are now common usage (Angrist and Pischke, 2009; Imbens and Rubin, 2015; Andrews et al., 2019).The coefficient for *Proximity* (Z) on the first-stage regression 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 suggests that those who grew up near a college have 4.2% higher earnings. The IV estimate is then given by the ratio,

[5] I.e, the set $\{\boldsymbol{X}, \boldsymbol{U}\}$ is sufficient to render the treatment assignment ignorable. In graphical terms, the set would satisfy the backdoor criterion (see, e.g, Pearl, 2009; Angrist and Pischke, 2009). Beyond ignorability, if the treatment effect is heterogeneous, this may affect the causal interpretation of $\hat{\tau}_{\text{OLS}}$ (e.g. Angrist and Pischke, 2009).

[6] 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 conditions for a "weakly causal" interpretation of the traditional IV estimand. Here we start from the premise that the researcher has already decided she is interested in the results of Equations 4-6.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 our tools to nonparametric settings leveraging recent results in Chernozhukov et al. (2022). We leave this to future work.

$\hat{\tau}_{\text{res}} := \hat{\lambda}_{\text{res}}/\hat{\theta}_{\text{res}} \approx 0.042/0.319 \approx 0.132$. 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 $\hat{\lambda}_{\text{res}}/\hat{\theta}_{\text{res}}$ is sometimes called the *indirect least squares* (ILS) estimator, or the "ratio of coefficients" estimator. Inference in the ILS framework is usually performed using the delta-method. A closely related approach 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) one can readily show that 2SLS and ILS are numerically identical.

2.3. *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, known as the "weak instrument" problem.[7] The Anderson-Rubin (AR) regression (Anderson and Rubin, 1949) provides one approach to constructing confidence intervals with correct coverage, regardless of the "strength" of the first stage. It 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

$$\textbf{Anderson-Rubin:} \quad Y_{\tau_0} = \hat{\phi}_{\tau_0,\text{res}} Z + \boldsymbol{X}\hat{\beta}_{\tau_0,\text{res}} + \hat{\varepsilon}_{\tau_0,\text{res}} \tag{3}$$

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 $\hat{\tau}_{\text{AR,res}}$ is numerically identical to 2SLS and ILS.

2.4. *The IV estimate 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 is often the case, neither assumption is easy to defend. First, 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 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.

Therefore, instead of adjusting for $\boldsymbol{X}$ only, as in the previous regressions, we should have adjusted for both the observed covariates $\boldsymbol{X}$ and unobserved covariates $\boldsymbol{W}$ as in

$$\textbf{First Stage:} \quad D = \hat{\theta} Z + \boldsymbol{X}\hat{\psi} + \boldsymbol{W}\hat{\delta} + \hat{\varepsilon}_d \tag{4}$$

$$\textbf{Reduced Form:} \quad Y = \hat{\lambda} Z + \boldsymbol{X}\hat{\beta} + \boldsymbol{W}\hat{\gamma} + \hat{\varepsilon}_y \tag{5}$$

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

[7] 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).

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*).[8] Our task is thus to characterize how the IV point estimates and confidence intervals, as given by Equations 4-6, would have changed due to the inclusion of omitted variables $\boldsymbol{W}$.

2.5. *Linearity, heterogeneity, and inference*

Before proceeding with our main results, two remarks are in order. First, regarding statistical inference, throughout the paper we focus solely on exact algebraic results pertaining to "classical" (homoskedastic) standard errors. Conditions under which classical confidence intervals have nominal coverage are well-known and thus omitted. While this may seem restrictive to some readers, we note that the main concern of sensitivity analysis is *systematic bias*, and not sampling uncertainty. Moreover, as we explain in Section 3.1, inference with robust standard errors is straightforward using the bootstrap or the delta-method. Second, we note again that our target parameter is the traditional IV estimand, defined as the ratio $\tau := \lambda/\theta$. While treatment effect heterogeneity may affect the precise causal interpretation of τ, it has no bearing on the partial identification results we present here, which consist of bounds on (ratios of) linear projection coefficients. These bounds hold whether the true conditional expectation functions are linear or not, and even in the latter case, (ratios of) linear projections may still recover interesting causal parameters (Angrist and Pischke, 2009). See also Footnote 6 and Section 6.

3. OVB WITH THE PARTIAL R^2 PARAMETERIZATION

Our proposed sensitivity analysis of the IV estimate requires first extending recently developed tools for sensitivity analysis of OLS (Cinelli and Hazlett, 2020). These extensions are not only useful on their own, but importantly, for present purposes, they greatly simplify the development of a suite sensitivity analysis tools for IV in Section 4. Toward this end, in this section we first propose *bias-adjusted* critical values for OLS, which allows sensitivity analysis to be performed by simply substituting traditional critical values with adjusted ones (which we later apply to IV in the Anderson-Rubin setting). Next, we introduce new sensitivity statistics for routine reporting, such as extreme robustness values, characterizing the bare minimum strength that omitted variables must have to overturn certain conclusions. Finally, we derive a novel bound on the strength of omitted variables on the basis of comparison with observed variables.

3.1. *Bias-adjusted estimates and standard errors*

We first establish key ideas, formulae, and notations from prior work on OVB (Cinelli and Hazlett, 2020). For concreteness, suppose we are interested in the regression coefficient $\hat{\lambda}$ and the (estimated) standard error $\widehat{\mathrm{se}}(\hat{\lambda})$ of Equation 5, namely, the OLS regression of the outcome Y on the instrument Z, adjusting for a set of observed covariates $\boldsymbol{X}$ and (for now) a single *unobserved* covariate W (we generalize to multivariate W below). Here Y, Z and W are $(n \times 1)$ vectors, $\boldsymbol{X}$ is an $(n \times p)$ matrix (including a constant), with n observations; $\hat{\lambda}$, $\hat{\beta}$ and $\hat{\gamma}$ are the OLS coefficient estimates and $\hat{\varepsilon}_y$ the corresponding residuals. As W is unobserved, the investigator instead estimates the *restricted* model of Equation 2 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 $\hat{\varepsilon}_{y,\text{res}}$ the corresponding residuals. The OVB framework seeks to answer the following question: how do the inferences from the restricted OLS model compare with the inferences from the full OLS model?

[8] See Supplementary Material for "canonical" causal diagrams illustrating settings in which $\{\boldsymbol{X}, \boldsymbol{W}\}$ renders Z a valid instrument for the effect of D and Y. Equivalent assumptions can be articulated in the potential outcomes framework (Angrist et al., 1996; Pearl, 2009; Swanson et al., 2018).

Let $R^2_{Y\sim W|Z,\boldsymbol{X}}$ denote the (sample) 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,

$$|\widehat{\text{bias}}(\lambda)| = \sqrt{\frac{R^2_{Y\sim W|Z,\boldsymbol{X}} R^2_{Z\sim W|\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}} \times \frac{\widehat{\text{sd}}(Y^{\perp \boldsymbol{X},D})}{\widehat{\text{sd}}(D^{\perp \boldsymbol{X}})} = \text{BF} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}})\sqrt{\text{df}} \tag{7}$$

Where here $\widehat{\text{sd}}(Y^{\perp \boldsymbol{X},D})$ is the (sample) residual standard deviation of Y after removing the part linearly explained by $\{\boldsymbol{X}, D\}$, and $\widehat{\text{sd}}(D^{\perp \boldsymbol{X}})$ is the (sample) residual standard deviation of D after removing the part linearly explained by $\boldsymbol{X}$. To aid interpretation, we define the term $\text{BF} := \sqrt{\frac{R^2_{Y\sim W|Z,\boldsymbol{X}} R^2_{Z\sim W|\boldsymbol{X}}}{1-R^2_{Z\sim W|\boldsymbol{X}}}}$ 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 $R^2_{Z\sim W|\boldsymbol{X}}$. The second equality follows from the fact that the classical (estimated) standard error equals $\widehat{\text{se}}(\hat{\lambda}_{\text{res}}) = \frac{\widehat{\text{sd}}(Y^{\perp \boldsymbol{X},D})}{\widehat{\text{sd}}(D^{\perp \boldsymbol{X}})}\text{df}^{-1/2}$ (here $\text{df} = n - p - 1$ is the residual degrees of freedom from the restricted model actually run). We note the standard error in the bias formula is used mainly for computational convenience. Likewise, the classical (estimated) standard error of the full model can be recovered with

$$\widehat{\text{se}}(\hat{\lambda}) = \sqrt{\frac{1 - R^2_{Y\sim W|Z,\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}})\sqrt{\text{df}\,/(\text{df}\,-1)} = \text{SEF} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}})\sqrt{\text{df}\,/(\text{df}\,-1)} \tag{8}$$

where we similarly define $\text{SEF} := \sqrt{\frac{1-R^2_{Y\sim W|Z,\boldsymbol{X}}}{1-R^2_{Z\sim W|\boldsymbol{X}}}}$ 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 $R^2_{Z\sim W|\boldsymbol{X}}$.

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 $\boldsymbol{W} = [W_1, W_2, \ldots, W_l]$, 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).

Finally, we note sensitivity analyses could alternatively be made in terms of population parameters. As per Equation 7, the partially identified region for λ is given by:

$$\lambda_{\pm} = \lambda_{\text{res}} \pm \text{BF} \times \frac{\text{sd}(Y^{\perp \boldsymbol{X},D})}{\text{sd}(D^{\perp \boldsymbol{X}})} \tag{9}$$

where in Equation 9 all terms (including BF) stand for population quantities. Confidence intervals for the partially identified region $[\lambda_{-},\ \lambda_{+}]$ can then be constructed using traditional statistical inference methods, such as the bootstrap or the delta-method; see, for instance, Chernozhukov et al. (2022, Theorem 4). Throughout the paper we keep the analysis at the sample level, with the understanding that similar analyses at the population level can be easily done as outlined above.

3.2. *Bias-adjusted critical values*

We now introduce a novel correction to traditional critical values that researchers can use to account for omitted variable bias. Let $t^*_{\alpha,\text{df}-1}$ denote the critical value for a standard t-test with significance level α and $\text{df}-1$ degrees of freedom. Now let $\text{LL}_{1-\alpha}(\lambda)$ be the lower limit and $\text{UL}_{1-\alpha}(\lambda)$ be the upper limit of a $1-\alpha$ confidence interval for λ in the full model, i.e.,

$$\text{LL}_{1-\alpha}(\lambda) := \hat{\lambda} - t^*_{\alpha,\text{df}-1} \times \widehat{\text{se}}(\hat{\lambda}), \quad \text{UL}_{1-\alpha}(\lambda) := \hat{\lambda} + t^*_{\alpha,\text{df}-1} \times \widehat{\text{se}}(\hat{\lambda}), \tag{10}$$

Considering the direction of the bias that further reduces the lower limit, as well as the direction that further increases the upper limit, Equations 7 and 8 imply that both quantities can be written as a function of the restricted estimates and a new multiplier

$$\text{LL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\text{res}} - t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}), \quad \text{UL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\text{res}} + t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2} \times \widehat{\text{se}}(\hat{\lambda}_{\text{res}}) \tag{11}$$

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

$$t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2} := \text{SEF}\sqrt{\text{df}/(\text{df}-1)} \times t^*_{\alpha,\text{df}-1} + \text{BF}\sqrt{\text{df}}. \tag{12}$$

As the subscript $\boldsymbol{R}^2 = \{R^2_{Y\sim W|Z,\boldsymbol{X}}, R^2_{Z\sim W|\boldsymbol{X}}\}$ conveys, $t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ depends on both sensitivity parameters. Notably, this correction does not depend on the data (but for the degrees of freedom). 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 $\text{SEF} = 1$ and $\text{BF} = R^2/\sqrt{1-R^2}$ resulting in a very simple correction formula,

$$t^{\dagger}_{\alpha,\text{df}-1,R^2,R^2} \approx t^*_{\alpha,\text{df}-1} + \frac{R^2}{\sqrt{1-R^2}}\sqrt{\text{df}}, \tag{13}$$

where we employ the approximation $\sqrt{\text{df}/(\text{df}-1)} \approx 1$. Table 1 shows the adjusted critical values for this case, considering different strengths of the omitted variable and various sample sizes.

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

Table 1: Bias-adjusted critical values, $t^{\dagger}_{\alpha,\text{df}-1,R^2,R^2}$, 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; $\alpha = 5\%$.

Tests using these new critical values thus account both for sampling uncertainty and residual biases with the postulated strength. Note how $t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ *increases* the larger the sample size. This behaviour is simply a consequence of the well-known, but often overlooked fact that in large samples any signal will eventually be detected, 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.

3.3. *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 exactly how the inclusion of W with such strength would change inference regarding the parameter of interest. 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 $R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}}$. Writing $t^{\dagger}_{\alpha,\text{df}\,-1,\boldsymbol{R}^2}$ as a function of the sensitivity parameters $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$, we then solve the maximization problem

$$\max_{R^2_{Y\sim W|Z,\boldsymbol{X}},R^2_{Z\sim W|\boldsymbol{X}}} \quad t^{\dagger}_{\alpha,\text{df}\,-1,\boldsymbol{R}^2} \quad \text{s.t.} \quad R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}},\ R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \quad (14)$$

Denoting the solution to the optimization problem in expression (14) 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

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

The interval composed of such limits,

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

retrieves all confidence intervals for λ that are compatible with an omitted variable with such strengths. If the confidence interval adjusting for W has nominal coverage, and if the true sample partial R^2 of W lies within the posited bounds (note that here the judgment is made at the *sample* level), then it immediately follows that $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$ is also a confidence interval with at least $1-\alpha$ coverage.

3.4. *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 interval $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$. This framework facilitates extending these metrics to other contexts, in particular to the IV setting in Section 4.

The extreme robustness value

Our first inverse question is: 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)? 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 14 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,

$$\text{XRV}_{q^*,\alpha}(\lambda) := \inf \left\{ \text{XRV};\ (1 - q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\lambda) \right\} \tag{15}$$

The solution to this problem gives,

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

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 $f^*_{\alpha,\text{df}-1} := t^*_{\alpha,\text{df}-1}/\sqrt{\text{df}-1}$).[9] 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 $\text{XRV}_{q^*=1,\alpha=1}(\lambda) = \frac{f^2_{Y\sim Z|\boldsymbol{X}}}{1+f^2_{Y\sim Z|\boldsymbol{X}}} = R^2_{Y\sim Z|\boldsymbol{X}}$. 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.

The robustness value

An alternative measure of robustness of the OLS estimate is to consider the minimal 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, $\text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\lambda)$, 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

$$\text{RV}_{q^*,\alpha}(\lambda) := \inf \left\{ \text{RV};\ (1 - q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\lambda) \right\} \tag{16}$$

We then have that,

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

Where $f_{q^*,\alpha}(\lambda) := q^*|f_{Y\sim Z|\boldsymbol{X}}| - f^*_{\alpha,\text{df}-1}$. 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. In the second case, both associations of W reach the bound. The last case is an interior point solution—when the constraint on the partial R^2 with the outcome is not binding, the RV reduces to the XRV.

3.5. *Bounding the plausible strength of omitted variables*

One final 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

$$k_Z := \frac{R^2_{Z\sim W|\boldsymbol{X}_{-j}}}{R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{R^2_{Y\sim X_j|Z\boldsymbol{X}_{-j}}}. \tag{17}$$

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

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. Cinelli and Hazlett (2020) derived bounds considering the part of W not linearly explained by $\boldsymbol{X}$. These are particularly useful when contemplating X_j and W both *confounders* of Z (violations of the ignorability of the instrument). In the IV setting, however, W and X_j may be *side-effects* of Z, instead of causes of Z. In such cases, it may be more natural to reason about the orthogonality of $\boldsymbol{X}$ and W after conditioning on Z. Therefore, here we additionally provide bounds under the condition $R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0$. We then have that

$$R^2_{Z\sim W|\boldsymbol{X}} \leq \eta f^2_{Z\sim X_j|\boldsymbol{X}_{-j}}, \qquad R^2_{Y\sim W|Z,\boldsymbol{X}} = k_Y f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} \tag{18}$$

where η is a multiplier function of k_y, k_Z and $R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}$. These results allow investigators to leverage knowledge of *relative importance* of variables (Kruskal and Majors, 1989) when making plausibility judgments regarding sensitivity parameters.

4. An OVB 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 the sensitivity analysis of the IV estimate itself within the Anderson-Rubin approach.

4.1. *Sensitivity analysis of the reduced form and first stage*

The critical examination of the first stage and the reduced form plays an important role for supporting the causal story behind a particular instrumental variable. Researchers are thus advised to report and interpret the first stage and the reduced form by, e.g., assessing whether those results are consistent with theory and the postulated mechanisms that justify the choice of instrument (Angrist and Krueger, 2001; Angrist and Pischke, 2009; Imbens, 2014). While investigating these separate regressions, all sensitivity analysis results discussed in the previous section can be readily deployed. Fortunately, such sensitivity analyses also answer many pivotal questions regarding the IV estimate itself. First, 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. Second, the sensitivity of the first stage (to confounding that could change its sign) reveals whether the IV estimate could be arbitrarily large in either direction. We now formalize these claims.

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

Recall that all IV estimators under consideration are equivalent, equal to the ratio of the reduced-form and the first-stage regression coefficients, $\hat{\tau} := \hat{\lambda}/\hat{\theta}$. This simple algebraic fact leads to 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 $\hat{\tau}$. Second, if we cannot rule out confounders or side-effects able to *change the sign* of the first stage, we cannot rule out that the IV point estimate $\hat{\tau}$ could be *arbitrarily large* in either direction. 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.

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

Consider now the IV estimand $\tau = \lambda/\theta$ (i.e, the population parameter). 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 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, namely, $\hat{\phi}_{\tau_0} = \hat{\lambda} - \tau_0\hat{\theta}$ and, $\widehat{\text{se}}(\hat{\phi}_{\tau_0}) = \sqrt{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2\widehat{\text{var}}(\hat{\theta}) - 2\tau_0\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})}$. 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,

$$\underbrace{\left(\hat{\theta}^2 - \widehat{\text{var}}(\hat{\theta})t^{*2}_{\alpha,\text{df}-1}\right)}_{a}\tau_0^2 + \underbrace{2\left(\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})t^{*2}_{\alpha,\text{df}-1} - \hat{\lambda}\hat{\theta}\right)}_{b}\tau_0 + \underbrace{\left(\hat{\lambda}^2 - \widehat{\text{var}}(\hat{\lambda})t^{*2}_{\alpha,\text{df}-1}\right)}_{c} \leq 0 \quad (19)$$

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. Also note that arbitrarily large values for τ_0 will satisfy the inequality in Equation 19 if, and only if, $a < 0$, meaning that we cannot reject the null hypothesis $H_0 : \theta = 0$ in the first-stage regression. 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. 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.

It is not uncommon for frequentist statistical tests to lead to logically incoherent decisions (Schervish, 1996). 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 in the Anderson-Rubin approach*

We now build a complete set of sensitivity tools for IV within the Anderson-Rubin approach.

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 given by the test of 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 IV. In detail, a sensitivity analysis for the null hypothesis that the IV estimate τ equals τ_0 can be performed by: (i) constructing $Y_{\tau_0} = Y - \tau_0 D$ under the null value $H_0 : \tau = \tau_0$; (ii) running the OLS model $Y_{\tau_0} = \hat{\phi}_{\text{res},\tau_0} Z + \boldsymbol{X}\hat{\beta}_{\text{res},\tau_0} + \hat{\varepsilon}_{\tau_0,\text{res}}$; and (iii) performing regular OLS sensitivity analysis for the null $H_0 : \phi_{\tau_0} = 0$. This tells 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 $H_0 : \tau = \tau_0$, as well as the minimal strength of $\mathbf{R}^2$ required to not reject the null $H_0 : \tau = \tau_0$, as given by the RV or XRV.

Compatible inferences given bounds on partial R^2

More broadly, analysts can recover the set of inferences compatible with plausibility judgments on the maximum strength of W. For a critical threshold $t^*_{\alpha,\text{df}-1}$, the confidence interval for τ in the Anderson-Rubin framework is given by $\text{CI}_{1-\alpha}(\tau) = \{\tau_0;\ t^2_{\phi_{\tau_0}} \leq t^{*2}_{\alpha,\text{df}-1}\}$. Thus, 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

$$\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) = \left\{\tau_0;\ t^2_{\hat{\phi}_{\text{res},\tau_0}} \leq \left(t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right\} \tag{20}$$

This interval can be found analytically using the same inequality as in Equation 19, now with the parameters of the restricted regression actually run, and $t^*_{\alpha,\text{df}-1}$ replaced by $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$. 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 $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$.

Here it is useful to discuss the possible shapes of $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ 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 $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2} = (-\infty, r_{\min}] \cup [r_{\max}, +\infty)$.[10]

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.

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

Extreme robustness values for IV.

The extreme robustness value $\mathrm{XRV}_{q^*,\alpha}(\tau)$ 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,

$$\mathrm{XRV}_{q^*,\alpha}(\tau) := \inf\left\{\mathrm{XRV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \mathrm{CI}^{\max}_{1-\alpha,1,\mathrm{XRV}}(\tau)\right\}. \tag{21}$$

It then follows immediately from Equation 20 that $\mathrm{XRV}_{q^*,\alpha}(\tau) = \mathrm{XRV}_{1,\alpha}(\phi_{\tau^*})$, where $\tau^* = (1-q^*)\hat{\tau}_{\text{res}}$. 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 $\mathrm{XRV}_{1,\alpha}(\tau) = \mathrm{XRV}_{1,\alpha}(\lambda)$. 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 logically concluded in the prior section.

The $\mathrm{XRV}_{q^*,\alpha}(\tau)$ 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 $\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ be $\mathrm{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ and $\mathrm{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ respectively. The extreme robustness value $\mathrm{XRV}_{\geq q^*,\alpha}(\tau)$ 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 cannot reject a change of $(100 \times q^*)\%$ *or worse* of the original IV estimate;

$$\mathrm{XRV}_{\geq q^*,\alpha}(\tau) := \inf\left\{\mathrm{XRV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \left[\mathrm{LL}^{\max}_{1-\alpha,1,\mathrm{XRV}}(\tau),\quad \mathrm{UL}^{\max}_{1-\alpha,1,\mathrm{XRV}}(\tau)\right]\right\} \tag{22}$$

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

$$\mathrm{XRV}_{\geq q^*,\alpha}(\tau) = \min\{\mathrm{XRV}_{1,\alpha}(\phi_{\tau^*}),\quad \mathrm{XRV}_{1,\alpha}(\theta)\} \tag{23}$$

corroborating our conclusion that, the robustness of IV estimates against biases as large or larger than a certain amount 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 $\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}$ varying simultaneously. That is,

$$\mathrm{RV}_{q^*,\alpha}(\tau) := \inf\left\{\mathrm{RV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \mathrm{CI}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\tau)\right\} \tag{24}$$

Again from Equation 20 we have that $\mathrm{RV}_{q^*,\alpha}(\tau) = \mathrm{RV}_{1,\alpha}(\phi_{\tau^*})$, which for the special case of $q^* = 1$ simplifies to $\mathrm{RV}_{1,\alpha}(\tau) = \mathrm{RV}_{1,\alpha}(\lambda)$, as before. We can also define the RV for not rejecting the null of a reduction of $(100 \times q^*)\%$ *or worse*

$$\mathrm{RV}_{\geq q^*,\alpha}(\tau) := \inf\left\{\mathrm{RV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \left[\mathrm{LL}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\tau),\quad \mathrm{UL}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\tau)\right]\right\} \tag{25}$$

By the same arguments articulated above, $\mathrm{RV}_{\geq q^*,\alpha}(\tau)$ 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

$$\mathrm{RV}_{\geq q^*,\alpha}(\tau) = \min\{\mathrm{RV}_{1,\alpha}(\phi_{\tau^*}),\quad \mathrm{RV}_{1,\alpha}(\theta)\} \tag{26}$$

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

Bounds on the strength of omitted variables

When testing a specific null hypothesis $H_0 : \tau = \tau_0$ in the AR regression, we have k_Z as in Section 3.5, and instead of k_Y we now have $k_{Y_{\tau_0}} := R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}_{-j}} / R^2_{Y_{\tau_0}\sim X_j|Z\boldsymbol{X}_{-j}}$. The plausibility judgment one is making here is that of how W is relative to observed covariates, under $H_0 : \tau = \tau_0$. Since the judgment is made under a specific null, the bounds will be different when testing different hypotheses. Therefore, it is useful to compute bounds under a slightly more *conservative* assumption. 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, i.e,

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

This has the useful property of providing a unique bound for any null hypothesis, and can be used to place bounds on the 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 of IV

In this section we return to our running example and show how these tools can be deployed to assess the robustness of those findings to violations of the IV assumptions. 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). Here we focus on the sensitivity of the IV estimate, separate analyses of the reduced form and first stage are provided in the Supplementary Materials.

5.1. *Minimal sensitivity reporting*

Outcome: *Earnings* (log)

Treatment	Estimate	$\mathrm{LL}_{1-\alpha}$	$\mathrm{UL}_{1-\alpha}$	t-value	$\mathrm{XRV}_{\geq q^*,\alpha}$	$\mathrm{RV}_{\geq q^*,\alpha}$
Education (years)	0.132	0.025	0.285	2.33	0.05%	0.67%

Bound (1x SMSA): $R^2_{Y_0\sim W|Z,\boldsymbol{X}} = 2\%$, $R^2_{W\sim Z|\boldsymbol{X}} = 0.6\%$, $t^{\dagger\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} = 2.55$

Note: $\mathrm{df} = 2994, \quad q^* = 1, \quad \alpha = 0.05$

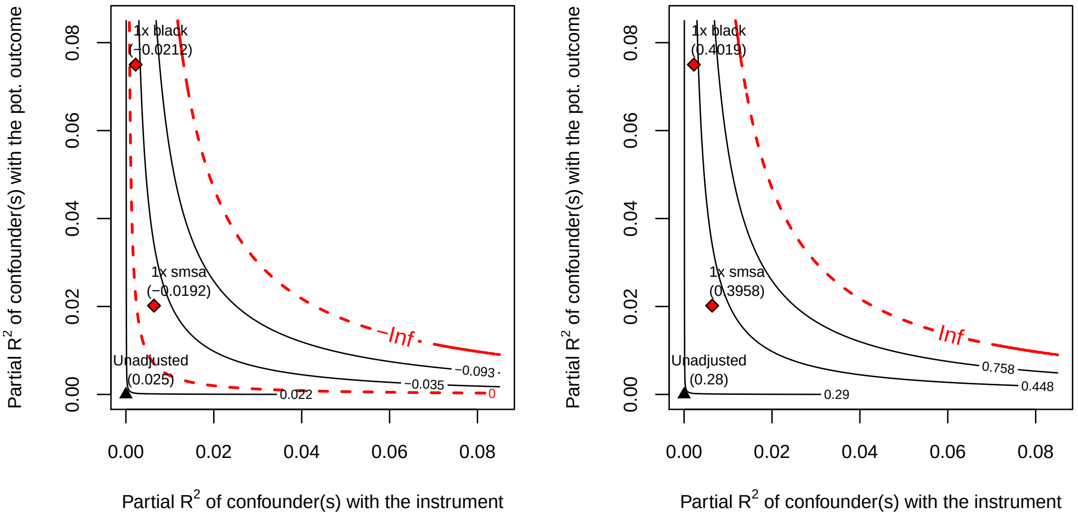
Table 2: Minimal sensitivity reporting of the IV estimate (Anderson-Rubin).

Table 2 shows our proposed minimal sensitivity reporting for IV estimates. It starts by replicating the usual statistics, such as the point estimate (0.132), as well as the lower and upper limits of the Anderson-Rubin confidence interval [0.025, 0.285] , and the t-value against the null hypothesis of zero effect (2.33). Next, we propose researchers report the extreme robustness value ($\mathrm{XRV}_{\geq q^*,\alpha} = 0.05\%$) and the robustness value ($\mathrm{RV}_{\geq q^*,\alpha} = 0.67\%$) required to bring the lower limit of the confidence interval to or beyond zero (or another meaningful threshold), at the 5% significance level. As derived in the previous section, the (extreme) robustness value of the IV estimate required to bring the lower limit of the confidence interval to zero or below is the minimum of the (extreme) robustness value of the reduced form and the (extreme) robustness value of the first stage. In our running example, the reduced form is more fragile, thus the sensitivity of the IV hinges critically on the sensitivity of the reduced form.

The RV reveals that confounders explaining 0.67% of the residual variation both of *proximity* and of (log) *Earnings* are already sufficient to make the IV estimate statistically insignificant. Further, the XRV shows that, if we are not willing to impose constraints on the partial R^2 of confounders with the outcome, they need only explain 0.05% of the residual variation instrument to "lose statistical significance." To aid users in making plausibility judgments, the note of the table provides bounds on 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, this immediately reveals that confounding as strong as *SMSA* (e.g. residual geographic confounding) is already sufficiently strong to be problematic.

5.2. *Sensitivity contours plots*

It will often be valuable to assess the sensitivity of the IV against hypothesis *other than zero.* To that end, investigators may wish to examine sensitivity contour plots showing the whole range of adjusted lower and upper limits of the AR confidence interval against various strengths of the omitted variables W. These contours are shown in Figure 1. Here the horizontal axis indicates the bounds on the partial R^2 of the confounder with the instrument, and the vertical axis indicates the bounds on $R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}$, i.e, the partial R^2 of the confounder with $Y_{\tau_0} := Y - \tau_0 D$ (the outcome after subtracting the "putative causal effect" of D). Under a constant treatment effects model, this has a simple interpretation—it is the untreated potential outcome. For simplicity, of exposition, we adopt this interpretation here. The contour lines show the worst lower (or upper) limit of the $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$, with omitted variables bounded by such strength. Red dashed lines shows a



(a) Sensitivity contours for the lower limit.

(b) Sensitivity contours for the upper limit.

Figure 1: Sensitivity contours of the lower and upper limits of AR confidence interval.

critical contour line of interest (such as zero) as well as the boundary beyond confidence intervals become unbounded. The red diamonds places bounds on strength of W as strong as *Black* (an indicator for race) and, again, *SMSA*, as per Section 4.2. As the plot reveals, both confounding as strong as *SMSA*, or as strong as *black*, could lead to an interval for the target parameter of $\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) = [-0.02, 0.40]$, which includes not only implausibly high values (40%), but also negative values (-2%), and is thus too wide for any meaningful conclusions. 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 call into question the strength of evidence provided by this IV study.

6. Discussion

Here we focused on the sensitivity of the "traditional" IV estimate, consisting of the ratio of two OLS regression coefficients. We chose to do so because this reflects current practices, and encompasses the vast majority of applied work. These tools can thus be immediately put to use to improve the robustness of current research, without requiring any additional assumptions, beyond those that already justified the IV estimate (including W) as the target of interest. Recent papers, however, have correctly 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 offers an interesting direction for future work.

Bibliography

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.

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

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.

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.

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

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

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.

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.

Felton, C. and Stewart, B. M. (2022). Handle with care: A sociologist's guide to causal inference with instrumental variables.

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

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

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 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. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

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.

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

Masten, M. A. and Poirier, A. (2021). Salvaging falsified instrumental variable models. *Econometrica*, 89(3):1449–1469.

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.*

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.

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. (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.

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.

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

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

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

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, $Y^{\perp X} := Y - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'Y$. 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),

$$\textbf{First stage:} \quad D = \hat{\theta} Z + \boldsymbol{X}\hat{\psi} + \boldsymbol{W}\hat{\delta} + \hat{\varepsilon}_d \tag{31}$$

$$\textbf{Reduced form:} \quad Y = \hat{\lambda} Z + \boldsymbol{X}\hat{\beta} + \boldsymbol{W}\hat{\gamma} + \hat{\varepsilon}_y \tag{32}$$

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 $\hat{\theta}$

$$\hat{\tau}_{\text{ILS}} := \frac{\hat{\lambda}}{\hat{\theta}} \tag{33}$$

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

$$\widehat{\text{var}}(\hat{\tau}_{\text{ILS}}) := \frac{1}{\hat{\theta}^2}\left(\widehat{\text{var}}(\hat{\lambda}) + \hat{\tau}_{\text{ILS}}^2\widehat{\text{var}}(\hat{\theta}) - 2\hat{\tau}_{\text{ILS}}\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})\right) \tag{34}$$

where, using the FWL formulation,

$$\widehat{\text{var}}(\hat{\lambda}) = \frac{\text{var}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1}, \qquad \widehat{\text{var}}(\hat{\theta}) = \frac{\text{var}(D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{35}$$

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

$$\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) = \frac{\text{cov}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}, D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{36}$$

is the estimated covariance of $\hat{\lambda}$ and $\hat{\theta}$. Here var$(\cdot)$ and 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 31 to produce fitted values for the treatment ($\widehat{D}$), then regressing the outcome on these fitted values,

$$\textbf{Second stage:} \quad Y = \hat{\tau}_{\text{2SLS}} \widehat{D} + \boldsymbol{X} \hat{\beta}_{\text{2SLS}} + \boldsymbol{W} \hat{\gamma}_{\text{2SLS}} + \hat{\varepsilon}_{\text{2SLS}} \tag{37}$$

The 2SLS estimate corresponds to the coefficient $\hat{\tau}_{\text{2SLS}}$ in Equation 37, called the "second-stage" regression.

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

$$\hat{\tau}_{\text{2SLS}} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} \tag{38}$$

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

$$\begin{aligned} \hat{\tau}_{\text{2SLS}} &= \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} & (39) \\ &= \frac{\hat{\theta} \times \text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, Z^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\hat{\theta}^2 \times \text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\hat{\lambda}}{\hat{\theta}} & (40) \end{aligned}$$

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

$$\widehat{\text{var}}(\hat{\tau}_{\text{2SLS}}) := \frac{\text{var}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}} - \hat{\tau} D^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{41}$$

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 41 can be expanded to

$$\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}}) = \text{var}(\hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}}) = \hat{\theta}^2 \, \text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}}) \tag{42}$$

Finally, the numerator can be written as,

$$\begin{aligned}
\operatorname{var}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \hat{\tau} D^{\perp \boldsymbol{X},\boldsymbol{W}}) &= \operatorname{var}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \hat{\tau}(\hat{\theta} Z^{\boldsymbol{X},\boldsymbol{W}} + D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})) && (43)\\
&= \operatorname{var}((Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \hat{\lambda} Z^{\boldsymbol{X},\boldsymbol{W}}) - \hat{\tau} D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) && (44)\\
&= \operatorname{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}} - \hat{\tau} D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) && (45)\\
&= \operatorname{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) + \hat{\tau}^2 \operatorname{var}(D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) - 2\hat{\tau} \operatorname{cov}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}, D^{\perp Z,\boldsymbol{X},\boldsymbol{W}}) && (46)
\end{aligned}$$

Plugging in Equations 46 and 42 back in Equation 41, then using Equations 35 and 36 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

$$\textbf{Anderson-Rubin:} \quad Y_{\tau_0} = \hat{\phi}_{\tau_0} Z + \boldsymbol{X}\hat{\beta}_{\tau_0} + \boldsymbol{W}\hat{\gamma}_{\tau_0} + \hat{\varepsilon}_{\tau_0} \tag{47}$$

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,

$$\hat{\tau}_{\text{AR}} = \{\tau_0;\ \hat{\phi}_{\tau_0} = 0\} \tag{48}$$

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,

$$\begin{aligned}
\hat{\phi}_{\tau_0} &= \frac{\operatorname{cov}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \tau_0 D^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\operatorname{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} && (49)\\
&= \frac{\operatorname{cov}(Y^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\operatorname{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} - \tau_0 \frac{\operatorname{cov}(D^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\operatorname{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} && (50)\\
&= \hat{\lambda} - \tau_0 \hat{\theta} && (51)
\end{aligned}$$

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

$$\hat{\tau}_{AR} = \frac{\hat{\lambda}}{\hat{\theta}} \tag{52}$$

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

$$\text{CI}_{1-\alpha}(\tau) = \{\tau_0; t^2_{\hat{\phi}_{\tau_0}} \leq t^{*2}_{\alpha,\text{df}}\} \tag{53}$$

This confidence interval can be obtained analytically as functions of the estimates of the first-stage and reduced form regressions. As shown in Equation 51, $\hat{\phi}_{\tau_0}$ can be written as the linear combination

$$\hat{\phi}_{\tau_0} = \hat{\lambda} - \tau_0\hat{\theta} \tag{54}$$

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

$$\begin{aligned}
\widehat{\text{var}}(\hat{\phi}_{\tau_0}) &= \frac{\text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}} - \tau_0 D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} \times \text{df}^{-1} && (55)\\
&= \left(\frac{\text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} + \tau_0^2\frac{\text{var}(D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} - 2\tau_0\frac{\text{cov}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}, D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})}\right) \times \text{df}^{-1} && (56)\\
&= \widehat{\text{var}}(\hat{\lambda}) + \tau_0^2\widehat{\text{var}}(\hat{\theta}) - 2\tau_0\widehat{\text{cov}}(\hat{\lambda},\hat{\theta}) && (57)
\end{aligned}$$

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

$$t_{\hat{\phi}_{\tau_0}} = \frac{\hat{\lambda} - \tau_0\hat{\theta}}{\sqrt{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2\widehat{\text{var}}(\hat{\theta}) - 2\tau_0\widehat{\text{cov}}(\hat{\lambda},\hat{\theta})}} \tag{58}$$

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

$$\frac{(\hat{\lambda} - \tau_0\hat{\theta})^2}{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2\widehat{\text{var}}(\hat{\theta}) - 2\tau_0\widehat{\text{cov}}(\hat{\lambda},\hat{\theta})} \leq t^{*2}_{\alpha,\text{df}} \tag{59}$$

First, note that the empty set is not possible here. If we pick $\tau_0 = \hat{\tau}_{\text{AR}}$, then the numerator in Equation 59 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

$$\underbrace{\left(\hat{\theta}^2 - \widehat{\text{var}}(\hat{\theta}) \times t^{*2}_{\alpha,\text{df}}\right)}_{a}\tau_0^2 + \underbrace{2\left(\widehat{\text{cov}}(\hat{\lambda},\hat{\theta}) \times t^{*2}_{\alpha,\text{df}} - \hat{\lambda}\hat{\theta}\right)}_{b}\tau_0 + \underbrace{\left(\hat{\lambda}^2 - \widehat{\text{var}}(\hat{\lambda}) \times t^{*2}_{\alpha,\text{df}}\right)}_{c} \leq 0 \tag{60}$$

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 57, resulting in a test statistic equal to Equation 58, and thus numerically identical to the AR approach.

B 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 8 and 9 imply

$$\begin{aligned}
\mathrm{LL}_{1-\alpha}(\lambda) &:= \hat{\lambda} - t^{*}_{\alpha,\mathrm{df}-1} \times \widehat{\mathrm{se}}(\hat{\lambda}) && (61)\\
&= \hat{\lambda}_{\mathrm{res}} - \mathrm{BF}\,\sqrt{\mathrm{df}} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) - t^{*}_{\alpha,\mathrm{df}-1} \times \mathrm{SEF}\,\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) && (62)\\
&= \hat{\lambda}_{\mathrm{res}} - \left(\mathrm{SEF}\,\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^{*}_{\alpha,\mathrm{df}-1} + \mathrm{BF}\,\sqrt{\mathrm{df}}\right) \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) && (63)
\end{aligned}$$

Similarly, now let $\mathrm{UL}_{1-\alpha}(\lambda)$ 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

$$\mathrm{UL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\mathrm{res}} + \left(\mathrm{SEF}\,\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^{*}_{\alpha,\mathrm{df}-1} + \mathrm{BF}\,\sqrt{\mathrm{df}}\right) \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) \tag{64}$$

Note that, in both Equations 63 and 64, 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*,

$$t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} := \mathrm{SEF}\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^{*}_{\alpha,\mathrm{df}-1} + \mathrm{BF}\sqrt{\mathrm{df}}. \tag{65}$$

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

$$\max_{R^2_{Y\sim W|Z,\boldsymbol{X}},\,R^2_{Z\sim W|\boldsymbol{X}}} \; t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \quad \text{s.t.} \quad R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}, \quad R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \tag{66}$$

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

$$\begin{aligned}
\frac{\partial(t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}/\sqrt{\mathrm{df}})}{\partial R^2_{Z\sim W|\boldsymbol{X}}} &= \frac{\partial\,\mathrm{BF}}{\partial R^2_{Z\sim W|\boldsymbol{X}}} + f^{*}_{\alpha,\mathrm{df}-1} \times \frac{\partial\,\mathrm{SEF}}{\partial R^2_{Z\sim W|\boldsymbol{X}}} && (67)\\
&= \frac{(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} + f^{*}_{\alpha,\mathrm{df}-1}\frac{(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}} && (68)\\
&= \frac{(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2} + f^{*}_{\alpha,\mathrm{df}-1}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \geq 0 && (69)
\end{aligned}$$

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,

$$\frac{\partial(t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}/\sqrt{\mathrm{df}})}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} = \frac{\partial\,\mathrm{BF}}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} + f^*_{\alpha,\mathrm{df}-1} \times \frac{\partial\,\mathrm{SEF}}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} \tag{70}$$

$$= \frac{(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}} + \frac{-f^*_{\alpha,\mathrm{df}-1}}{2(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \tag{71}$$

$$= \frac{(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2} - f^*_{\alpha,\mathrm{df}-1}(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \tag{72}$$

That is, due to the variance reduction factor of the omitted variable (VRF in Equation 9), 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 72 must be positive when evaluated at the solution. Rearranging and squaring, we see that this happens when

$$R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \geq f^{*2}_{\alpha,\mathrm{df}-1} \times f^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}} \tag{73}$$

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 73 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 72 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,

$$R^{2*}_{Y\sim W|Z,\boldsymbol{X}} = \frac{R^{2\,\max}_{Z\sim W|\boldsymbol{X}}}{f^{*2}_{\alpha,\mathrm{df}-1} + R^{2\,\max}_{Z\sim W|\boldsymbol{X}}} \tag{74}$$

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

$$\mathrm{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) = \hat{\lambda}_{\mathrm{res}} - t^{\dagger\,\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}), \quad \mathrm{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2} = \hat{\lambda}_{\mathrm{res}} + t^{\dagger\,\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) \tag{75}$$

And interval composed of such limits

$$\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) = \left[\mathrm{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda), \quad \mathrm{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)\right] \tag{76}$$

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* $\mathrm{XRV}_{q^*,\alpha}(\lambda)$ 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:

$$\text{XRV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{XRV};\ (1 - q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\lambda)\right\} \tag{77}$$

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 73 fails (since $f^2_{Y\sim W|Z,\boldsymbol{X}}$ is unbounded). Therefore, XRV will be given by an interior point solution. Using Equation 74 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

$$\text{XRV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if } \ f_{q^*}(\lambda) \leq f^*_{\alpha,\text{df}-1} \\ \dfrac{f^2_{q^*}(\lambda) - f^{*2}_{\alpha,\text{df}-1}}{1 + f^2_{q^*}(\lambda)}, & \text{otherwise.} \end{cases} \tag{78}$$

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

$$\text{RV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{RV};\ (1 - q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,\text{RV},\text{RV}}(\lambda)\right\} \tag{79}$$

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,

$$\text{RV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if } \ f_{q^*}(\lambda) \leq f^*_{\alpha,\text{df}-1} \\ \dfrac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right), & \text{if } \ f^*_{\alpha,\text{df}-1} < f_{q^*}(\lambda) < f^{*-1}_{\alpha,\text{df}-1} \\ \text{XRV}_{q^*,\alpha}(\lambda), & \text{otherwise.} \end{cases} \tag{80}$$

The condition $f_{q*}(\lambda) < f^{*-1}_{\alpha,\text{df}-1}$, stems from the fact that the XRV solution cannot satisfy Equation 73. 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,

$$
\begin{aligned}
\mathrm{RV}_{q^*,\alpha}(\lambda) &= \frac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right) \qquad (81)\\
&= \frac{1}{2}\left(\sqrt{(f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^4 + 4(f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^2} - (f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^2\right) \qquad (82)\\
&> \frac{1}{2}\left(\sqrt{(f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^4 + 4(f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^2} - (f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^2\right) \qquad (83)\\
&= \frac{1}{2}\left(\sqrt{\left(\frac{f^2_{q}(\lambda) - 1}{f_{q^*}(\lambda)}\right)^4 + 4\left(\frac{f^2_{q^*}(\lambda) - 1}{f_{q^*}(\lambda)}\right)^2} - \left(\frac{f^2_{q^*}(\lambda) - 1}{f_{q^*}(\lambda)}\right)^2\right) \qquad (84)\\
&= \left(\frac{1}{2}\right)\left(\frac{f^2_{q^*}(\lambda) - 1}{f^2_{q^*}(\lambda)}\right)\left(\sqrt{(f^2_{q}(\lambda) - 1)^2 + 4f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) \qquad (85)\\
&= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(\sqrt{f^4_{q}(\lambda) + 1 - 2f^2_{q^*}(\lambda) + 4f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) \qquad (86)\\
&= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(\sqrt{f^4_{q}(\lambda) + 1 + 2f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) \qquad (87)\\
&= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(f^2_{q^*}(\lambda) + 1 - f^2_{q^*}(\lambda) + 1\right) \qquad (88)\\
&= 1 - 1/f^2_{q^*}(\lambda) \qquad (89)
\end{aligned}
$$

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

$$k_Z := \frac{R^2_{Z\sim W|\boldsymbol{X}_{-j}}}{R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{R^2_{Y\sim X_j|Z\boldsymbol{X}_{-j}}}. \qquad (90)$$

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 straightforward. 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

$$\left|R_{Y\sim W|Z,\boldsymbol{X}}\right| = \left|\frac{R_{Y\sim W|Z,\boldsymbol{X}_{-j}} - R_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} R_{W\sim X_j|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}}}}\right| \tag{91}$$

$$= \left|\frac{R_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}}\right| \tag{92}$$

$$= \left|\frac{\sqrt{k_Y} R_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}}\right| \tag{93}$$

$$= \sqrt{k_Y}\left|f_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}\right| \tag{94}$$

Hence,

$$R^2_{Y\sim W|Z,\boldsymbol{X}} = k_Y \times f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}} \tag{95}$$

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

$$R_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0 \implies \frac{R_{W\sim X_j|\boldsymbol{X}_{-j}} - R_{W\sim Z|\boldsymbol{X}_{-j}} R_{X_j\sim Z|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{W\sim Z|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{X_j\sim Z|\boldsymbol{X}_{-j}}}} = 0 \tag{96}$$

$$\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} - R_{W\sim Z|\boldsymbol{X}_{-j}} R_{X_j\sim Z|\boldsymbol{X}_{-j}} = 0 \tag{97}$$

$$\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} = R_{W\sim Z|\boldsymbol{X}_{-j}} R_{X_j\sim Z|\boldsymbol{X}_{-j}} \tag{98}$$

$$\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} = R_{Z\sim W|\boldsymbol{X}_{-j}} R_{Z\sim X_j|\boldsymbol{X}_{-j}} \tag{99}$$

$$\implies |R_{W\sim X_j|\boldsymbol{X}_{-j}}| = \sqrt{k_Z} R^2_{Z\sim X_j|\boldsymbol{X}_{-j}} \tag{100}$$

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

$$\left|R_{Z\sim W|\boldsymbol{X}}\right| = \left|\frac{R_{Z\sim W|\boldsymbol{X}_{-j}} - R_{Z\sim X_j|\boldsymbol{X}_{-j}} R_{W\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|\boldsymbol{X}_{-j}}}}\right| \tag{101}$$

$$\leq \frac{\left|R_{Z\sim W|\boldsymbol{X}_{-j}}\right| + \left|R_{Z\sim X_j|\boldsymbol{X}_{-j}} R_{W\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|\boldsymbol{X}_{-j}}}} \tag{102}$$

$$= \frac{\left|\sqrt{k_Z} R_{Z\sim X_j|\boldsymbol{X}_{-j}}\right| + \left|\sqrt{k_Z} R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}} \tag{103}$$

$$= \left(\frac{\sqrt{k_Z} + \left|R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right) \times \left(\frac{\left|R_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right) \tag{104}$$

$$= \eta' |f_{Z\sim X_j|\boldsymbol{X}_{-j}}| \tag{105}$$

Hence we have that

$$R^2_{Z\sim W|\boldsymbol{X}} \leq {\eta'}^2 f^2_{Z\sim X_j|\boldsymbol{X}_{-j}} \tag{106}$$

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 $\boldsymbol{X}$):

$$Y_i = \tau D_i + \eta Z_i + \varepsilon_i \tag{107}$$

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,

$$\tau = \frac{\lambda - \eta}{\theta}. \tag{108}$$

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 conclusion 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 107, 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

$$Y_i = \tau D_i + \eta Z_i + \gamma W + \varepsilon_i \tag{109}$$

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 Results for the Empirical Example

F.1 Minimal reporting and sensitivity contours of the reduced form

Table 5 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-error and t-value, we recommend that researchers also report the: (i) partial R^2 of the instrument with the outcome ($R^2_{Y\sim Z|\boldsymbol{X}} = 0.18\%$), 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, $\alpha = 0.05$).

Outcome: *Earnings* (log)

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

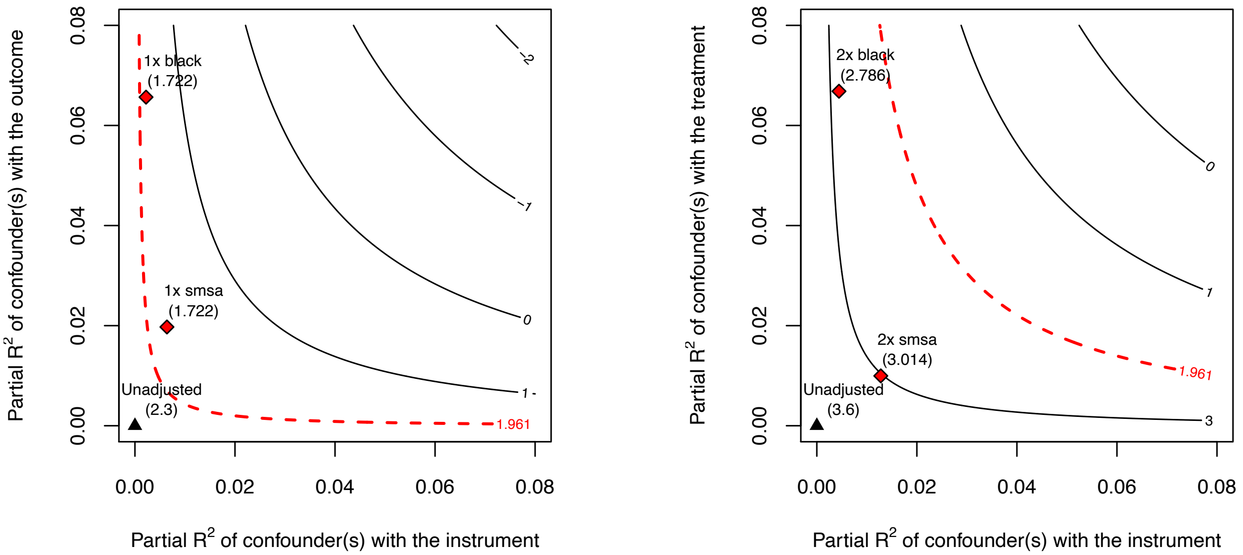
Bound (1x SMSA): $R^2_{Y\sim W|Z,\boldsymbol{X}} = 2\%$, $R^2_{W\sim Z|\boldsymbol{X}} = 0.6\%$, $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2} = 2.55$

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

Table 5: Minimal sensitivity reporting of the reduced-form regression.

In our running example, the RV reveals that confounders explaining 0.67% of the residual variation both of *proximity* and of (log) *Earnings* are already sufficient to make the reduced-form estimate statistically insignificant. Further, the XRV and the $R^2_{Y\sim Z|\boldsymbol{X}}$ show that, if we are not willing to impose constraints on the partial R^2 of confounders with the outcome, they need only explain 0.05% of the residual variation instrument to "lose significance," or 0.18% to fully eliminating the point estimate. To aid users in making plausibility judgments, the note of Table 5 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, this immediately reveals that confounding as strong as *SMSA* (e.g. residual geographic confounding) is sufficiently strong to be problematic.

Beyond the results of Table 5, 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 3a. 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 3: Sensitivity contour plots of the reduced form and first stage.

F.2 Minimal reporting and sensitivity contours of the first stage

Table 6 performs the same sensitivity exercises 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. This is 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 $\text{RV}_{q^*,\alpha} = 3.02\%$). 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.

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 $= 2994$, $q^* = 1$, $\alpha = 0.05$						

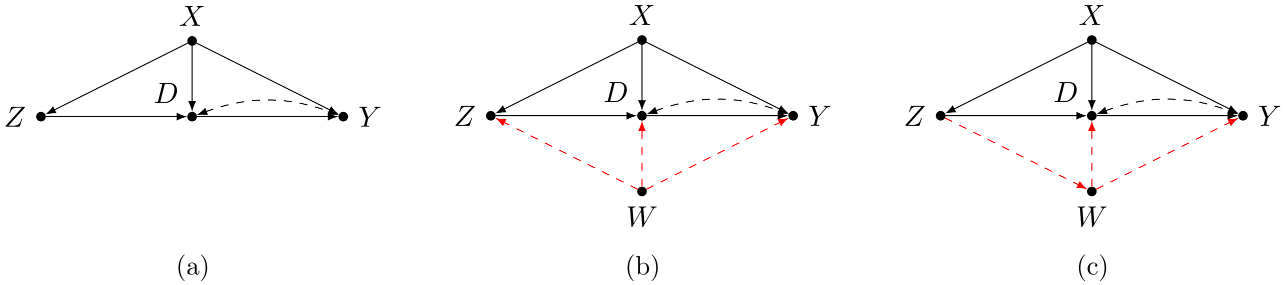
Table 6: Minimal sensitivity reporting of the first-stage regression.

Figure 3b 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*.

G Supplementary Tables and Figures

	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 7: 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 Indirect Least Squares/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.



(a)

(b)

(c)

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.