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 and Data Science, University of California, Los Angeles.
chazlett@ucla.edu

Abstract

We develop an omitted variable bias framework for sensitivity analysis of instrumental variable 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 instrumental variables, describing the minimum strength that omitted variables need to have to change the conclusions of a study. Next we provide visual displays that fully characterize the sensitivity of point estimates and confidence intervals to violations of the standard instrumental variable 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 instrumental variable 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 proximity to college as an instrumental variable 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, instrumental variable 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. In recent 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). 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. Although in certain cases such 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 instrumental variable 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 catalogued several plausible violations of the exclusion restriction for such instruments (e.g Gallen, 2020; Mellon, 2020). More worrisome, if traditional 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 (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 a study.

While a number of sensitivity analyses for instrumental variables have been proposed (e.g., Small, 2007; Small and Rosenbaum, 2008; Conley et al., 2012; Wang et al., 2018; Cinelli et al., 2019; Masten and Poirier, 2021), they have rarely been employed in practice. For example, in economics, only 1 out of 27 papers using instrumental variables, published in the *American Economic Review* in 2020, performed formal sensitivity analysis to unobserved variables. In political science, this number was 1 out of 12 papers across the top three general interest journals in 2019 (Cinelli and Hazlett, 2020). In Sociology, none of the 34 papers published between 2004 and 2022 in the *American Journal of Sociology* and the *American Sociology Review* did so (Felton and Stewart, 2022). Note that sensitivity to *unobserved* variables is distinct from (though related to) sensitivity to analytical choices of the investigator, which is more commonly found in the applied literature; these include, for example, sensitivity to different estimators, the presence of outliers, effect heterogeneity, or covariate selection (see, Blundell et al., 2001; Belzil and Hansen, 2002; Jaeger and Parys, 2009 for examples applied to returns to education).

We suggest several reasons for this slow uptake. First, the traditional approach for sensitivity analysis of instrumental variable estimates has focused on parameterizing violations of the IV assumptions with a single parameter summarizing the overall 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 traditional approaches). Second, while users of IV methods are instructed to routinely report quantities to diagnose certain inferential problems such as “weak instruments” (e.g. the F-statistic, Stock and Yogo, 2002), we lack sensitivity statistics that can quickly communicate how robust a study is to violations in the form of omitted confounders or side-effects of the instrument. Additionally, 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 (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) naturally handles violations due to multiple side-effects and confounders, possibly acting non-linearly; (ii) is well suited for routine reporting; and (iii) exploits expert knowledge to bound sensitivity parameters. (Here we focus on the just-identified case of one treatment and one instrument for two reasons. First, examining violations of identification assumptions is already challenging enough with a single instrument (Angrist and Pischke, 2009). Second, most applied work falls into this category: for instance, Young (2022) finds that 80% of instrumental variable regressions in the *American Economic Review* and 15 other journals of the *American Economic Association* used a single instrument. Even in multiple-instrument studies, it is not uncommon for researchers to report and give special focus to their best single instrument.)

Specifically, we introduce two main sensitivity statistics for instrumental variable estimates: (i) the *robustness value* 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 hypothesis tests, as well as lower or upper limits of confidence intervals. 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 observed variables.

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 observed 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 regression. Applied researchers 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. All proofs and details can be found in the Supplementary Material. Open-source software for `R` implements the methods discussed in this paper: `https://github.com/carloscinelli/iv.sensemakr`.

2. Background and running example

2.1. *Ordinary least squares and the omitted variable bias 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.

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 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 (if the treatment effect is heterogeneous, this may affect the causal interpretation of τ_{OLS}, see, e.g. Angrist and Pischke, 2009). 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 omitted variable bias, likely overestimating the true returns to schooling.

2.2. *Instrumental variables as a solution to the omitted variable bias 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 $\boldsymbol{X}$, whether one lives near a college is not itself confounded with earnings, nor does proximity to college affect earnings apart from its effect on years of education. If we believe such assumptions hold it is possible to recover a valid estimate of the (weigthed average of local) average treatment effect(s) of *Education* on *Earnings* by simply taking the ratio of two OLS coefficients, one measuring the effect of *Proximity* on *Earnings*, and another measuring the effect of *Proximity* on *Education*, as in the two regression 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* (1) and the *reduced form* (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, $\hat{\tau}_{\text{res}} := \hat{\lambda}_{\text{res}}/\hat{\theta}_{\text{res}} \approx 0.042/0.319 \approx 0.132$. 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%. (Conditions that allow a causal interpretation of the traditional IV estimand are extensively discussed elsewhere and will not be reviewed here, see Angrist et al. (1996); Angrist and Pischke (2009); Swanson et al. (2018); Słoczyński (2020) and Blandhol et al. (2022). See also Section 6.)

The ratio $\hat{\lambda}_{\text{res}}/\hat{\theta}_{\text{res}}$ is sometimes called the *indirect least squares* estimator. Inference in this framework is usually performed using the delta-method. A closely related approach is denoted by *two-stage least squares*, 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 two-stage least squares and indirect least squares yield numerically identical estimates and standard errors.

2.3. *Anderson-Rubin regression, Fieller's theorem and weak instruments*

Inference using the previous methods may prove unreliable when the first-stage coefficient is too close to zero relative to the sampling variability of its estimator. This is known as the "weak instrument" problem. The Anderson-Rubin regression (Anderson and Rubin, 1949) provides one approach to constructing confidence intervals with correct coverage, regardless of the strength of the first stage. Additionally, it also yields the uniformly most powerful unbiased test under this setup (Moreira, 2009).

The approach 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 regression

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

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). 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 also numerically identical to the indirect least squares and two-stage least squares estimates.

The literature on weak instruments is extensive (see, e.g., Nelson and Startz, 1990; Staiger and Stock, 1994; Kleibergen, 2002; Moreira, 2003, 2009; Andrews et al., 2019), and users are routinely advised to report diagnostic measures (e.g. the F-statistic of the first stage). It is important to note, however, that sensitivity to unobserved confounders or side-effects is distinct from issues posed by weak instruments. In particular, the latter depends on sample size, whereas the former does not. Thus, instruments deemed "strong" by conventional statistics may still be fragile in the face of unobserved variables—see Remark 4.

2.4. *The instrumental variable estimate may still suffer from omitted variable bias*

The previous instrumental variable 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

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

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

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

where $\boldsymbol{W}$ stands for all unobserved factors necessary to make *Proximity* a valid instrument for the effect of *Education* on *Earnings*. 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).

2.5. *Problem statement*

Our task is to characterize how instrumental variable estimates, as given by the OLS regressions in (4)-(6), would have changed due to the inclusion of omitted variables $\boldsymbol{W}$. As such, we should be able to leverage sensitivity analysis tools for OLS to examine the sensitivity of IV. The next section thus extends and refines several results for the sensitivity analysis of arbitrary OLS estimates. These results are not only useful on their own right, but, importantly, they will later be applied to the development of a suite sensitivity analysis tools for instrumental variables in Section 4. Finally, throughout the paper, we impose the following regularity condition.

Assumption 1 (Full Rank). The matrices of independent variables in (4)-(6) have full rank.

This ensures all relevant quantities discussed below are finite.

3. EXTENSIONS TO THE OMITTED VARIABLE BIAS FRAMEWORK FOR OLS

3.1. *Preliminaries*

We start by briefly establishing key ideas, formulae, and notation from prior work (Cinelli and Hazlett, 2020). For concreteness, in this section we discuss the omitted variable bias framework in the context of the reduced-form regression. Readers should keep in mind, however, that all results presented here hold for *arbitrary OLS estimates*—including, but not limited to, the first stage and the Anderson-Rubin regression. The logical implications of the sensitivity of these auxiliary regressions for the sensitivity of IV itself are deferred to Section 4.

Consider the regression coefficient estimate $\hat{\lambda}$ and the classical (i.e, homoskedastic) standard error estimate $\widehat{\text{se}}(\hat{\lambda})$ of Equation (5), namely, the 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 regression 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 adjusting for Z and $\boldsymbol{X}$ alone, and $\hat{\varepsilon}_{y,\text{res}}$ the corresponding residuals. The omitted variable bias framework seeks to answer the following question: how do the estimates from the restricted model compare with the estimates from the full model?

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 $\boldsymbol{X}$. It is also useful to define Cohen's partial f^2, e.g, $f^2_{Z\sim W|\boldsymbol{X}} := \frac{R^2_{Z\sim W|\boldsymbol{X}}}{1-R^2_{Z\sim W|\boldsymbol{X}}}$ which will appear frequently throughout derivations. Given the point estimate and (estimated) standard error of the restricted model actually run, $\hat{\lambda}_{\text{res}}$ and $\widehat{\text{se}}(\hat{\lambda}_{\text{res}})$, these two R^2 values are sufficient to recover $\hat{\lambda}$ and $\widehat{\text{se}}(\hat{\lambda})$.

THEOREM 1 (OVB IN THE PARTIAL R^2 PARAMETERIZATION). *Under Assumption 1, the absolute difference between the restricted and full OLS estimates is given by,*

$$|\hat{\lambda}_{res} - \hat{\lambda}| = \underbrace{\sqrt{R^2_{Y\sim W|Z,\boldsymbol{X}} f^2_{Z\sim W|\boldsymbol{X}}}}_{BF} \times \frac{sd(Y^{\perp \boldsymbol{X},Z})}{sd(Z^{\perp \boldsymbol{X}})}; \tag{7}$$

moreover, the (classical) standard error of the full OLS estimate is given by

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

where $sd(Y^{\perp \boldsymbol{X},Z})$ is the (sample) residual standard deviation of Y after removing the part linearly explained by $\{\boldsymbol{X}, Z\}$, $sd(Z^{\perp \boldsymbol{X}})$ is the (sample) residual standard deviation of Z after removing the part linearly explained by $\boldsymbol{X}$, and $\text{df} = n - p - 1$ is the residual degrees of freedom from the restricted model (2). To aid interpretation, we call the term BF in (7) the "bias factor" of W, and the term SEF in (8) the "standard error factor" of W.

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).

Note Theorem 1 is stated in terms of sample estimates. All results presented in this paper are of this type: they are exact algebraic results of how traditional OLS coefficients and standard error estimates change due to the inclusion of omitted variables. Conditions under which traditional estimates yield valid inferences are well-known and thus omitted.

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 $t^*_{\alpha,\text{df}-1} > 0$ denote the (absolute value of) 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.,

$$\text{LL}_{1-\alpha}(\lambda) := \hat{\lambda} - t^*_{\alpha,\text{df}-1} \times \widehat{\text{se}}(\hat{\lambda}), \quad \text{UL}_{1-\alpha}(\lambda) := \hat{\lambda} + t^*_{\alpha,\text{df}-1} \times \widehat{\text{se}}(\hat{\lambda}). \tag{9}$$

Considering the worst-case direction of the bias that further reduces the lower limit (or increases the upper limit) in (9), Equations (7) and (8) of Theorem 1 imply that both quantities can be written as a function of the restricted estimates and a new multiplier.

THEOREM 2 (BIAS ADJUSTED CRITICAL VALUE). *Under Assumption 1, for given $\boldsymbol{R}^2 = \{R^2_{Y\sim W|Z,\boldsymbol{X}}, R^2_{Z\sim W|\boldsymbol{X}}\}$, and α, consider the direction of the bias that reduces $LL_{1-\alpha}(\lambda)$. Then*

$$LL_{1-\alpha}(\lambda) = \hat{\lambda}_{res} - t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2} \times \widehat{\text{se}}(\hat{\lambda}_{res}). \tag{10}$$

Conversely, considering the direction of the bias that increases $UL_{1-\alpha}(\lambda)$, we have

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

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

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

where BF and SEF are the bias and standard error factors of Theorem 1.

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 observed data, but for the degrees of freedom. In other words, the bias correction is a function of the strength of unobserved confounding and the sample size alone. This allows one to quickly assess the robustness of reported findings to omitted variables of any postulated strength $\boldsymbol{R}^2$, by simply comparing the reported t-statistic with the desired adjusted critical value, even without access to the original data.

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

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

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 account both for sampling uncertainty and residual biases with the postulated strength. Note $t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$ *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 it is spurious. Thus, as the sample size grows, a higher threshold is needed in order to protect inferences against systematic biases.

We note Table 1 picks $R^2_{Y\sim W|Z,\boldsymbol{X}} = R^2_{Z\sim W|\boldsymbol{X}} = R^2$ for illustrative purposes only. Researchers can construct bias-adjusted critical values for any arbitrary pair of R^2 values—see, e.g, Supplementary Material for 2×2 tables of $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ where both $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$ are varied simultaneously.

Remark 1. Sensitivity analysis cannot reveal the strength of confounding present, only the strength of confounding required to alter a research conclusion. For instance, Table 1 reveals that in a study with 1 million observations, one needs a t-value of at least 12 in order to guarantee that the results are robust to latent variables that explain 1% of the residual variation both of the dependent and independent variables. The table also tells us that any study with a t-value less than 12 is vulnerable to such biases. The table *does not* tell us whether latent variables with such strength do exist in any particular study—this needs to be adjudicated using expert knowledge. Note, however, that knowing what one needs to know is useful, and represents an improvement over conventional analysis, which assumes $R^2 = 0$. See Section 6 for additional discussion.

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 $t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$ as a function of the sensitivity parameters $R^2_{Y\sim W|Z,\boldsymbol{X}}$ and $R^2_{Z\sim W|\boldsymbol{X}}$, we then solve the maximization problem,

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

Denoting the solution to the optimization problem in expression (14) as $t^{\dagger\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$, we obtain the maximum bias-adjusted critical value.

THEOREM 3 (MAXIMUM BIAS-ADJUSTED CRITICAL VALUE). *Fix* α, $R^{2\max}_{Y\sim W|Z,\boldsymbol{X}}$ *and* $R^{2\max}_{Z\sim W|\boldsymbol{X}} < 1$ *in the optimization problem (14). Then,*

$$t^{\dagger\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} = t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^{2*}},$$

with $\boldsymbol{R}^{2*} = \{R^{2\max}_{Y\sim W|Z,\boldsymbol{X}},\ R^{2\max}_{Z\sim W|\boldsymbol{X}}\}$ *if* $R^{2\max}_{Z\sim W|\boldsymbol{X}} \geq f^{*2}_{\alpha,df-1} f^{2\max}_{Y\sim W|Z,\boldsymbol{X}}$, *and* $\boldsymbol{R}^{2*} = \{R^{2\max}_{Z\sim W|\boldsymbol{X}}/(f^{*2}_{\alpha,df-1} + R^{2\max}_{Z\sim W|\boldsymbol{X}}),\ R^{2\max}_{Z\sim W|\boldsymbol{X}}\}$ *otherwise, where here we define* $f^{*}_{\alpha,\mathrm{df}-1} := t^{*}_{\alpha,\mathrm{df}-1}/\sqrt{\mathrm{df}-1}$.

Once in possession of $t^{\dagger\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}$, the most extreme possible lower and upper limits of confidence intervals after adjusting for W are then given by

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

The interval composed of such limits,

$$\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda) := \left[\mathrm{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda),\quad \mathrm{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)\right], \tag{15}$$

retrieves the union of all confidence intervals for λ that are compatible with an omitted variable with such strengths.

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). In what follows, 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.

3.4. *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,

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

The solution to this problem gives the following result.

THEOREM 4 (EXTREME ROBUSTNESS VALUE—OLS). *Under Assumption 1, for given q^* and α, the extreme robustness value equals*

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

where $f_{q^*}(\lambda) := q^*|f_{Y\sim Z|\boldsymbol{X}}|$, and $f^*_{\alpha,\text{df}-1} := t^*_{\alpha,\text{df}-1}/\sqrt{\text{df}-1}$.

Remark 2. Beyond its procedural interpretation, $\text{XRV}_{q^*,\alpha}(\lambda)$ can also be interpreted as an "adjusted partial R^2" of Z with Y. To see why, 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.

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

$$\mathrm{RV}_{q^*,\alpha}(\lambda) := \inf\left\{\mathrm{RV};\ (1-q^*)\hat{\lambda}_{\text{res}} \in \mathrm{CI}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\lambda)\right\}. \tag{17}$$

The RV of OLS estimates has then the following characterization.

THEOREM 5 (ROBUSTNESS VALUE—OLS). *Under Assumption 1, for given q^* and α, the robustness value equals*

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

where $f_{q^*,\alpha}(\lambda) := q^*|f_{Y\sim Z|\boldsymbol{X}}| - f^*_{\alpha,\mathrm{df}-1}$, *and* $f^*_{\alpha,\mathrm{df}-1} := t^*_{\alpha,\mathrm{df}-1}/\sqrt{\mathrm{df}-1}$.

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 (here, when the f statistic is very large, it may be numerically convenient to use the equivalent expression $2/\left(1+\sqrt{1+4/f^2_{q^*,\alpha}(\lambda)}\right)$ which avoids catastrophic cancellations). 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.

3.6. *Bounding the plausible strength of omitted variables*

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

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

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

THEOREM 6 (RELATIVE BOUNDS ON THE STRENGTH OF W). *Under Assumption 1, for fixed k_Z and k_Y as defined in (18), if $R^2_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0$ then*

$$R^2_{Z\sim W|\boldsymbol{X}} \leq \eta f^2_{Z\sim X_j|\boldsymbol{X}_{-j}}, \qquad R^2_{Y\sim W|Z,\boldsymbol{X}} = k_Y f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}, \tag{19}$$

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

These results allow investigators to leverage knowledge of *relative importance* of variables (Kruskal and Majors, 1989) when making plausibility judgments regarding sensitivity parameters, by setting $R^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}} = k_Y f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}$, $R^{2\,\max}_{Z\sim W|\boldsymbol{X}} = \eta f^2_{Z\sim X_j|\boldsymbol{X}_{-j}}$ in $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\lambda)$.

4. AN OMITTED VARIABLE BIAS FRAMEWORK FOR THE SENSITIVITY OF IV

4.1. *A suite of sensitivity analysis tools for instrumental variables*

We are now ready to develop a suite of sensitivity analysis tools for instrumental variable regression. In what follows, 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 the instrumental variable estimate. We then construct a complete omitted variable bias framework for sensitivity analysis of instrumental variables within the Anderson-Rubin approach.

4.2. *What can be learned from 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 (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 effect 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 (in the context of randomization inference, similar observations have been noted by Imbens and Rosenbaum, 2005; Small and Rosenbaum, 2008; Keele et al., 2017). We elaborate on these claims below.

Starting with the point estimate, all estimators under consideration here 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 immediate and practically 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 $\hat{\tau}$. Or, more worrisome, if analysis reveals that it takes weak confounding or side-effects to explain away $\hat{\lambda}$, the same holds for $\hat{\tau}$. Second, if we cannot rule out confounders or side-effects able to *change the sign* of the first stage, we cannot rule out that $\hat{\tau}$ 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.

Moving to inferential concerns, 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,

$$\underbrace{(\hat\theta^2 - \widehat{\text{var}}(\hat\theta)t^{*2}_{\alpha,\text{df}-1})}_{a}\tau_0^2 + \underbrace{2(\widehat{\text{cov}}(\hat\lambda,\hat\theta)t^{*2}_{\alpha,\text{df}-1} - \hat\lambda\hat\theta)}_{b}\tau_0 + \underbrace{(\hat\lambda^2 - \widehat{\text{var}}(\hat\lambda)t^{*2}_{\alpha,\text{df}-1})}_{c} \leq 0. \quad (20)$$

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 (20) 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.

4.3. *Sensitivity analysis for a specific null hypothesis*

Within the Anderson-Rubin approach, a sensitivity analysis for the null hypothesis $H_0 : \tau = \tau_0$, for any arbitrary value τ_0 can be performed as follows.

Algorithm 1. Sensitivity analysis for a specific null hypothesis.

```
(1) Set $H_0 : \tau = \tau_0$, $\alpha$, and $\mathbf{R}^2 = \{R^2_{Z\sim W|\boldsymbol{X}}, R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}\}$;
(2) Construct $Y_{\tau_0} = Y - \tau_0 D$;
(3) Fit the Anderson-Rubin regression $Y_{\tau_0} = \hat\phi_{\text{res},\tau_0} Z + \boldsymbol{X}\hat\beta_{\text{res},\tau_0} + \hat\varepsilon_{\tau_0,\text{res}}$;
(4) Compare the t-value for testing $H_0 : \phi_{\text{res},\tau_0} = 0$ against the critical value $t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$;
(5) Compute $\text{XRV}_{q^*=1,\alpha}(\phi_{\tau_0})$ and $\text{RV}_{q^*=1,\alpha}(\phi_{\tau_0})$;
(6) Report the results of (4) and (5).
```

The procedure above 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. Note the bounds on $\mathbf{R}^2$ can be chosen to reflect the assumption that the omitted variables are no stronger than certain observed covariates, as per Section 3.6.

4.4. *Compatible inferences for IV 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 $\text{CI}_{1-\alpha}(\tau) = \{\tau_0;\ t^2_{\hat\phi_{\tau_0}} \leq t^{*2}_{\alpha,\text{df}-1}\}$. 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 the IV estimate $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ is then defined as

$$\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) := \left\{\tau_0;\ t^2_{\hat\phi_{\text{res},\tau_0}} \leq \left(t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right\}. \quad (21)$$

This interval can be found analytically using the same inequality as in Equation (20), but 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 $t^{\dagger\,\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}$. 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.

4.5. *Extreme robustness values for IV*

The extreme robustness value $\text{XRV}_{q^*,\alpha}(\tau)$ for the instrumental variable 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 estimate; that is,

$$\text{XRV}_{q^*,\alpha}(\tau) := \inf\left\{\text{XRV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\tau)\right\}. \tag{22}$$

The $\text{XRV}_{q^*,\alpha}(\tau)$ computes the minimal strength of W required to not reject a particular null hypothesis of interest. However, 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, confidence intervals can sometimes consist of disjoint intervals. Therefore, let the upper and lower limits of $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ be $\text{LL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ and $\text{UL}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ respectively. The extreme robustness value $\text{XRV}_{\geq q^*,\alpha}(\tau)$ for the IV estimate is defined as the minimum strength of association that confounders or side-effects need to have with the instrument so that we cannot reject a change of $(100 \times q^*)\%$ *or worse* of the original estimate,

$$\text{XRV}_{\geq q^*,\alpha}(\tau) := \inf\left\{\text{XRV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \left[\text{LL}^{\max}_{1-\alpha,1,\text{XRV}}(\tau),\quad \text{UL}^{\max}_{1-\alpha,1,\text{XRV}}(\tau)\right]\right\}. \tag{23}$$

Both quantities can be obtained via the Anderson-Rubin and first-stage regressions as follows.

THEOREM 7 (EXTREME ROBUSTNESS VALUE—IV). *Under Assumption 1, for given q^* and α, the extreme robustness values for IV are given by*

$$XRV_{q^*,\alpha}(\tau) = XRV_{1,\alpha}(\phi_{\tau^*}), \quad \textit{and}, \tag{24}$$

$$XRV_{\geq q^*,\alpha}(\tau) = \min\{XRV_{1,\alpha}(\phi_{\tau^*}),\quad XRV_{1,\alpha}(\theta)\}, \tag{25}$$

where $\tau^* = (1-q^*)\hat{\tau}_{res}$.

Remark 3. Theorem 7 corroborates the discussion of Section 4.2. 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. Moreover, for the special case of the null hypothesis of zero effect, $H_0 : \tau = 0$, we obtain $\text{XRV}_{\geq 1,\alpha}(\tau) = \min\{\text{XRV}_{1,\alpha}(\lambda),\quad \text{XRV}_{1,\alpha}(\theta)\}$, that is, the XRV of the IV estimate, against biases that bring it to zero or worse, is equal to the minimum of the XRV of the first stage and the reduced form, both evaluated at the zero null ($q^* = 1$).

Remark 4. Note that the XRV of the first stage $\text{XRV}_{1,\alpha}(\theta)$ can be arbitrarily different from traditional metrics of instrument strength. For a simple numerical example, consider $\text{df} = 100,000$ and suppose the first stage F statistic is $F = t^2 \approx 100$, which could be considered a strong instrument for statistical inference purposes. In this case, we still have $\text{XRV}_{1,\alpha}(\theta) \approx 0.001$.

4.6. *Robustness values for IV*

The definitions of the robustness value for instrumental variables follow the same logic discussed above, but now considering both bounds on $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$ varying simultaneously. That is, the RV for not rejecting a bias of exactly q^* is defined as

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

and the RV for not rejecting the null of a reduction of $(100 \times q^*)\%$ *or worse* is defined as,

$$\mathrm{RV}_{\geq q^*,\alpha}(\tau) := \inf\left\{\mathrm{RV};\ (1-q^*)\hat{\tau}_{\text{res}} \in \left[\mathrm{LL}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\tau),\quad \mathrm{UL}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\tau)\right]\right\}. \tag{27}$$

We then have analogous results for robustness values, and similar discussion applies.

THEOREM 8 (ROBUSTNESS VALUE—IV). *Under Assumption 1, for given q^* and α, the robustness values for IV are given by*

$$RV_{q^*,\alpha}(\tau) = RV_{1,\alpha}(\phi_{\tau^*}), \quad \textit{and}, \tag{28}$$

$$RV_{\geq q^*,\alpha}(\tau) = \min\{RV_{1,\alpha}(\phi_{\tau^*}), \quad RV_{1,\alpha}(\theta)\}, \tag{29}$$

where $\tau^* = (1-q^*)\hat{\tau}_{res}$.

4.7. *Conservative bounds on the strength of omitted variables*

When testing a specific null hypothesis $H_0 : \tau = \tau_0$ in the Anderson-Rubin regression, we have k_Z as in Section 3.6, 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 thus under the null $H_0 : \tau = \tau_0$. 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 Anderson-Rubin confidence intervals, as we show next.

5. USING THE OMITTED VARIABLE BIAS FRAMEWORK FOR THE SENSITIVITY OF IV

We return to our running example of Section 2 and show how the tools developed here can be deployed to assess the robustness of the original 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).

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. We also show these same statistics for the first stage and reduced form. As derived in Theorems 7 and 8, 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 evaluated at the zero null. 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 (see Supplementary Material for separate detailed analyses of the robustness of the reduced form and first stage).

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 instrumental variable estimate

Model	Param.	Estimate	$\mathrm{LL}_{1-\alpha}$	$\mathrm{UL}_{1-\alpha}$	t-value	$\mathrm{XRV}_{\geq q^*,\alpha}$	$\mathrm{RV}_{\geq q^*,\alpha}$
Inst. Variable	τ	0.132	0.025	0.285	2.33	0.05%	0.67%
First Stage	θ	0.320	0.148	0.492	3.64	0.31%	3.02%
Reduced Form	λ	0.042	0.007	0.078	2.33	0.05%	0.67%

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

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

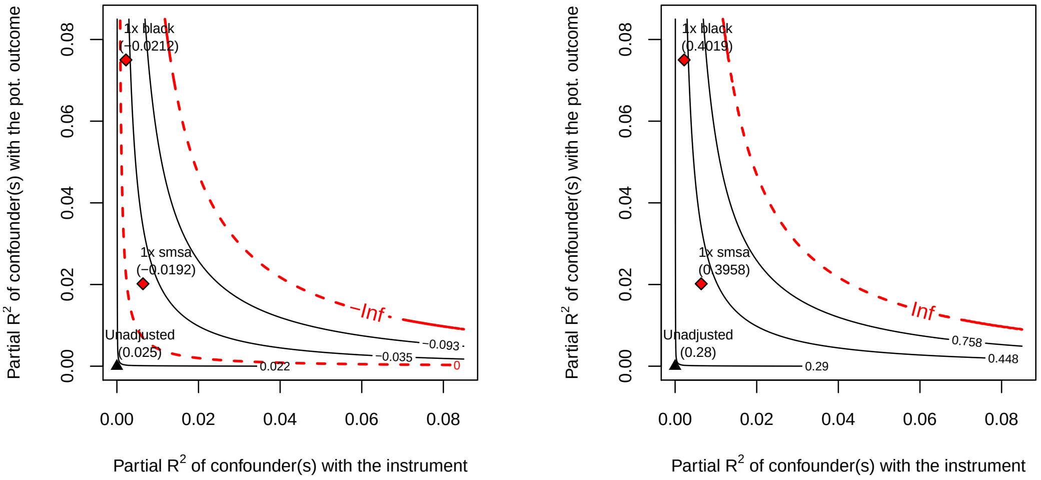
Table 2: Minimal sensitivity reporting.

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 of the instrument to be problematic. 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, $t^{\dagger\,\max}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} = 2.55$. 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.

It will often be valuable to assess the sensitivity of the instrumental variable estimate 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 Anderson-Rubin 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 $R^2_{Z\sim W|\boldsymbol{X}}$ and the vertical axis indicates the bounds on $R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}$. Under a constant treatment effects model, $R^2_{Y_{\tau_0}\sim W|Z,\boldsymbol{X}}$ has a simple interpretation—it stands for how much residual variation confounders explain of 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 $\mathrm{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$, with omitted variables bounded by such strength. Red dashed lines shows a critical contour line of interest (such as zero) as well as the boundary beyond confidence intervals become unbounded. The red diamonds places bounds on strength of W as strong as *Black* (an indicator for race) and, again, *SMSA*, as per Section 4.7. 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 study.

6. DISCUSSION

Sensitivity analysis tools, such as those introduced in this paper, provide logical deductions aimed at: (i) revealing the consequences of varying degrees of violation of identifying assumptions (e.g., via bias-adjusted critical values), and (ii) determining the minimal degree of violation of those assumptions necessary to overturn certain conclusions (e.g., via robustness values). This shifts the scientific debate from arguing whether, say, latent confounders of an instrumental variable have exactly zero strength—an indefensible claim in most settings—to a more realistic discussion about whether we can confidently rule out strengths that are shown to be problematic.



(a) Sensitivity contours for the lower limit.

(b) Sensitivity contours for the upper limit.

Figure 1: Sensitivity contours of the Anderson-Rubin confidence interval.

The results of sensitivity analyses are not always self-evident and can often be surprising. They may reveal that certain studies are highly sensitive to plausible perturbations of identifying assumptions, while others remain robust despite such perturbations. Even when results fall in between these two extremes, sensitivity analyses still represent an improvement over simply assuming away the problem. They clarify what one needs to know, by transparently revealing how vulnerable the results are to violations of the exclusion and independence restrictions. This provides policymakers a better understanding of what remains unknown about an estimated effect, and offers researchers a roadmap for improving their analyses in future inquiries.

It is important to emphasize that plausibility judgments on the maximum strength of latent variables inevitably depend on expert knowledge and can thus vary substantially across scientific disciplines, fields of study, and the quality of the research design. For that reason, we do not propose any universal thresholds for the sensitivity statistics we propose here. For instance, in an observational study without randomization nor a rich set of measured confounders, it would be hard to rule out latent confounders that explain, say, 1% of the residual variation of the instrument. This indeed seems to be the case in our running example (Card, 1993), where residual geographic confounders could plausibly attain such strength. In other scientific contexts, however, a value of 1% may in fact be large. For example, in a Mendelian randomization study where the main concern is pleiotropy, it may be defensible to argue against genetic variants explaining 1% of the variation of a latent complex pleiotropic trait (Cinelli et al., 2022).

Finally, in this paper we focused on the traditional instrumental variable estimand, consisting of the ratio of two regression coefficients. We chose to do so because this reflects current practices for IV analysis 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 traditional IV analysis in the first place. Recent papers, however, have usefully questioned the causal interpretation of this estimand, as it relies on strong parametric assumptions (Słoczyński, 2020; Blandhol et al., 2022). Extending 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.

ACKNOWLEDGEMENT

The authors thank the editor, associated editor, and reviewers for helpful comments. Cinelli's research was supported in part by the Royalty Research Fund at the University of Washington, and by the National Science Foundation under Grant No. MMS-2417955.

SUPPLEMENTARY MATERIAL

The Supplementary Material provides proofs for all results, compares our proposal with alternative approaches, and includes additional analyses of the empirical example.

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.

Belzil, C. and Hansen, J. (2002). Unobserved ability and the return to schooling. *Econometrica*, 70(5):2075–2091.

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

Blundell, R., Dearden, L., and Sianesi, B. (2001). Estimating the returns to education: Models, methods and results.

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)*.

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

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

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.

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.

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.

Jaeger, D. A. and Parys, J. (2009). On the sensitivity of return to schooling estimates to estimation methods, model specification, and influential outliers if identification is weak.

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.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803.

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.

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

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048.

Moreira, M. J. (2009). Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics*, 152(2):131–140.

Nelson, C. and Startz, R. (1990). Some further results on the exact small sample properties of the instrumental variable estimator.

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.

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.

Staiger, D. O. and Stock, J. H. (1994). Instrumental variables regression with weak instruments.

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.

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

Carlos Cinelli & Chad Hazlett

A The mechanics of IV estimation

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

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

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),

$$\textbf{First stage:} \quad D = \hat{\theta} Z + \boldsymbol{X}\hat{\psi} + \boldsymbol{W}\hat{\delta} + \hat{\varepsilon}_d, \tag{1}$$

$$\textbf{Reduced form:} \quad Y = \hat{\lambda} Z + \boldsymbol{X}\hat{\beta} + \boldsymbol{W}\hat{\gamma} + \hat{\varepsilon}_y, \tag{2}$$

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 $\hat{\theta}$,

$$\hat{\tau}_{\text{ILS}} := \frac{\hat{\lambda}}{\hat{\theta}}. \tag{3}$$

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

$$\widehat{\text{var}}(\hat{\tau}_{\text{ILS}}) := \frac{1}{\hat{\theta}^2}\left(\widehat{\text{var}}(\hat{\lambda}) + \hat{\tau}_{\text{ILS}}^2\widehat{\text{var}}(\hat{\theta}) - 2\hat{\tau}_{\text{ILS}}\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})\right) \tag{4}$$

where, using the FWL formulation,

$$\widehat{\text{var}}(\hat{\lambda}) = \frac{\text{var}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1}, \qquad \widehat{\text{var}}(\hat{\theta}) = \frac{\text{var}(D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{5}$$

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

$$\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) = \frac{\text{cov}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}, D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1} \tag{6}$$

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

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

A closely related approach for instrumental variable estimation is denoted by "two-stage least squares" (2SLS). As its name suggests, this involves two nested steps of OLS estimation: a first-stage regression given by Equation (1) to produce fitted values for the treatment ($\widehat{D}$), then regressing the outcome on these fitted values,

$$\textbf{Second stage:} \quad Y = \hat{\tau}_{\text{2SLS}} \widehat{D} + \boldsymbol{X} \hat{\beta}_{\text{2SLS}} + \boldsymbol{W} \hat{\gamma}_{\text{2SLS}} + \hat{\varepsilon}_{\text{2SLS}}. \tag{7}$$

The 2SLS estimate corresponds to the coefficient $\hat{\tau}_{\text{2SLS}}$ in Equation (7), called the "second-stage" regression.

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

$$\hat{\tau}_{\text{2SLS}} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}. \tag{8}$$

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

$$\hat{\tau}_{\text{2SLS}} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, \hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\hat{\theta} \times \text{cov}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}}, Z^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\hat{\theta}^2 \times \text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}})} = \frac{\hat{\lambda}}{\hat{\theta}}, \tag{9}$$

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

$$\widehat{\text{var}}(\hat{\tau}_{\text{2SLS}}) := \frac{\text{var}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}} - \hat{\tau} D^{\perp \boldsymbol{X}, \boldsymbol{W}})}{\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}})} \times \text{df}^{-1}. \tag{10}$$

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 (10) can be expanded to

$$\text{var}(\widehat{D}^{\perp \boldsymbol{X}, \boldsymbol{W}}) = \text{var}(\hat{\theta} Z^{\perp \boldsymbol{X}, \boldsymbol{W}}) = \hat{\theta}^2 \, \text{var}(Z^{\perp \boldsymbol{X}, \boldsymbol{W}}). \tag{11}$$

Finally, the numerator can be written as,

$$\begin{aligned}
\text{var}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}} - \hat{\tau} D^{\perp \boldsymbol{X}, \boldsymbol{W}}) &= \text{var}(Y^{\perp \boldsymbol{X}, \boldsymbol{W}} - \hat{\tau}(\hat{\theta} Z^{\boldsymbol{X}, \boldsymbol{W}} + D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}})) && (12) \\
&= \text{var}((Y^{\perp \boldsymbol{X}, \boldsymbol{W}} - \hat{\lambda} Z^{\boldsymbol{X}, \boldsymbol{W}}) - \hat{\tau} D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}) && (13) \\
&= \text{var}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}} - \hat{\tau} D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}) && (14) \\
&= \text{var}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}) + \hat{\tau}^2 \, \text{var}(D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}) - 2\hat{\tau} \, \text{cov}(Y^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}, D^{\perp Z, \boldsymbol{X}, \boldsymbol{W}}). && (15)
\end{aligned}$$

Plugging in Equations (15) and (11) back in Equation (10), then using Equations (5) and (6) 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

$$\textbf{Anderson-Rubin:} \quad Y_{\tau_0} = \hat{\phi}_{\tau_0} Z + \boldsymbol{X}\hat{\beta}_{\tau_0} + \boldsymbol{W}\hat{\gamma}_{\tau_0} + \hat{\varepsilon}_{\tau_0}, \tag{16}$$

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,

$$\hat{\tau}_{\text{AR}} = \{\tau_0;\ \hat{\phi}_{\tau_0} = 0\}. \tag{17}$$

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

$$\begin{aligned} \hat{\phi}_{\tau_0} &= \frac{\text{cov}(Y^{\perp \boldsymbol{X},\boldsymbol{W}} - \tau_0 D^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} && (18)\\ &= \frac{\text{cov}(Y^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} - \tau_0 \frac{\text{cov}(D^{\perp \boldsymbol{X},\boldsymbol{W}}, Z^{\perp \boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} && (19)\\ &= \hat{\lambda} - \tau_0 \hat{\theta}. && (20) \end{aligned}$$

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

$$\hat{\tau}_{AR} = \frac{\hat{\lambda}}{\hat{\theta}}, \tag{21}$$

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

$$\text{CI}_{1-\alpha}(\tau) = \{\tau_0; t^2_{\hat{\phi}_{\tau_0}} \leq t^{*2}_{\alpha,\text{df}}\}. \tag{22}$$

This confidence interval can be obtained analytically as functions of the estimates of the first-stage and reduced form regressions. As shown in Equation (20), $\hat{\phi}_{\tau_0}$ can be written as the linear combination

$$\hat{\phi}_{\tau_0} = \hat{\lambda} - \tau_0 \hat{\theta}. \tag{23}$$

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

$$\widehat{\text{var}}(\hat{\phi}_{\tau_0}) = \frac{\text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}} - \tau_0 D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} \times \text{df}^{-1} \tag{24}$$

$$= \left(\frac{\text{var}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} + \tau_0^2 \frac{\text{var}(D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})} - 2\tau_0 \frac{\text{cov}(Y^{\perp Z,\boldsymbol{X},\boldsymbol{W}}, D^{\perp Z,\boldsymbol{X},\boldsymbol{W}})}{\text{var}(Z^{\perp \boldsymbol{X},\boldsymbol{W}})}\right) \times \text{df}^{-1} \tag{25}$$

$$= \widehat{\text{var}}(\hat{\lambda}) + \tau_0^2 \widehat{\text{var}}(\hat{\theta}) - 2\tau_0 \widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}). \tag{26}$$

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

$$t_{\hat{\phi}_{\tau_0}} = \frac{\hat{\lambda} - \tau_0 \hat{\theta}}{\sqrt{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2 \widehat{\text{var}}(\hat{\theta}) - 2\tau_0 \widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})}}. \tag{27}$$

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

$$\frac{(\hat{\lambda} - \tau_0 \hat{\theta})^2}{\widehat{\text{var}}(\hat{\lambda}) + \tau_0^2 \widehat{\text{var}}(\hat{\theta}) - 2\tau_0 \widehat{\text{cov}}(\hat{\lambda}, \hat{\theta})} \leq t^{*2}_{\alpha,\text{df}}. \tag{28}$$

First, note that the empty set is not possible here. If we pick $\tau_0 = \hat{\tau}_{\text{AR}}$, then the numerator in Equation (28) 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

$$\underbrace{\left(\hat{\theta}^2 - \widehat{\text{var}}(\hat{\theta}) \times t^{*2}_{\alpha,\text{df}}\right)}_{a} \tau_0^2 + 2 \underbrace{\left(\widehat{\text{cov}}(\hat{\lambda}, \hat{\theta}) \times t^{*2}_{\alpha,\text{df}} - \hat{\lambda}\hat{\theta}\right)}_{b} \tau_0 + \underbrace{\left(\hat{\lambda}^2 - \widehat{\text{var}}(\hat{\lambda}) \times t^{*2}_{\alpha,\text{df}}\right)}_{c} \leq 0. \tag{29}$$

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 F.1, 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 (26), resulting in a test statistic equal to Equation (27), and thus numerically identical to the AR approach.

B OVB with the partial R^2 parameterization

As in the main text, we use the reduced form as an example, with the understanding that all results here hold for arbitrary OLS estimates.

B.1 Proof of Theorem 1

Theorem 1 was derived in Cinelli and Hazlett (2020). For completeness, we reproduce the proof here. As before var(·), sd(·), cov(·), cor(·) denote *sample* variances, standard deviations, covariances and correlations respectively. By the FWL theorem,

$$
\begin{aligned}
|\hat{\lambda}_{\text{res}} - \hat{\lambda}| &= \left(\frac{\mathrm{cov}(Z^{\perp \boldsymbol{X}},\ W^{\perp \boldsymbol{X}})}{\mathrm{var}(Z^{\perp \boldsymbol{X}})}\right)\left(\frac{\mathrm{cov}(Y^{\perp \boldsymbol{X},Z},\ W^{\perp \boldsymbol{X},Z})}{\mathrm{var}(W^{\perp \boldsymbol{X},Z})}\right) \\
&= \left(\frac{\mathrm{cor}(Z^{\perp \boldsymbol{X}},\ W^{\perp \boldsymbol{X}})\mathrm{sd}(W^{\perp \boldsymbol{X}})}{\mathrm{sd}(Z^{\perp \boldsymbol{X}})}\right)\left(\frac{\mathrm{cor}(Y^{\perp \boldsymbol{X},Z},\ W^{\perp \boldsymbol{X},D})\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(W^{\perp \boldsymbol{X},D})}\right) \\
&= \left(\frac{\mathrm{cor}(Y^{\perp \boldsymbol{X},Z},\ W^{\perp \boldsymbol{X},Z})\mathrm{cor}(Z^{\perp \boldsymbol{X}},\ W^{\perp \boldsymbol{X}})}{\frac{\mathrm{sd}(W^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(W^{\perp \boldsymbol{X}})}}\right)\left(\frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(Z^{\perp \boldsymbol{X}})}\right). \qquad (30)
\end{aligned}
$$

Noting that $\mathrm{cor}(Y^{\perp \boldsymbol{X},Z},\ W^{\perp \boldsymbol{X},Z})^2 = R^2_{Y\sim W|Z,\boldsymbol{X}}$, that $\mathrm{cor}(W^{\perp \boldsymbol{X}},\ Z^{\perp \boldsymbol{X}})^2 = R^2_{D\sim W|\boldsymbol{X}}$, and that $\frac{\mathrm{var}(W^{\perp \boldsymbol{X},Z})}{\mathrm{var}(W^{\perp \boldsymbol{X}})} = 1 - R^2_{W\sim D|\boldsymbol{X}} = 1 - R^2_{D\sim W|\boldsymbol{X}}$, we can write (30) as

$$
|\hat{\lambda}_{\text{res}} - \hat{\lambda}| = \sqrt{\frac{R^2_{Y\sim W|Z,\boldsymbol{X}}\ R^2_{D\sim W|\boldsymbol{X}}}{1 - R^2_{D\sim W|\boldsymbol{X}}}} \times \frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(Z^{\perp \boldsymbol{X}})} = \sqrt{R^2_{Y\sim W|Z,\boldsymbol{X}}\ f^2_{D\sim W|\boldsymbol{X}}} \times \frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(Z^{\perp \boldsymbol{X}})}. \qquad (31)
$$

Note this is a general bias result for linear projections, and it holds both in the sample as well as in the population, by simply replacing sample quantities with the analogous population quantities.

As for the classical standard errors, recall that, again by the FWL theorem, they can be written as:

$$
\widehat{\mathrm{se}}(\hat{\lambda}_{\text{res}}) = \frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(Z^{\perp \boldsymbol{X}})}\sqrt{\frac{1}{\mathrm{df}}}, \qquad \widehat{\mathrm{se}}(\hat{\lambda}) = \frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z,W})}{\mathrm{sd}(Z^{\perp \boldsymbol{X},W})}\sqrt{\frac{1}{\mathrm{df}-1}}. \qquad (32)
$$

Where here $\mathrm{df} = n - p - 1$. Taking the ratio we obtain

$$
\frac{\widehat{\mathrm{se}}(\hat{\lambda})}{\widehat{\mathrm{se}}(\hat{\lambda}_{\text{res}})} = \left(\frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z,W})}{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}\right)\left(\frac{\mathrm{sd}(Z^{\perp \boldsymbol{X}})}{\mathrm{sd}(Z^{\perp \boldsymbol{X},W})}\right)\sqrt{\frac{\mathrm{df}}{\mathrm{df}-1}}. \qquad (33)
$$

Using the same partial R^2 identities as before, we have

$$
\widehat{\mathrm{se}}(\hat{\lambda}) = \sqrt{\frac{1 - R^2_{Y\sim W|Z,\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\text{res}}) \times \sqrt{\frac{\mathrm{df}}{\mathrm{df}-1}} = \sqrt{\frac{1 - R^2_{Y\sim W|Z,\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}} \times \frac{\mathrm{sd}(Y^{\perp \boldsymbol{X},Z})}{\mathrm{sd}(Z^{\perp \boldsymbol{X}})} \times \sqrt{\frac{1}{\mathrm{df}-1}}. \qquad (34)
$$

To aid interpretation, we define $\mathrm{BF} := \sqrt{\frac{R^2_{Y\sim W|Z,\boldsymbol{X}} R^2_{Z\sim W|\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}}$ 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 $R^2_{Z\sim W|\boldsymbol{X}}$, and $\mathrm{SEF} := \sqrt{\frac{1 - R^2_{Y\sim W|Z,\boldsymbol{X}}}{1 - R^2_{Z\sim W|\boldsymbol{X}}}}$ 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 $R^2_{Z\sim W|\boldsymbol{X}}$.

C Bias-adjusted critical values

As in the main text, we use the reduced form as an example, with the understanding that all results here hold for arbitrary OLS estimates. Here $\mathrm{df} = n - p - 1$.

C.1 Proof of Theorem 2

Let $\mathrm{LL}_{1-\alpha}(\lambda) := \hat{\lambda} - t^*_{\alpha,\mathrm{df}-1} \times \widehat{\mathrm{se}}(\hat{\lambda})$ be the lower limit of a $1-\alpha$ level confidence interval of the full reduced form regression, where $t^*_{\alpha,\mathrm{df}-1}$ 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 we obtain,

$$\begin{aligned}
\mathrm{LL}_{1-\alpha}(\lambda) &:= \hat{\lambda} - t^*_{\alpha,\mathrm{df}-1} \times \widehat{\mathrm{se}}(\hat{\lambda}) && (35)\\
&= \hat{\lambda}_{\mathrm{res}} - \mathrm{BF}\sqrt{\mathrm{df}} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) - t^*_{\alpha,\mathrm{df}-1} \times \mathrm{SEF}\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}) && (36)\\
&= \hat{\lambda}_{\mathrm{res}} - \left(\mathrm{SEF}\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^*_{\alpha,\mathrm{df}-1} + \mathrm{BF}\sqrt{\mathrm{df}}\right) \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}). && (37)
\end{aligned}$$

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

$$\mathrm{UL}_{1-\alpha}(\lambda) = \hat{\lambda}_{\mathrm{res}} + \left(\mathrm{SEF}\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^*_{\alpha,\mathrm{df}-1} + \mathrm{BF}\sqrt{\mathrm{df}}\right) \times \widehat{\mathrm{se}}(\hat{\lambda}_{\mathrm{res}}). \tag{38}$$

Note that, in both Equations (37) and (38), 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*,

$$t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} := \mathrm{SEF}\sqrt{\mathrm{df}/(\mathrm{df}-1)} \times t^*_{\alpha,\mathrm{df}-1} + \mathrm{BF}\sqrt{\mathrm{df}}. \tag{39}$$

C.2 Proof of Theorem 3

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

$$\max_{R^2_{Y\sim W|Z,\boldsymbol{X}},R^2_{Z\sim W|\boldsymbol{X}}} t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2} \quad \text{s.t.} \quad R^2_{Y\sim W|Z,\boldsymbol{X}} \leq R^{2\max}_{Y\sim W|Z,\boldsymbol{X}}, \quad R^2_{Z\sim W|\boldsymbol{X}} \leq R^{2\max}_{Z\sim W|\boldsymbol{X}}. \tag{40}$$

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

$$\begin{aligned}
\frac{\partial(t^{\dagger}_{\alpha,\mathrm{df}-1,\boldsymbol{R}^2}/\sqrt{\mathrm{df}})}{\partial R^2_{Z\sim W|\boldsymbol{X}}} &= \frac{\partial\,\mathrm{BF}}{\partial R^2_{Z\sim W|\boldsymbol{X}}} + f^*_{\alpha,\mathrm{df}-1} \times \frac{\partial\,\mathrm{SEF}}{\partial R^2_{Z\sim W|\boldsymbol{X}}} && (41)\\
&= \frac{(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} + f^*_{\alpha,\mathrm{df}-1}\frac{(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}} && (42)\\
&= \frac{(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2} + f^*_{\alpha,\mathrm{df}-1}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{3/2}(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \geq 0. && (43)
\end{aligned}$$

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,

$$\frac{\partial(t^{\dagger}_{\alpha,\text{df}-1,\boldsymbol{R}^2}/\sqrt{\text{df}})}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} = \frac{\partial\,\text{BF}}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} + f^{*}_{\alpha,\text{df}-1} \times \frac{\partial\,\text{SEF}}{\partial R^2_{Y\sim W|Z,\boldsymbol{X}}} \tag{44}$$

$$= \frac{(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}{2(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}} + \frac{-f^{*}_{\alpha,\text{df}-1}}{2(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}} \tag{45}$$

$$= \frac{(R^2_{Z\sim W|\boldsymbol{X}})^{1/2}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2} - f^{*}_{\alpha,\text{df}-1}(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}}{2(R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Y\sim W|Z,\boldsymbol{X}})^{1/2}(1-R^2_{Z\sim W|\boldsymbol{X}})^{1/2}}. \tag{46}$$

That is, due to the variance reduction factor of the omitted variable, 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 (46) must be positive when evaluated at the solution. Rearranging and squaring, we see that this happens when

$$R^{2\,\max}_{Z\sim W|\boldsymbol{X}} \geq f^{*2}_{\alpha,\text{df}-1} \times f^{2\,\max}_{Y\sim W|Z,\boldsymbol{X}}. \tag{47}$$

Clearly, when considering the sensitivity of the point estimate, we have $f^{*}_{\alpha,\text{df}-1} = 0$, and the condition always holds. If condition of Equation (47) 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 (46) 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,

$$R^{2*}_{Y\sim W|Z,\boldsymbol{X}} = \frac{R^{2\,\max}_{Z\sim W|\boldsymbol{X}}}{f^{*2}_{\alpha,\text{df}-1} + R^{2\,\max}_{Z\sim W|\boldsymbol{X}}}. \tag{48}$$

D (Extreme) Robustness Values for OLS

As in the main text, we use the reduced form as an example, with the understanding that all results here hold for arbitrary OLS estimates.

D.1 Proof of Theorem 4

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

$$\text{XRV}_{q^*,\alpha}(\lambda) := \inf\left\{\text{XRV};\ (1-q^*)\hat{\lambda}_{\text{res}} \in \text{CI}^{\max}_{1-\alpha,1,\text{XRV}}(\lambda)\right\}. \tag{49}$$

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 (47) fails (since $f^2_{Y\sim W|Z,\boldsymbol{X}}$ is unbounded). Therefore, XRV will be given by an interior point solution. Using Equation (48) 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

$$\mathrm{XRV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if } \ f_{q^*}(\lambda) \leq f^*_{\alpha,\mathrm{df}-1}, \\ \dfrac{f^2_{q^*}(\lambda) - f^{*2}_{\alpha,\mathrm{df}-1}}{1 + f^2_{q^*}(\lambda)}, & \text{otherwise.} \end{cases} \tag{50}$$

D.2 Proof of Theorem 5

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

$$\mathrm{RV}_{q^*,\alpha}(\lambda) := \inf\left\{\mathrm{RV};\ (1 - q^*)\hat{\lambda}_{\mathrm{res}} \in \mathrm{CI}^{\max}_{1-\alpha,\mathrm{RV},\mathrm{RV}}(\lambda)\right\}, \tag{51}$$

where now we consider both sensitivity parameters bounded by RV. Again, consider the case where $f_{q^*}(\lambda) < f^*_{\alpha,\mathrm{df}\,-1}$. The RV then must be zero, since we already cannot reject the null hypothesis $H_0 : \lambda = (1 - q^*)\hat{\lambda}_{\mathrm{res}}$ 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,

$$\mathrm{RV}_{q^*,\alpha}(\lambda) = \begin{cases} 0, & \text{if } \ f_{q^*}(\lambda) \leq f^*_{\alpha,\mathrm{df}-1}, \\ \dfrac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right), & \text{if } \ f^*_{\alpha,\mathrm{df}-1} < f_{q^*}(\lambda) < f^{*-1}_{\alpha,\mathrm{df}-1}, \\ \mathrm{XRV}_{q^*,\alpha}(\lambda), & \text{otherwise.} \end{cases} \tag{52}$$

The condition $f_{q*}(\lambda) < f^{*-1}_{\alpha,\mathrm{df}-1}$, stems from the fact that the XRV solution cannot satisfy Equation (47). We now show that this is equivalent to the condition $\mathrm{RV}_{q^*,\alpha}(\lambda) > 1 - 1/f^2_{q^*}(\lambda)$ that Cinelli and Hazlett (2020) had previously established. If $f_{q^*}(\lambda) < 1/f^*_{\alpha,\mathrm{df}-1}$ then,

$$\begin{aligned}
\mathrm{RV}_{q^*,\alpha}(\lambda) &= \frac{1}{2}\left(\sqrt{f^4_{q^*,\alpha}(\lambda) + 4f^2_{q^*,\alpha}(\lambda)} - f^2_{q^*,\alpha}(\lambda)\right) && (53)\\
&= \frac{1}{2}\left(\sqrt{(f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^4 + 4(f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^2} - (f_{q^*}(\lambda) - f^*_{\alpha,\mathrm{df}-1})^2\right) && (54)\\
&> \frac{1}{2}\left(\sqrt{(f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^4 + 4(f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^2} - (f_{q^*}(\lambda) - 1/f_{q^*}(\lambda))^2\right) && (55)\\
&= \frac{1}{2}\left(\sqrt{\left(\frac{f^2_q(\lambda) - 1}{f_{q^*}(\lambda)}\right)^4 + 4\left(\frac{f^2_{q^*}(\lambda) - 1}{f_{q^*}(\lambda)}\right)^2} - \left(\frac{f^2_{q^*}(\lambda) - 1}{f_{q^*}(\lambda)}\right)^2\right) && (56)\\
&= \left(\frac{1}{2}\right)\left(\frac{f^2_{q^*}(\lambda) - 1}{f^2_{q^*}(\lambda)}\right)\left(\sqrt{(f^2_q(\lambda) - 1)^2 + 4f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) && (57)\\
&= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(\sqrt{f^4_q(\lambda) + 1 - 2f^2_{q^*}(\lambda) + 4f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) && (58)\\
&= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(\sqrt{f^4_q(\lambda) + 1 + 2f^2_{q^*}(\lambda)} - f^2_{q^*}(\lambda) + 1\right) && (59)\\
&= \left(\frac{1}{2}\right)\left(1 - 1/f^2_{q^*}(\lambda)\right)\left(f^2_{q^*}(\lambda) + 1 - f^2_{q^*}(\lambda) + 1\right) && (60)\\
&= 1 - 1/f^2_{q^*}(\lambda) && (61)
\end{aligned}$$

Therefore, $f_{q^*}(\lambda) < 1/f^*_{\alpha,\text{df}-1} \implies \text{RV}_{q^*,\alpha}(\lambda) > 1 - 1/f^2_{q^*}(\lambda)$. By the same argument one can derive $\text{RV}_{q^*,\alpha}(\lambda) > 1 - 1/f^2_{q^*}(\lambda) \implies f_q(\lambda) > 1/f^*_{\alpha,\text{df}-1}$. Hence, both conditions are equivalent.

E Bounds on the strength of W

As in the main text, we use the reduced form as an example, with the understanding that all results here hold for arbitrary OLS estimates.

E.1 Proof of Theorem 6

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

$$k_Z := \frac{R^2_{Z\sim W|\boldsymbol{X}_{-j}}}{R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim W|Z,\boldsymbol{X}_{-j}}}{R^2_{Y\sim X_j|Z\boldsymbol{X}_{-j}}}. \tag{62}$$

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.

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

$$\begin{aligned}
\left|R_{Y\sim W|Z,\boldsymbol{X}}\right| &= \left|\frac{R_{Y\sim W|Z,\boldsymbol{X}-j} - R_{Y\sim X_j|Z,\boldsymbol{X}-j}R_{W\sim X_j|Z,\boldsymbol{X}-j}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}-j}}\sqrt{1-R^2_{W\sim X_j|Z,\boldsymbol{X}-j}}}\right| &(63)\\
&= \left|\frac{R_{Y\sim W|Z,\boldsymbol{X}-j}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}-j}}}\right| &(64)\\
&= \left|\frac{\sqrt{k_Y}R_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Y\sim X_j|Z,\boldsymbol{X}-j}}}\right| &(65)\\
&= \sqrt{k_Y}\left|f_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}\right|. &(66)
\end{aligned}$$

Hence,

$$R^2_{Y\sim W|Z,\boldsymbol{X}} = k_Y \times f^2_{Y\sim X_j|Z,\boldsymbol{X}_{-j}}. \tag{67}$$

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

$$\begin{aligned}
R_{W\sim X_j|Z,\boldsymbol{X}_{-j}} = 0 &\implies \frac{R_{W\sim X_j|\boldsymbol{X}_{-j}} - R_{W\sim Z|\boldsymbol{X}_{-j}}R_{X_j\sim Z|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{W\sim Z|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{X_j\sim Z|\boldsymbol{X}_{-j}}}} = 0 &(68)\\
&\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} - R_{W\sim Z|\boldsymbol{X}_{-j}}R_{X_j\sim Z|\boldsymbol{X}_{-j}} = 0 &(69)\\
&\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} = R_{W\sim Z|\boldsymbol{X}_{-j}}R_{X_j\sim Z|\boldsymbol{X}_{-j}} &(70)\\
&\implies R_{W\sim X_j|\boldsymbol{X}_{-j}} = R_{Z\sim W|\boldsymbol{X}_{-j}}R_{Z\sim X_j|\boldsymbol{X}_{-j}} &(71)\\
&\implies |R_{W\sim X_j|\boldsymbol{X}_{-j}}| = \sqrt{k_Z}R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}. &(72)
\end{aligned}$$

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

$$\begin{aligned}
\left|R_{Z\sim W|\boldsymbol{X}}\right| &= \left|\frac{R_{Z\sim W|\boldsymbol{X}_{-j}} - R_{Z\sim X_j|\boldsymbol{X}_{-j}} R_{W\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|\boldsymbol{X}_{-j}}}}\right| && (73)\\
&\leq \frac{\left|R_{Z\sim W|\boldsymbol{X}_{-j}}\right| + \left|R_{Z\sim X_j|\boldsymbol{X}_{-j}} R_{W\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{W\sim X_j|\boldsymbol{X}_{-j}}}} && (74)\\
&= \frac{\left|\sqrt{k_Z} R_{Z\sim X_j|\boldsymbol{X}_{-j}}\right| + \left|\sqrt{k_Z} R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}} && (75)\\
&= \left(\frac{\sqrt{k_Z} + \left|R^3_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-k_Z R^4_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right) \times \left(\frac{\left|R_{Z\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}}\right) && (76)\\
&= \eta |f_{Z\sim X_j|\boldsymbol{X}_{-j}}|. && (77)
\end{aligned}$$

Hence we have that

$$R^2_{Z\sim W|\boldsymbol{X}} \leq \eta^2 f^2_{Z\sim X_j|\boldsymbol{X}_{-j}}, \tag{78}$$

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

F (Extreme) Robustness Values for IV

F.1 Proof of Theorems 7 and 8

Recall that in the AR framework, testing the null hypothesis $H_0 : \tau = \tau_0$ is equivalent to the testing the null hypothesis $H_0 : \phi_{\tau_0} = 0$. Therefore, by the definition of $(\text{X})\text{RV}_{q^*,\alpha}(\tau)$, it immediately follows that $\text{XRV}_{q^*,\alpha}(\tau) = \text{XRV}_{1,\alpha}(\phi_{\tau^*})$ and that $\text{RV}_{q^*,\alpha}(\tau) = \text{RV}_{1,\alpha}(\phi_{\tau^*})$, where $\tau^* = (1-q^*)\hat{\tau}_{\text{res}}$. We now have to show that:

$$\begin{aligned}
\text{XRV}_{\geq q^*,\alpha}(\tau) &= \min\{\text{XRV}_{1,\alpha}(\phi_{\tau^*}),\quad \text{XRV}_{1,\alpha}(\theta)\}, && (79)\\
\text{RV}_{\geq q^*,\alpha}(\tau) &= \min\{\text{RV}_{1,\alpha}(\phi_{\tau^*}),\quad \text{RV}_{1,\alpha}(\theta)\}. && (80)
\end{aligned}$$

That is, in words, that when we are interested in biases as large or larger than a certain amount, the (X)RV of the IV estimate is bounded by the the (X)RV of the first stage evaluated at the zero null.

To see why this is the case, consider the possible shapes of the adjusted AR confidence interval $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau)$. The interval is obtained by solving the following quadratic equation,

$$\begin{aligned}
&\underbrace{\left(\hat{\theta}^2_{\text{res}} - \widehat{\text{var}}(\hat{\theta}_{\text{res}}) \times \left(t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right)}_{a}\tau_0^2 + \underbrace{2\left(\widehat{\text{cov}}(\hat{\lambda}_{\text{res}}, \hat{\theta}_{\text{res}}) \times \left(t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2 - \hat{\lambda}_{\text{res}}\hat{\theta}_{\text{res}}\right)}_{b}\tau_0\\
&\qquad + \underbrace{\left(\hat{\lambda}^2_{\text{res}} - \widehat{\text{var}}(\hat{\lambda}_{\text{res}}) \times \left(t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2}\right)^2\right)}_{c} \leq 0. \qquad (81)
\end{aligned}$$

Here $\text{df} = n - p - 1$. Now 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 $a > 0$, 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 $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) = [r_{\min}, r_{\max}]$. When $a < 0$, 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 $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) = (-\infty, +\infty)$; and, (ii) when $\Delta > 0$ the confidence interval will be union of two disjoint intervals $\text{CI}^{\max}_{1-\alpha,\boldsymbol{R}^2}(\tau) = (-\infty, r_{\min}] \cup [r_{\max}, +\infty)$.[1] In both cases, arbitrarily large negative or positive values are part of the confidence interval, which is important for our discussion.

Thus we have the following conclusion: whenever $\text{CI}^{\max}_{1-\alpha,\text{df}-1}(\tau)$ is connected, we must have the same solution as before, namely, that the (X)RV of the IV estimate equals the (X)RV of the AR regression evaluated at the zero null with the appropriate outcome transformation. However, as discussed above, biases arbitrary larger than q^* can happen when the AR confidence interval is unbounded. Unbounded intervals occur if and only if $a < 0$, which is equivalent to $t^2_{\hat{\theta}_{\text{res}}} \leq (t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2})^2$. This is precisely the same condition for the (extreme) robustness value of the first stage evaluated at the zero null. Therefore, the (X)RV for IV, considering biases as large or larger than a certain amount, must be the minimum of these two values.

G 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 $\boldsymbol{X}$):

$$Y_i = \tau D_i + \eta Z_i + \varepsilon_i, \tag{82}$$

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,

$$\tau = \frac{\lambda - \eta}{\theta}. \tag{83}$$

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 (82), 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

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

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 2).

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

$$Y_i = \tau D_i + \eta Z_i + \gamma W + \varepsilon_i. \tag{84}$$

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.

H Supplementary Results for the Empirical Example

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

Table 1 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%

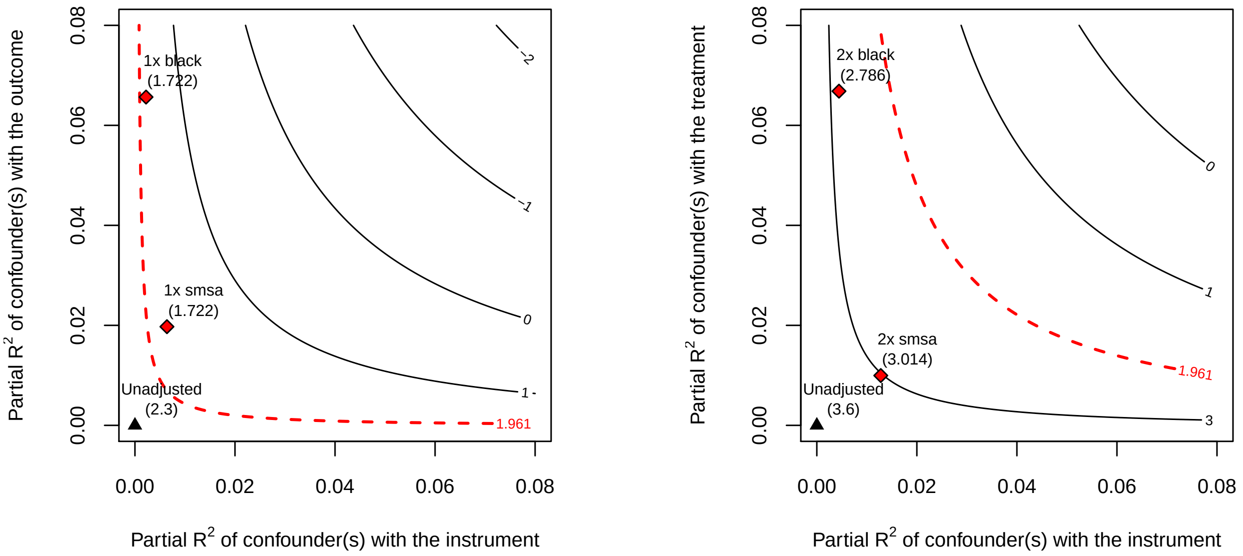
Bound (1x SMSA): $R^2_{Y\sim W|Z,\boldsymbol{X}} = 2\%$, $R^2_{W\sim Z|\boldsymbol{X}} = 0.6\%$, $t^{\dagger\max}_{\alpha,\text{df}-1,\boldsymbol{R}^2} = 2.55$

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

Table 1: 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 1 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 1, 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 1a. The



(a) Sensitivity contours of the reduced form.

(b) Sensitivity contours of the first stage.

Figure 1: Sensitivity contour plots of the reduced form and first stage.

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.

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

Table 2 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 2: Minimal sensitivity reporting of the first-stage regression.

Figure 1b 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*.

I Data on the use of sensitivity for IV

We collected data from all papers published in the American Economic Review during the year of 2020. If the paper performed an instrumental variable analysis, even if it was not the main analysis of the paper, we coded this as a paper using IV. Of this subset, we then checked whether the paper performed or mentioned any type sensitivity analysis against violations of the exclusion or independence restrictions.

J Connections to Frost (1979)

Here we briefly discuss the connection of our OVB formula with that of Frost (1979). Using our notation, Frost's formula can be written as:

$$|\lambda - \lambda_r| = |\gamma|\sqrt{R^2_{Z\sim W|X}} \times \frac{\mathrm{sd}(W^{\perp X})}{\mathrm{sd}(D^{\perp X})}.$$

Notice that Frost's formula has three sensitivity parameters related to the omitted variable W: (1) the regression coefficient of W in the long regression of the outcome γ; (2) the residual standard deviation of W after adjusting for the remaining covariates; and (3) the partial correlation of W with the independent variable of interest.

This parameterization has some shortcomings for sensitivity analysis. For instance, these three sensitivity parameters are not variation independent. That is, fixing $\sqrt{R^2_{Z\sim W|X}}$ constrains the feasible region of the product $|\gamma| \times \mathrm{sd}(W^{\perp X})$, which is now upper bounded by $\left(\frac{1}{\sqrt{1-R^2_{Z\sim W|X}}} \times \frac{\mathrm{sd}(Y^{\perp Z,X})}{\mathrm{sd}(D^{\perp X})}\right)$. Thus, for instance, a naive use of Frost's formula could lead to bounds larger than they need to be. Moreover, for the same reason, Frost's parameterization does not immediately reveal that bounding $R^2_{Z\sim W|X}$ alone is sufficient to bounds the bias, a key component for deriving some of the robustness metrics we propose in the paper, such as the XRV.

K Connections to over-identification tests

Our focus in this paper is on the just-identified case with one instrument and one treatment. The presence of multiple putative instruments, however, can be used to perform test instrument validity, provisional on the condition that at least some of the instruments are valid. Here we discuss some connections between over-identification tests and our proposal for sensitivity analysis.

First, if a researcher performs an over-identification test, and this test rejects that all instruments are valid, then this would provide a very compelling justification for running the sensitivity analysis of the type we propose in this paper—the data itself already points out that some invalid instruments must be present, and, therefore, result should not be taken at face value. We also note, however, that failing to reject that the instruments are valid does not mean the instruments are indeed valid, and unobserved confounding can still be a concern.

Another potential connection with our approach is that it may be possible to extend our sensitivity analysis to perform over-identification tests not to assess the exact validity of multiple instruments, but to assess whether multiple instruments are all approximately valid, within certain bounds. We leave further examination of sensitivity analysis with multiple instruments to future work.

L Supplementary Tables and Figures

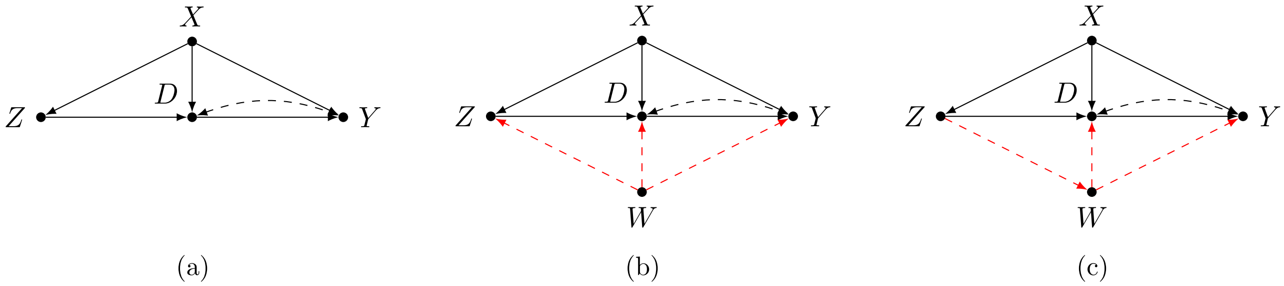


Figure 2: Causal diagrams illustrating traditional IV assumptions. Directed arrows, such as $X \to Y$, denote a possible direct causal effect of X on Y. Bidirected arrows, such as $D \leftrightarrow Y$, stand for latent common causes between D and Y. In Figure 2a, X is sufficient for rendering Z a valid instrumental variable. In Figures 2b and 2c, however, W is also needed to render Z a valid IV, either because it confounds the instrument-outcome relationship (Fig. 2b) or because it is a side-effect of the instrument affecting the outcome other than through its effect of on the treatment (Fig. 2c). In practice, all these violations will be happening simultaneously.

	Dependent variable:			
	Education	Earnings (log)		
	FS	RF	OLS	IV
	(1)	(2)	(3)	(4)
Proximity	0.320*** (0.088)	0.042** (0.018)		
Education			0.075*** (0.003)	0.132** (0.055)
Black	−0.936*** (0.094)	−0.270*** (0.019)	−0.199*** (0.018)	−0.147*** (0.054)
SMSA	0.402*** (0.105)	0.165*** (0.022)	0.136*** (0.020)	0.112*** (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 3: 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.

$R^2_{Z\sim W\mid\boldsymbol{X}}/R^2_{Y\sim W\mid Z,\boldsymbol{X}}$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
0.05	3.58	4.20	4.66	5.04	5.37	5.66	5.91	6.15	6.36	6.55	6.73	6.89	7.04	7.17	7.29	7.39	7.47	7.52	7.52
0.1	4.37	5.29	5.99	6.56	7.06	7.50	7.90	8.27	8.60	8.91	9.20	9.47	9.72	9.95	10.16	10.35	10.52	10.65	10.74
0.15	5.04	6.22	7.10	7.84	8.48	9.05	9.57	10.05	10.49	10.90	11.28	11.63	11.97	12.28	12.57	12.83	13.07	13.27	13.42
0.2	5.67	7.08	8.14	9.03	9.80	10.49	11.12	11.70	12.23	12.73	13.20	13.63	14.04	14.43	14.79	15.12	15.43	15.69	15.90
0.25	6.29	7.92	9.16	10.19	11.09	11.89	12.63	13.30	13.93	14.51	15.06	15.57	16.06	16.51	16.94	17.34	17.71	18.04	18.30
0.3	6.91	8.77	10.18	11.35	12.38	13.30	14.14	14.91	15.62	16.30	16.92	17.52	18.08	18.60	19.10	19.56	19.99	20.38	20.70
0.35	7.56	9.64	11.23	12.55	13.71	14.74	15.69	16.56	17.37	18.13	18.84	19.51	20.15	20.75	21.31	21.84	22.34	22.78	23.16
0.4	8.24	10.57	12.33	13.81	15.10	16.26	17.32	18.29	19.20	20.05	20.85	21.60	22.31	22.99	23.63	24.23	24.78	25.30	25.73
0.45	8.97	11.55	13.51	15.16	16.59	17.88	19.05	20.14	21.15	22.09	22.99	23.83	24.62	25.38	26.09	26.77	27.40	27.97	28.47
0.5	9.77	12.63	14.80	16.62	18.21	19.64	20.94	22.15	23.27	24.32	25.31	26.25	27.13	27.98	28.77	29.52	30.23	30.88	31.44
0.55	10.67	13.83	16.23	18.25	20.01	21.59	23.04	24.37	25.62	26.79	27.89	28.93	29.91	30.85	31.74	32.58	33.36	34.09	34.73
0.6	11.68	15.19	17.86	20.09	22.05	23.81	25.41	26.90	28.28	29.58	30.80	31.96	33.06	34.10	35.09	36.03	36.91	37.72	38.44
0.65	12.87	16.77	19.74	22.24	24.42	26.38	28.17	29.82	31.37	32.82	34.18	35.48	36.70	37.87	38.98	40.03	41.01	41.93	42.74
0.7	14.29	18.67	22.01	24.80	27.25	29.45	31.46	33.32	35.06	36.69	38.22	39.68	41.06	42.37	43.62	44.81	45.92	46.96	47.88
0.75	16.07	21.04	24.83	28.00	30.78	33.28	35.56	37.68	39.65	41.50	43.25	44.91	46.48	47.97	49.39	50.74	52.02	53.20	54.26
0.8	18.41	24.16	28.54	32.20	35.42	38.31	40.95	43.39	45.68	47.82	49.84	51.76	53.58	55.32	56.96	58.53	60.01	61.39	62.62
0.85	21.77	28.61	33.82	38.19	42.02	45.47	48.61	51.53	54.25	56.81	59.22	61.51	63.68	65.75	67.72	69.59	71.36	73.01	74.50
0.9	27.25	35.88	42.46	47.97	52.80	57.15	61.12	64.80	68.24	71.46	74.51	77.40	80.15	82.77	85.26	87.62	89.86	91.96	93.85
0.95	39.37	51.90	61.47	69.48	76.51	82.83	88.61	93.97	98.97	103.67	108.11	112.31	116.32	120.13	123.76	127.21	130.48	133.54	136.31

Table 4: Bias adjusted critical values: df $= 1,000$

References

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.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. 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.

Frost, P. A. (1979). Proxy variables and specification bias. *The review of economics and Statistics*, pages 323–325.

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

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

Replication File - An Omitted Variable Bias Framework for Sensitivity Analysis of Instrumental Variables

Carlos Cinelli and Chad Hazlett

2024-11-28

Main Text

Package and data

First install the package from GitHub, available at https://github.com/carloscinelli/iv.sensemakr.

```r
# loads package
library(iv.sensemakr)

# loads and prepare dataset
data("card")
y <- card$lwage  # outcome
d <- card$educ   # treatment
z <- card$nearc4 # instrument
x <- model.matrix( ~ exper + expersq + black + south + smsa + reg661 + reg662 + reg663 +
                     reg664 + reg665+ reg666 + reg667 + reg668 + smsa66,
                   data = card) # covariates
```

Anderson-Rubin regression

Fitting the Anderson-Rubin regression.

```r
# fits IV model
card.fit <- iv_fit(y,d,z,x)
card.fit
```

```
##
## Instrumental Variable Estimation
## (Anderson-Rubin Approach)
## =============================================
## IV Estimates:
##   Coef. Estimate: 0.132
##   t-value: 2.33
##   p-value: 0.02
##   Conf. Interval: [0.025, 0.285]
## Note: H0 = 0, alpha = 0.05, df = 2994.
## =============================================
## See summary for first stage and reduced form.
```

Sensitivity Analysis

Sensitivity analysis results in the main text.

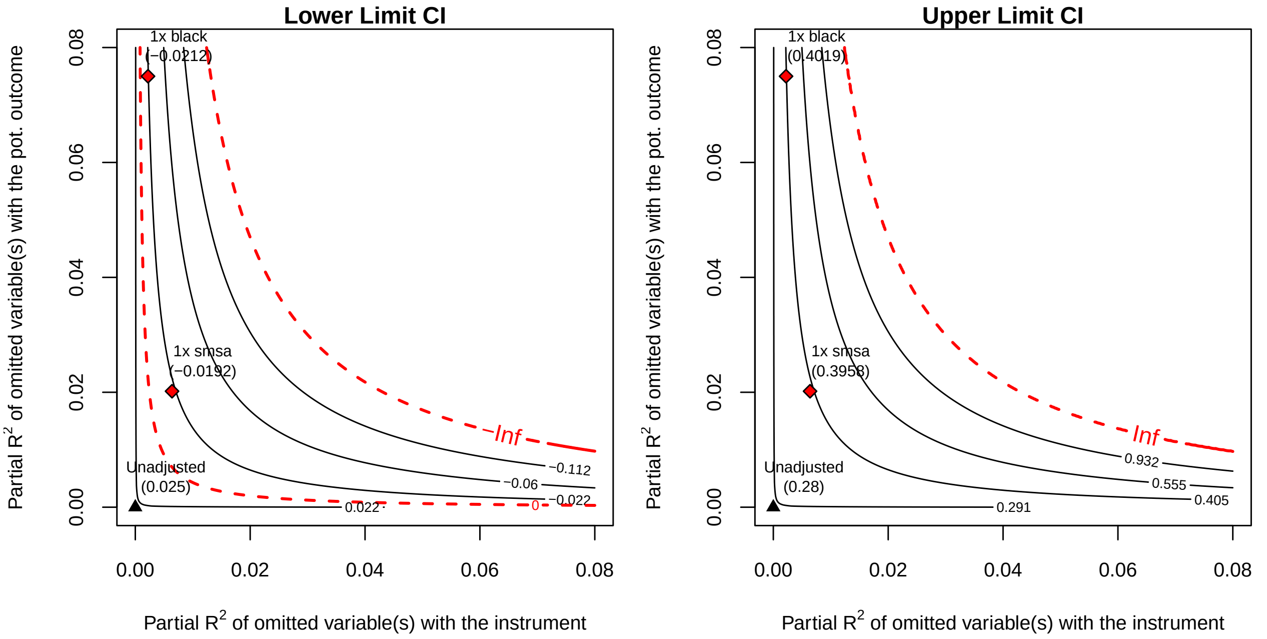
```
# runs sensitivity analysis
card.sens <- sensemakr(card.fit, benchmark_covariates = c("black", "smsa"))

# sensitivity statistics
all.stats <- sensitivity_stats(card.sens,parm = c("iv", "fs", "rf"))
round(all.stats, 4)[1:6]
```

```
##    estimate    lwr    upr t.value xrv_qa  rv_qa
## iv   0.1315 0.0248 0.2848  2.3271 0.0005 0.0067
## fs   0.3199 0.1476 0.4922  3.6408 0.0031 0.0302
## rf   0.0421 0.0066 0.0775  2.3271 0.0005 0.0067
```

```
# benchmarking
iv.bounds <- card.sens$bounds$iv
iv.bounds[-1] <- round(iv.bounds[-1], 4)
iv.bounds
```

```
##   bound_label r2zw.x r2y0w.zx     lwr    upr t.dagger
## 1    1x black 0.0022   0.0750 -0.0212 0.4019   2.5942
## 2     1x smsa 0.0064   0.0202 -0.0192 0.3958   2.5710
```

```
# sensitivity contour plot for AR
plot(card.sens, lim = 0.08)
```



Appendix – sensitivity of the Reduced Form

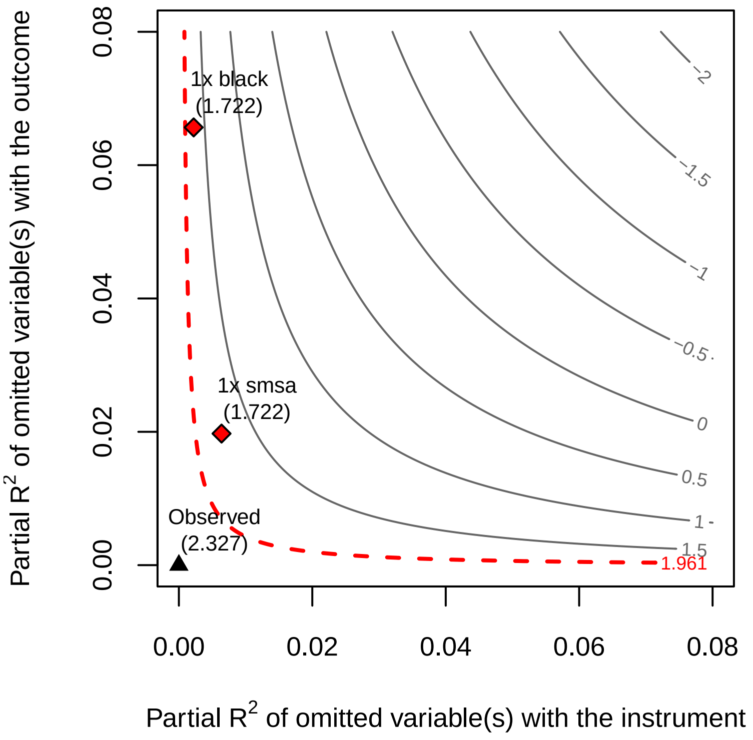
```
# sensitivity statistics
round(sensitivity_stats(card.sens, parm = "rf")[1:6], 4)
```

```
##    estimate    lwr    upr t.value xrv_qa  rv_qa
## rf   0.0421 0.0066 0.0775  2.3271 0.0005 0.0067
```

```
# benchmarking
rf.bounds <- card.sens$bounds$rf
rf.bounds[-1] <- round(rf.bounds[-1], 4)
rf.bounds
```

```
##   bound_label r2zw.x r2yw.zx     lwr    upr t.dagger
## 1    1x black 0.0022  0.0657 -0.0042 0.0883   2.5583
## 2     1x smsa 0.0064  0.0197 -0.0043 0.0884   2.5645
```

```
# sensitivity contour plot for RF
plot(card.sens, parm = "rf", sensitivity.of = "t-value", lim = 0.08, kz = 1, ky = 1)
```



Appendix – sensitivity of the First Stage

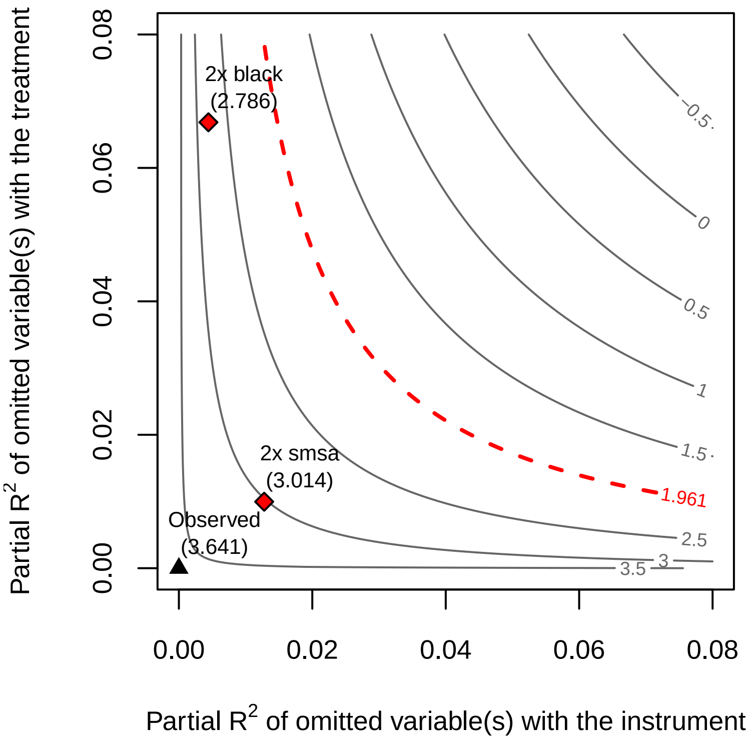
```r
# sensitivity statistics
round(sensitivity_stats(card.sens, parm = "fs")[1:6], 4)
```

```
##    estimate    lwr    upr t.value xrv_qa  rv_qa
## fs   0.3199 0.1476 0.4922  3.6408 0.0031 0.0302
```

```r
# benchmarking
fs.bounds <- card.sens$bounds$fs
fs.bounds[-1] <- round(fs.bounds[-1], 4)
fs.bounds
```

```
##   bound_label r2zw.x r2dw.zx    lwr    upr t.dagger
## 1    1x black 0.0022  0.0334 0.1089 0.5309   2.4014
## 2     1x smsa 0.0064  0.0050 0.1202 0.5195   2.2723
```

```r
# sensitivity contour plot for FS
plot(card.sens, parm = "fs", sensitivity.of = "t-value", lim = 0.08, kz = 2, kd = 2)
```


Partial R^2 of omitted variable(s) with the treatment
Partial R^2 of omitted variable(s) with the instrument
2x black
(2.786)
2x smsa
(3.014)
Observed
(3.641)