



## **sensemakr: Sensitivity Analysis Tools for OLS**

**Carlos Cinelli**

University of California, Los Angeles

**Jeremy Ferwerda**

Dartmouth College

**Chad Hazlett**

University of California, Los Angeles

---

### **Abstract**

This paper introduces the R and Stata package **sensemakr** for assessing the sensitivity of regression estimates to unobserved confounding. The package provides a suite of tools for sensitivity analysis in regression models developed in [Cinelli and Hazlett \(2020a\)](#). These tools are based on the familiar omitted variable bias framework, and can be easily computed using only standard regression results. Furthermore, they do not require assumptions on the functional form of the treatment assignment mechanism nor on the distribution of the unobserved confounders, naturally handle multiple confounders, possibly acting non-linearly, and enable bounding of sensitivity parameters employing domain knowledge.

*Keywords:* causal inference, sensitivity analysis, omitted variable bias, robustness value, R, Stata.

---

## **1. Introduction**

Investigators in a wide range of disciplines face the perennial challenge of making and defending causal claims using observational sources of data in which they could not ensure the “treatment” variable of interest was randomized or otherwise free of confounding. The most common identification strategy, in such circumstances, is to condition on or “adjust for” observed covariates in the hopes that there are not problematic unobserved covariates that remain. Linear regression remains among the most widely used approaches to realize such adjustments. However, supporting the claim that the coefficient from such a regression unbiasedly estimates a causal quantity requires defending an untestable assumption that there are no unobserved confounders. Moreover, such an assumption does not need to hold precisely (and never does) in order for a research result to remain substantively valid and informative. Sensitivity analyses quantify how strong unobserved confounding needs to be to change the research conclusions. They also aid in reasoning as whether confounding of such strength is plausible. This gives investigators the ability to rigorously describe the fragility of putative

causal estimates in the face of potential unobserved confounding.

The goal of this paper is to introduce the R and Stata package **sensemakr** (Cinelli and Hazlett 2020c; Cinelli, Ferwerda, and Hazlett 2020), which implements a suite of sensitivity analysis tools proposed in Cinelli and Hazlett (2020a) for assessing the sensitivity of a regression coefficient to the inclusion of omitted variables. The goal of **sensemakr** is to make it easy to understand the impact that omitted variables would have on a regression result. This allows analysts to investigate the robustness of their estimates to violations of the assumption of no unobserved confounding, answering questions such as:

- How strong would an unobserved confounder (or a group of confounders) have to be to change our research conclusions?
- In a worst-case scenario, how robust are our results to all unobserved confounders acting together, possibly non-linearly?
- How strong would confounding need to be relative to the strength of observed covariates, to change our answer by a certain amount?

Various sensitivity analyses have been proposed dating back to Cornfield, Haenszel, Hammond, Lilienfeld, Shimkin, and Wynder (1959), with more recent contributions including (Rosenbaum and Rubin 1983; Robins 1999; Frank 2000; Rosenbaum 2002; Imbens 2003; Brumback, Hernán, Haneuse, and Robins 2004; Frank, Sykes, Anagnostopoulos, Cannata, Chard, Krause, and McCrory 2008; Hosman, Hansen, and Holland 2010; Imai, Keele, Yamamoto *et al.* 2010; Vanderweele and Arah 2011; Blackwell 2013; Frank, Maroulis, Duong, and Kelcey 2013; Carnegie, Harada, and Hill 2016; Dorie, Harada, Carnegie, and Hill 2016; Middleton, Scott, Diakow, and Hill 2016; Oster 2017; Cinelli, Kumor, Chen, Pearl, and Bareinboim 2019; Franks, D’Amour, and Feller 2019). Yet, such sensitivity analyses remain underutilized. We argue that a number of factors contribute to this reluctant uptake. One is the complicated nature and strong assumptions many of these methods impose, sometimes involving restrictions on or even a complete description of the nature of the confounder. A second reason is that though users routinely report “regression tables” (or perhaps coefficient plots) to convey the results of a regression, until recently we have lacked “standard” quantities that can simply and correctly summarize sensitivity in the face of unobserved confounding. Third, and most fundamentally, connecting the results of a formal sensitivity analysis to a cogent argument about what types of confounders may exist in one’s research project is often difficult, particularly when there are no compelling arguments as to why the treatment assignment should be approximately “ignorable”, “exogeneous”, or “as-if random”. Further, some of the solutions offered by the literature can lead users to erroneous conclusions.

This paper is organized as follows. Section 2 briefly reviews the omitted variable bias framework for sensitivity sensitivity analysis proposed in Cinelli and Hazlett (2020a); this framework provides the theoretical foundations for the tools developed in **sensemakr**. Next, Section 3 goes over the the basic functionality and provides a practical introduction to sensitivity analysis using **sensemakr** for R. Section 4 looks at advanced usage of the R package, and shows how to leverage individual functions for customized sensitivity analyses. Finally, Section 5

describes `sensemkr` for Stata. We end the paper with final remarks...

## 2. Sensitivity analysis in an omitted variable bias framework

In this section, we briefly review the omitted variable bias (OVB) framework for sensitivity analysis presented in [Cinelli and Hazlett \(2020a\)](#). This method builds on a scale-free reparameterization of the OVB formula in terms of partial  $R^2$  values. Benefits of the parameterization include:

- assessing the sensitivity of multiple confounders acting together, possibly non-linearly;
- assessing the sensitivity to extreme scenarios in which all (or a designated portion) of the residual variation of the outcome is assumed to be explained by unobserved confounding;
- exploiting knowledge of relative strength of variables to bound the bias due to unobserved confounding;
- presenting all sensitivity results concisely, for easy routine reporting.

### 2.1. The OVB framework

The starting point of our analysis is a *full* (or “long”) linear regression model of an outcome  $Y$  on a treatment  $D$ , controlling for a set of covariates given by *both*  $\mathbf{X}$  and  $Z$ ,

$$Y = \hat{\tau}D + \mathbf{X}\hat{\beta} + \hat{\gamma}Z + \hat{\epsilon}_{\text{full}} \quad (1)$$

where  $Y$  is an  $(n \times 1)$  vector containing the outcome of interest for each of the  $n$  observations and  $D$  is an  $(n \times 1)$  treatment variable (which may be continuous or binary);  $\mathbf{X}$  is an  $(n \times p)$  matrix of *observed* covariates including the constant; and  $Z$  is a single  $(n \times 1)$  *unobserved* covariate (we discuss how to extend results for a multivariate  $Z$  below).

Equation 1 is the regression model that the investigator *wished* she had run to obtain a valid causal estimate of the effect of  $D$  on  $Y$ . Nevertheless,  $Z$  is unobserved. Therefore, the feasible regression the investigator is able to estimate is the *restricted* (or “short”) model *omitting*  $Z$ , that is,

$$Y = \hat{\tau}_{\text{res}}D + \mathbf{X}\hat{\beta}_{\text{res}} + \hat{\epsilon}_{\text{res}} \quad (2)$$

Given the discrepancy of what we wish to know and what we actually have, the main question we would like to answer is: how do the observed point estimate and standard error of the restricted regression,  $\hat{\tau}_{\text{res}}$  and  $\widehat{se}(\hat{\tau}_{\text{res}})$ , compare to the desired point estimate and standard error of the full regression,  $\hat{\tau}$  and  $\widehat{se}(\hat{\tau})$ ?

*OVB with the partial  $R^2$  parameterization*

Define as  $\widehat{\text{bias}}$  the difference between the full and restricted estimates,

$$\widehat{\text{bias}} := \hat{\tau}_{\text{res}} - \hat{\tau} \quad (3)$$

Now let: (i)  $R_{D \sim Z|\mathbf{X}}^2$  denote the share of residual variance of the *treatment*  $D$  explained by the omitted variable  $Z$ , after accounting for the remaining covariates  $\mathbf{X}$ ; and, (ii)  $R_{Y \sim Z|D, \mathbf{X}}^2$  denote the share of residual variance of the *outcome*  $Y$  explained by the omitted variable  $Z$ , after accounting for  $\mathbf{X}$  and  $D$ . Cinelli and Hazlett (2020a) have shown that these quantities are sufficient for determining the bias, adjusted estimate and adjusted standard errors of the full regression of Equation 1.

More precisely, the bias can be written as,

$$|\widehat{\text{bias}}| = \widehat{\text{se}}(\hat{\tau}_{\text{res}}) \sqrt{\frac{R_{Y \sim Z|D, \mathbf{X}}^2 R_{D \sim Z|\mathbf{X}}^2}{1 - R_{D \sim Z|\mathbf{X}}^2} (\text{df})} \quad (4)$$

Where df stands for the degrees of freedom of the restricted regression actually run. Moreover, the estimated standard error of  $\hat{\tau}$  can be recovered with,

$$\widehat{\text{se}}(\hat{\tau}) = \widehat{\text{se}}(\hat{\tau}_{\text{res}}) \sqrt{\frac{1 - R_{Y \sim Z|D, \mathbf{X}}^2}{1 - R_{D \sim Z|\mathbf{X}}^2} \left( \frac{\text{df}}{\text{df} - 1} \right)}. \quad (5)$$

Given hypothetical values of  $R_{D \sim Z|\mathbf{X}}^2$  and  $R_{Y \sim Z|D, \mathbf{X}}^2$ , Equations 4 and 5 allow investigators to examine the sensitivity of point estimates and standard-errors (and consequently also t-values, confidence intervals or p-values) to the inclusion of any omitted variable  $Z$  with such strengths. Or, conversely, given a critical threshold deemed to be problematic, one can find the strength of confounders capable of bringing about a bias of that amount. Another useful property of the OVB formula with the partial  $R^2$  parameterization is that the effect of  $R_{Y \sim Z|D, \mathbf{X}}^2$  on the bias is bounded. This allows investigators to contemplate extreme sensitivity scenarios, in which the parameter  $R_{Y \sim Z|D, \mathbf{X}}^2$  is set to 1 (or another conservative value), and see what happens as  $R_{D \sim Z|\mathbf{X}}^2$  varies.

## 2.2. Sensitivity statistics for routine reporting

The previous formulas fully determine the bias (or adjusted values) of the estimate and the standard error for any given degree of confounding, and as such can be used in numerous ways to explore the sensitivity of a regression result. For instance, sensitivity contour plots and sensitivity plots of extreme scenarios, which we demonstrate in the next sections using **sensemakr**, allow us to fully explore the sensitivity of an estimate as we vary both sensitivity parameters.

Nevertheless, making sensitivity analysis standard practice benefits from simple and interpretable statistics which can quickly describe the sensitivity of a study to unobserved confounding. These statistics serve two main purposes:

1. They can be easily displayed alongside other usual summary statistics in regression tables, making a minimal sensitivity analysis to unobserved confounding simple, accessible and standardized;

2. They can be easily computed from quantities found in a regression table, thereby enabling readers and reviewers to assess the sensitivity of results they see in print, even if the original authors did not perform sensitivity analyses.

With this in mind, [Cinelli and Hazlett \(2020a\)](#) propose two main sensitivity statistics for *routine reporting*: (i) the (observed) partial  $R^2$  of the treatment with the outcome,  $R_{Y \sim D|\mathbf{X}}^2$ ; and, the *robustness value*.

#### *The partial $R^2$ of the treatment with the outcome*

Beyond being an effect measure that quantifies how much variation of the outcome the treatment explains, the partial  $R^2$  of the treatment with the outcome can also be used to convey how robust the point estimate is to unobserved confounding in an “extreme scenario.” More precisely, suppose the unobserved confounder  $Z$  explains *all* residual variance of the outcome, that is,  $R_{Y \sim Z|D,\mathbf{X}} = 1$ . Then, for this confounder to bring the point estimate to zero, it must explain *at least* as much residual variation of the treatment as the residual variation of the outcome that the treatment currently explains. In other words, if  $R_{Y \sim Z|D,\mathbf{X}} = 1$ , then we must have that  $R_{D \sim Z|\mathbf{X}}^2 \geq R_{Y \sim D|\mathbf{X}}^2$ , otherwise this confounder cannot logically account for all the observed association between the treatment and the outcome ([Cinelli and Hazlett 2020a](#)).

#### *The Robustness Value*

The second sensitivity statistics proposed in [Cinelli and Hazlett \(2020a\)](#) is the *robustness value*. The robustness value  $RV_{q,\alpha}$  quantifies the *minimal* strength of association that the confounder needs to have, *both* with the treatment and with the outcome, so that a confidence interval of level  $\alpha$  includes a change of  $q\%$  of the current estimated value.

Let  $f_q := q|f_{Y \sim D|\mathbf{X}}|$ , where  $|f_{Y \sim D|\mathbf{X}}|$  is the partial *Cohen's  $f$*  of the treatment with the outcome multiplied by the percentage reduction  $q$  deemed to be problematic.<sup>1</sup> Also, let  $|t_{\alpha,df-1}^*|$  denote the t-value threshold for a t-test with significance level of  $\alpha$  and  $df-1$  degrees of freedom, and define  $f_{\alpha,df-1}^* := |t_{\alpha,df-1}^*|/\sqrt{df-1}$ . Finally, construct  $f_{q,\alpha}$ , which “deducts” from  $f_{Y \sim D|\mathbf{X}}$  both the proportion of reduction  $q$  of the point estimate and the boundary below which statistical significance is lost at the level of  $\alpha$ . That is,  $f_{q,\alpha} := f_q - f_{\alpha,df-1}^*$ . We then have that  $RV_{q,\alpha}$  is given by ([Cinelli and Hazlett 2020a,b](#)),

$$RV_{q,\alpha} = \begin{cases} 0, & \text{if } f_{q,\alpha} < 0 \\ \frac{1}{2} \left( \sqrt{f_{q,\alpha}^4 + 4f_{q,\alpha}^2} - f_{q,\alpha}^2 \right), & \text{if } f_q < 1/f_{\alpha,df-1}^* \\ \frac{f_q^2 - f_{\alpha,df-1}^{*2}}{1 + f_q^2}, & \text{otherwise.} \end{cases} \quad (6)$$

Any confounder that explains  $RV_{q,\alpha}\%$  of the residual variance *both* of the treatment and of the outcome is sufficiently strong to make the adjusted t-test not reject the null hypothesis  $H_0 : \tau = (1 - q)|\hat{\tau}_{\text{res}}|$  at the  $\alpha$  level (or, equivalently, sufficiently strong to make the adjusted

---

<sup>1</sup>Cohen's  $f^2$  can be written as  $f_{Y \sim D|\mathbf{X}}^2 = R_{Y \sim D|\mathbf{X}}^2/(1 - R_{Y \sim D|\mathbf{X}}^2)$

$1 - \alpha$  confidence interval include  $(1 - q)|\hat{\tau}_{\text{res}}|$ ). Likewise, a confounder with both associations lower than  $RV_{q,\alpha}$  is not capable of overturning the conclusion of such a test. Setting  $\alpha = 1$  returns the robustness value for the point estimate. Further details on how to interpret the RV in practice are given in the next sections.

### 2.3. Bounds on the strength of confounding using observed covariates

Consider a confounder orthogonal to the observed covariates, ie.,  $Z \perp \mathbf{X}$ , or, equivalently, consider only the part of  $Z$  not linearly explained by  $\mathbf{X}$ . Now denote by  $X_j$  a specific covariate of the set  $\mathbf{X}$  and define

$$k_D := \frac{R_{D \sim Z | \mathbf{X}_{-j}}^2}{R_{D \sim X_j | \mathbf{X}_{-j}}^2}, \quad k_Y := \frac{R_{Y \sim Z | \mathbf{X}_{-j}, D}^2}{R_{Y \sim X_j | \mathbf{X}_{-j}, D}^2}. \quad (7)$$

where  $\mathbf{X}_{-j}$  represents the vector of covariates  $\mathbf{X}$  excluding  $X_j$ . That is, the terms  $k_D$  and  $k_Y$  represent how strong the confounder  $Z$  is relative to observed covariate  $X_j$ , where “strength” is measured by how much residual variation they explain of the treatment (for  $k_D$ ) and of the outcome (for  $k_Y$ ). Given  $k_D$  and  $k_Y$ , we can rewrite the strength of the confounders as (Cinelli and Hazlett 2020a),

$$R_{D \sim Z | \mathbf{X}}^2 = k_D f_{D \sim X_j | \mathbf{X}_{-j}}^2, \quad R_{Y \sim Z | D, \mathbf{X}}^2 \leq \eta^2 f_{Y \sim X_j | \mathbf{X}_{-j}, D}^2 \quad (8)$$

where  $\eta$  is a scalar which depends on  $k_Y$ ,  $k_D$  and  $R_{D \sim X_j | \mathbf{X}_{-j}}^2$ .

These equations allow the investigator to assess the maximum bias that a hypothetical confounder at most “k times” as strong as a particular covariate  $X_j$  could cause. This can be used to explore the relative strength of confounding necessary for bias to have been problematic and change the research conclusions. Furthermore, when the researcher has domain knowledge to argue that a certain covariate  $X_j$  is particularly important in explaining treatment or outcome variation, and that omitted variables cannot explain as much residual variance of  $D$  or  $Y$  as that observed covariate, these results can be used to set plausible bounds in the total amount of confounding.

### 2.4. Multiple or non-linear confounders

Finally, suppose that, instead of a single unobserved confounder  $Z$ , there are *multiple* unobserved confounders  $\mathbf{Z} = [Z_1, Z_2, \dots, Z_k]$ . In this case, the regression the investigator wished she had run becomes,

$$Y = \hat{\tau}D + \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma} + \hat{\epsilon}_{\text{full}}. \quad (9)$$

As Cinelli and Hazlett (2020a) show, the previous results considering a single unobserved confounder are in fact *conservative* when considering the impact of multiple confounders, barring an adjustment in the degrees of freedom of Equation 5. Moreover, since the vector  $\mathbf{Z}$  is arbitrary, this can also accommodate non-linear confounders or even misspecification of the functional form of the observed covariates  $\mathbf{X}$ . In other words, to assess the maximum bias that multiple, non-linear confounders could cause in our current estimates, it suffices to

think in terms of the *maximum explanatory power* that  $\mathbf{Z}$  could have have in the treatment and outcome regressions, as parameterized by  $R^2_{D \sim \mathbf{Z} | \mathbf{X}}$  and  $R^2_{Y \sim \mathbf{Z} | D, \mathbf{X}}$ .

### 3. sensemakr for R: basic functionality

In this section we illustrate the basic functionality of **sensemakr** for R. Given that sensitivity analysis requires contextual knowledge to be properly interpreted, we illustrate these tools with a real example. We use **sensemakr** to reproduce all results found in Section 5 of [Cinelli and Hazlett \(2020a\)](#), which estimates the effects of exposure to violence on attitudes towards peace, in Darfur. Further details about this application and the data can be found in [Hazlett \(2019\)](#).

#### 3.1. Violence in Darfur: data and research question

In 2003 and 2004, the Darfurian government orchestrated a horrific campaign of violence against civilians, killing an estimated two hundred thousand people. This application asks whether, on average, being directly injured or maimed in this episode made individuals more likely to feel “vengeful” and unwilling to make peace with those who perpetrated this violence. Or, might those who directly suffered such violence be motivated to see it end, supporting calls for peace?

The **sensemakr** package provides the data required for this example based on a survey among Darfurian refugees in eastern Chad ([Hazlett 2019](#)). To get started we first need to install the package. From within R, the **sensemakr** package can be installed from the Comprehensive R Archive Network (CRAN).

```
R> install.packages("sensemakr")
```

Next, after loading the package, the data can be loaded with the command `data("darfur")`.

```
R> library(sensemakr)
R> data("darfur")
```

The “treatment” variable of interest is `directlyharmed`, which indicates whether the individual was physically injured or maimed during the attack on her or his village in Darfur. The main outcome of interest is `peacefactor`, an index measuring pro-peace attitudes. Other covariates in the data include: `village` (a factor variable indicating the original village of the respondent), `female` (a binary indicator of gender), `age`, `herder_dar` (whether they were a herder in Darfur), `farmer_dar` (whether they were a farmer in Darfur), and `past_voted` (whether they report having voted in an earlier election, prior to the conflict). For further details, see `?darfur`.

[Hazlett \(2019\)](#) argues that the purpose of these attacks was to punish civilians from ethnic groups presumed to support the opposition and to kill or drive these groups out so as to reduce this support. Violence against civilians included aerial bombardments by the government as well as assaults by the *Janjaweed*, a pro-government militia. For this example, suppose a researcher argues that, while some villages were more or less intensively attacked, within village violence was largely indiscriminate. The bombings could not be finely targeted owing



	<i>Dependent variable:</i>
	peacefactor
directlyharmed	0.097*** (0.023)
female	-0.232*** (0.024)
Observations	1,276
R <sup>2</sup>	0.512
Adjusted R <sup>2</sup>	0.205
Residual Std. Error	0.310 (df = 783)
F Statistic	1.670*** (df = 492; 783)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 1: OLS results for `darfur.model`. Due to sapce, only the results for `directlyharmed` and `female` are shown.

to their crudeness, and there were not many reason to target them. Similarly, the *Janjaweed* had no reason to target certain individuals rather than others, and no information with which to do so—with one major exception: women were targeted and often subjected to sexual violence.

Given these considerations, this researcher may argue that adjusting for `village` and `female` is sufficient for control of confounding, and run the following linear regression model (in which other pre-treatment covariates, although not necessary for identification, are also included):

```
R> darfur.model <- lm(peacefactor ~ directlyharmed + village + female +
R+                               age + farmer_dar + herder_dar +
R+                               pastvoted + hhsizes_darfur,
R+                               data = darfur)
```

This regression model results in the estimates shown in Table 1. According to this model, those who were directly harmed in violence were on average more “pro-peace,” not less.

### *The threat of unobserved confounders*

The previous estimate requires the assumption of *no unobserved confounders* for unbiasedness. While supported by the claim that there is no targeting of violence within village and gender strata, not all investigators may agree with this account. For example, although the bombing was crude, perhaps bombs were still more likely to hit the center of the village, and those in the center were also likely to hold different attitudes towards peace. Or, it could be the case that the *Janjaweed* observed signals that indicate individual characteristics such as wealth, and targeted using this information. Or, in another vein, an individual’s prior political attitudes could have led them to take actions that exposed them to greater risk during the attack. To



complicate things, all these factors could interact with each other or otherwise have other non-linear effects.

These concerns suggest that, instead of the previous linear model (`darfur.model`), we should have run a model such as

```
R> darfur.complete.model <- lm(peacefactor ~ directlyharmed + village +
R>                                female + age + farmer_dar + herder_dar +
R>                                pastvoted + hhsizes_darfur +
R>                                center*wealth*political_attitudes,
R>                                data = darfur)
```

Where here `center*wealth*political_attitudes` indicates *fully interacted* terms for these three variables. However trying to fit the model `darfur.complete.model` will result in error: none of the variables `center`, `wealth` or `political_attitudes` were measured.

This naturally leads to the questions: given an assumption on how strongly omitted variables relate to the treatment and the outcome, how would including them have changed our inferences regarding the coefficient of `directlyharmed`? Or, what is the *minimal strength* that these unobserved confounders (or *all* remaining unobserved confounders) need to be to change our previous conclusions? Additionally, how plausible are such confounders? For instance, considering the special role of gender in targeting, even if variables such as `wealth` remained as confounders, one could argue that it is unreasonable to expect that they explain more of the variation of exposure to violence than does gender. How much bias could a confounder as strong or stronger than `female` cause? We show next how `sensemakr` answer each of these questions.

### 3.2. Violence in Darfur: sensitivity analysis

The main function in `sensemakr` for R is `sensemakr()`. This function performs the most commonly required sensitivity analyses and return an object of class `sensemakr`, which can then be further explored with the `print`, `summary` and `plot` methods (see details in `?print.sensemakr` and `?plot.sensemakr`). We begin the analysis by applying `sensemakr()` to the original regression model, `darfur.model`.

```
R> darfur.sensitivity <- sensmakr(model = darfur.model,
R+                                treatment = "directlyharmed",
R+                                benchmark_covariates = "female",
R+                                kd = 1:3,
R+                                ky = 1:3,
R+                                q = 1,
R+                                alpha = 0.05,
R+                                reduce = TRUE)
```

The arguments of this call are:

- **model**: the `lm` object with the outcome regression. In our case, `darfur.model`.
- **treatment**: the name of the treatment variable. In our case, `"directlyharmed"`.

- **benchmark\_covariates**: the names of covariates that will be used to bound the plausible strength of the unobserved confounders. Here, we put "female", which one could argue to be among the main determinants of exposure to violence. It also found to be among the strongest determinants of attitudes towards peace empirically here.
- **kd** and **ky**: these arguments parameterize how many times stronger the confounder is related to the treatment ( **kd** ) and to the outcome ( **ky** ) in comparison to the observed benchmark covariate ( "female" ). In our example, setting **kd** = 1:3 and **ky** = 1:3 means we want to investigate the maximum strength of a confounder once, twice, or three times as strong as female (in explaining treatment and outcome variation). If only **kd** is given, **ky** will be set equal to it by default.
- **q**: this allows the user to specify what fraction of the effect estimate would have to be explained away to be problematic. Setting **q** = 1, as we do here, means that a reduction of 100% of the current effect estimate, that is, a true effect of *zero*, would be deemed problematic. The default is **q** = 1.
- **alpha**: significance level of interest for making statistical inferences. The default is set to **alpha** = 0.05.
- **reduce**: should we consider confounders acting towards *increasing* or *reducing* the absolute value of the estimate? The default is **reduce** = TRUE, which means we are considering confounders that pull the estimate towards (or through) zero. Setting **reduce** = FALSE will consider confounders that pull the estimate *away* from zero.

Using the default arguments, one can simplify the previous call to

```
R> darfur.sensitivity <- sensemakr(model = darfur.model,
R+                               treatment = "directlyharmed",
R+                               benchmark_covariates = "female",
R+                               kd = 1:3)
```

After running `sensemakr()`, we can explore the sensitivity analysis results. We note that the function `sensemakr()` also has `formula` and `numeric` methods. See `?sensemakr` for details.

### *Sensitivity statistics for routine reporting*

The print method for `sensemakr` provides the original (unadjusted) estimate along with three summary sensitivity statistics suited for *routine reporting*: (1) the partial  $R^2$  of the treatment with the outcome; (2) the robustness value (RV) required to reduce the estimate entirely to zero (i.e.  $q = 1$ ); and, (3) the RV beyond which the estimate would no longer be statistically distinguishable from zero at the 5% level ( $q = 1$ ,  $\alpha = 0.05$ ).

```
R> darfur.sensitivity
```

### Sensitivity Analysis to Unobserved Confounding

Model Formula: `peacefactor ~ directlyharmed + village + female + age + farmer_dar +`

Outcome: <i>peacefactor</i>						
Treatment:	Est.	S.E.	t-value	$R^2_{Y \sim D \mathbf{X}}$	$RV_{q=1}$	$RV_{q=1, \alpha=0.05}$
<i>directlyharmed</i>	0.097	0.023	4.184	2.2%	13.9%	7.6%
df = 783	Bound (1x female): $R^2_{Y \sim Z \mathbf{X}, D} = 12.5\%$ , $R^2_{D \sim Z \mathbf{X}} = 0.9\%$					

Table 2: Minimal sensitivity analysis reporting.

```
herder_dar + pastvoted + hhsiz_darfur
```

Unadjusted Estimates of ' *directlyharmed* ':

Coef. estimate: 0.097

Standard Error: 0.023

t-value: 4.18

Sensitivity Statistics:

Partial R2 of treatment with outcome: 0.022

Robustness Value, q = 1 : 0.139

Robustness Value, q = 1 alpha = 0.05 : 0.076

For more information, check `summary`.

The package also provides a function that creates a latex or html table with these results, as shown in Table 2 (for the html table, simply change the argument to `format = "html"`).

```
R> ovb_minimal_reporting(darfur.sensitivity, format = "latex")
```

These three sensitivity statistics provide a *minimal reporting* style for sensitivity analysis. More precisely:

- The robustness value for bringing the point estimate of *directlyharmed* exactly to zero ( $RV_{q=1}$ ) is 13.9%. This means that unobserved confounders that explain 13.9% of the residual variance *both* of the treatment and of the outcome are sufficiently strong to explain away all the observed effect. On the other hand, unobserved confounders that *do not* explain at least 13.9% of the residual variance *both* of the treatment and of the outcome are not sufficiently strong to do so.
- The robustness value for testing the null hypothesis that the coefficient of *directlyharmed* is zero ( $RV_{q=1, \alpha=0.05}$ ) falls to 7.6%. This means that unobserved confounders that explain 7.6% of the residual variance *both* of the treatment and of the outcome are sufficiently strong to bring the lower bound of the confidence interval to zero (at the chosen significance level of 5%). On the other hand, unobserved confounders that *do not* explain at least 7.6% of the residual variance *both* of the treatment and of the outcome are not sufficiently strong to do so.
- Finally, the partial  $R^2$  of *directlyharmed* with *peacefactor* means that, in an *extreme scenario*, in which we assume that unobserved confounders explain *all* of the left out variance of the outcome, these unobserved confounders would need to explain at least 2.2% of the residual variance of the treatment to fully explain away the observed effect.

These are useful quantities that summarize *what we need to know* in order to safely rule out confounders that are deemed to be problematic. Interpreting these values requires domain knowledge about the data generating process. Therefore, we encourage researchers to argue about what are plausible bounds on the maximum explanatory power that unobserved confounders could have in a given application.

Where investigators are unable to offer strong arguments limiting the *absolute strength* of confounding, it is often informative to make *relative claims*, for instance, by arguing that unobserved confounders are likely not multiple times stronger than a certain observed covariate. In our application, this is indeed the case. One could argue that, given the nature of the attacks, it is hard to imagine that unobserved confounding could explain much more of the residual variance of targeting than what is explained by the observed variable **female**. The lower corner of the table, thus, provides bounds on confounding as strong as female,  $R_{Y \sim Z|X,D}^2 = 12.5\%$ , and  $R_{D \sim Z|X}^2 = 0.9\%$ . Since both of those are below the RV, the table reveals that confounders as strong as **female** are not sufficient to explain away the observed estimate. Moreover, the bound on  $R_{D \sim Z|X}^2$  is below the partial  $R^2$  of the treatment with the outcome,  $R_{Y \sim D|X}^2$ . This means that even an extreme confounder explaining *all* residual variation of the outcome and as strongly associated with the treatment as **female** would not overturn the research conclusions. As noted in Section 2.4, these results are exact for a single unobserved confounder, and conservative for multiple confounders, possibly acting non-linearly.

Finally, the summary method for **sensemakr** provides an extensive report with verbal descriptions of all these analyses. Here, for instance, entering the command `summary(darfur.sensitivity)` produces verbose output similar to the text explanations in the last several paragraphs (and thus not reproduced here), so that researchers can directly cite or include such text in their reports.

### *Sensitivity contour plots of point estimates and t-values*

The minimal sensitivity reporting of Table 2 offers a useful summary of how robust the current estimate is to unobserved confounding. For refining their analyses, researchers can visually explore the whole range of possible estimates that confounders with different strengths could cause. These plots can also represent different bounds on the plausible strength of confounding based on different assumptions on how they compare to observed covariates. For these, one can use the plot method for **sensemakr**.

We begin by examining the default plot type, contour plots for the point estimate.

```
R> plot(darfur.sensitivity)
```

The resulting plot is shown in the left of Figure 1. The horizontal axis shows the hypothetical residual share of variation of the treatment that unobserved confounding explains,  $R_{D \sim Z|X}^2$ . The vertical axis shows the hypothetical partial  $R^2$  of unobserved confounding with the outcome,  $R_{Y \sim Z|X,D}^2$ . The contours show what would be the estimate for **directlyharmed** that one would have obtained in the full regression model including unobserved confounders with such hypothetical strengths. Note the plot is parameterized in way that hurts our preferred hypothesis, by pulling the estimate towards zero. Recall that the direction of the bias was set in the argument `reduce = TRUE` of `sensemakr()`.

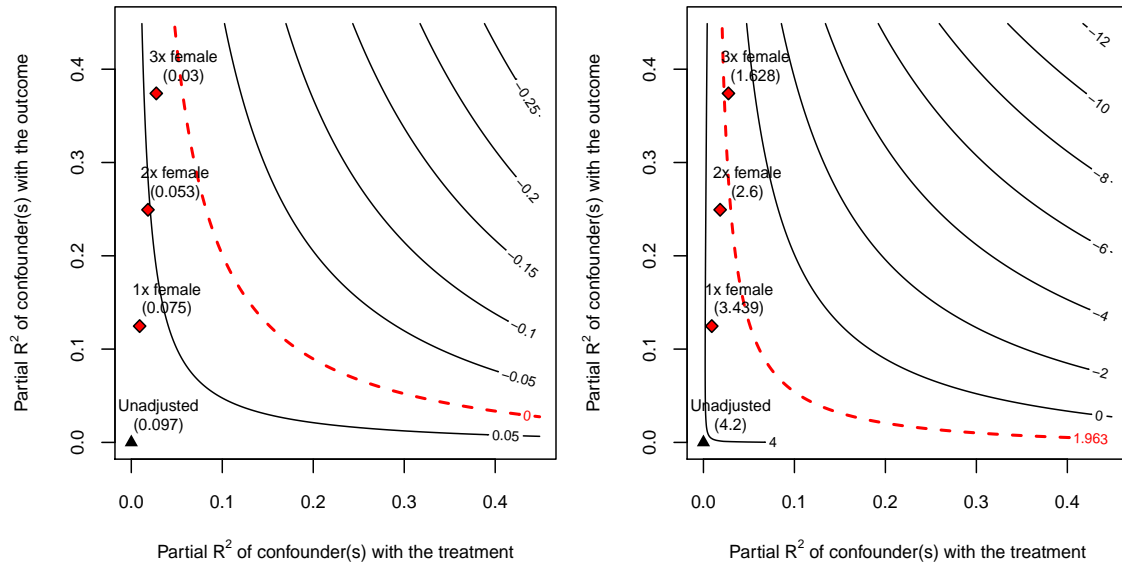


Figure 1: Sensitivity contour plots of point estimate (left) and t-value (right)

The bounds on the strength of confounding, determined by the parameter `kd = 1:3` in the call for `sensemakr()`, are also shown in the plot. The plot reveals that the direction of the effect (positive) is robust to confounding once, twice or even three times as strong as the observed covariate `female`, although in this last case the magnitude of the effect is reduced to a third of the original estimate.

We now examine the sensitivity of the *t-value* for testing the null hypothesis of zero effect. For this, it suffices to change the option `sensitivity.of = "t-value"`.

```
R> plot(darfur.sensitivity, sensitivity.of = "t-value")
```

The resulting plot is shown in the right of Figure 1. At the 5% significance level, the null hypothesis of zero effect would still be rejected given confounders once or twice as strong as `female`. However, while the point-estimate remains positive, accounting for sampling uncertainty now means that the null hypothesis of zero effect *would not* be rejected with the inclusion of a confounder three times as strong as `female`.

### *Sensitivity plots of extreme scenarios*

Sometimes researchers may be better equipped to make plausibility judgments about the strength of determinants of the treatment assignment mechanism, and have less knowledge about the determinants of the outcome. In those cases, sensitivity plots using *extreme scenarios* are a useful option. These are produced with the option `type = extreme`. Here one assumes confounding explains **all** or some large fraction of the residual variance of the outcome, then vary how strongly such confounding is hypothetically related to the treatment, to see how this affects the resulting point estimate.

```
R> plot(darfur.sensitivity, type = "extreme")
```

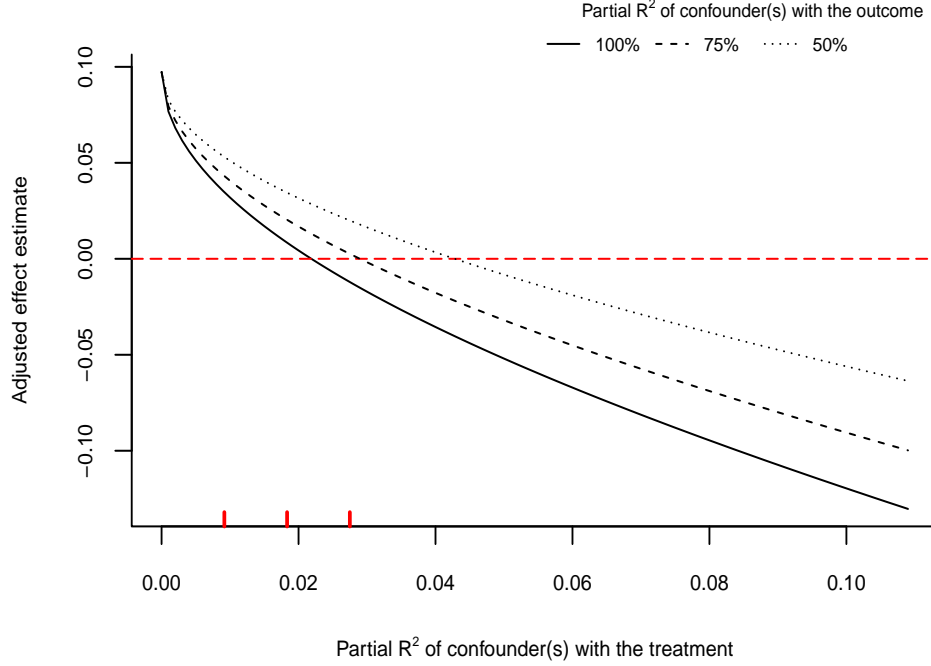


Figure 2: Sensitivity analysis to extreme scenarios.

Figure 2 shows the result. By default these plots consider confounding that explains 100%, 75%, and 50% of variation in the residual outcome, producing three separate curves for each scenario. This is equivalent to setting the argument `r2yz.dx = c(1, .75, .5)`. The bounds on the strength of association of a confounder once, twice or three times as strongly associated with the treatment as `female` are shown as red ticks in the horizontal axis. As the plot shows, even in the most extreme case ( $R_{Y \sim Z|X,D}^2 = 100\%$ ), confounders would need to be more than twice as strongly associated with the treatment as `female` to fully explain away the point estimate. Moving to the scenarios  $R_{Y \sim Z|X,D}^2 = 75\%$  and  $R_{Y \sim Z|X,D}^2 = 50\%$ , confounders would need to be more than three times as strongly associated with the treatment as `female` to fully explain away the point estimate.

### 3.3. A disciplined discussion about confounding

Having demonstrated the basic functionality of the package, we end this section by recalling some important caveats on interpretation that apply to any sensitivity analyses. Readers may refer to [Cinelli and Hazlett \(2020a\)](#) for further discussion. The results computed by `sensemakr()` tell us what we need to be prepared to believe in order to sustain that a given conclusion is not due to confounding. In particular, the results of the sensitivity analysis performed here show that, to explain all the observed estimated effect, even in a worst case scenario where the unobserved confounder explains all residual variation of the outcome, this unobserved confounder would need to be more than twice as strongly associated with the treatment as the covariate `female`. This is a *true quantitative statement* that describes the strength of confounding needed to overturn the research conclusions.

To counter a common misconception, we emphasize that such an analysis says nothing about whether such a confounder does or does not exist. The role of sensitivity analysis is, instead,

to *discipline the discussion* regarding the causal interpretation of the effect estimate. In particular,

1. A causal interpretation of the research conclusion may be defended by articulating that a confounder with such strength is unlikely to exist. For instance, one could argue that, given the way injuries (the “treatment”) occurred, the scope for targeting particular types of individuals was quite limited; aircraft dropped makeshift and unguided bombs and other objects over villages, and militia raided without concern for who they would attack—the only known major exception to this, due to sexual assaults, was targeting gender, which is also one of the most visually apparent characteristics of an individual.
2. Likewise, similar grounds are required to persuasively dismiss a causal interpretation of the result. There are requirements, in terms of (relative) strength, that the hypothesized unobserved confounders need to meet in order to be problematic. For instance, helpful skepticism must articulate why a confounder that explains at least more than twice of the variation of the treatment assignment than the covariate `female` is plausible. Otherwise, the putative confounder cannot logically account for all the observed association, even in an extreme scenario.

Robustness to confounding is thus claimed to the extent one agrees with the arguments in 1 (which rely on domain knowledge about attacks in Darfur), while a result can be deemed fragile insofar as alternative stories meeting the requirements in 2 can be offered. Sensitivity analyses should not be used to obviate discussions about confounding by engaging in automatic procedures; rather, they should stimulate a more disciplined, quantitative argument about confounding, in which such statements are made and debated.

## 4. `sensemkr` for R: advanced use

The functionality demonstrated in the previous section will suffice for most users, most of the time. Sometimes, however, more flexibility will be needed, and can be obtained by employing additional functions of `sensemkr` for R. The functions most likely to be called upon can be divided into the following categories:

- *functions for computing the bias, adjusted estimates and standard errors*: these comprise, among others, the functions `bias()`, `adjusted_estimate()`, `adjusted_se()` and `adjusted_t()`. They take as input the original (unadjusted) estimate (in the form of a linear model or numeric values) and a pair of sensitivity parameters (the partial  $R^2$  of the omitted variable with the treatment and the outcome), and return the new quantity adjusted for omitted variable bias.
- *functions for computing sensitivity statistics*: these comprise, among others, the functions `partial_r2()`, `robustness_value()`, and `sensitivity_stats()`. These functions compute sensitivity statistics suited for routine reporting, as proposed in [Cinelli and Hazlett \(2020a\)](#). They take as input the original (unadjusted) estimate (in the form of a linear model or numeric values), and return the corresponding sensitivity statistic.
- *sensitivity plots*: These functions provide direct access to sensitivity contour plots, `ovb_contour_plot()`, and sensitivity plots of extreme scenarios, `ovb_extreme_plot()`,



for customization. There is also a convenience function `add_bound_to_contour()` which facilitates the placement of user computed bounds on contour plots. All plot functions return invisibly the data needed to replicate the plot, so users that prefer fully customized plots can also easily do so. The default options for plots work best with width and height around 4 to 5 inches.

- *bounding functions*: these functions compute bounds on the strength of confounding “k times” as strong as certain observed covariates. The main high level function is `ovb_bounds()`, and there is also the auxiliary function `ovb_partial_r2_bound()`.

We demonstrate the use of these functions below through examples chosen to illustrate important features of sensitivity analysis.

#### 4.1. Formal versus informal benchmarking: customizing bounds

Informal “benchmarking” procedures have been suggested as aids to interpretation for numerous sensitivity analyses. These approaches are usually described as revealing how an unobserved confounder  $Z$  “not unlike” some observed covariate  $X_j$  would alter the results of a study (Imbens 2003; Blackwell 2013; Hosman *et al.* 2010; Carnegie *et al.* 2016; Dorie *et al.* 2016; Hong, Qin, and Yang 2018). As shown in Cinelli and Hazlett (2020a), these informal proposals may lead users to erroneous conclusions, even when they have correct knowledge of how unobserved confounders compare to observed covariates. Here we replicate Section 6.1 of Cinelli and Hazlett (2020a) using *sensemakr*, and provide a numerical example that illustrates the potential for misleading results from informal benchmarking. This example also demonstrates advanced usage of the package, including how to construct sensitivity contour plots with customized bounds.

##### *Data and model*

We begin by simulating the data generating process which will be used in our example, as given by Equations 10 to 13 below. Here we have a treatment variable  $D$ , an outcome variable  $Y$ , one observed confounder  $X$ , and one *unobserved* confounder  $Z$ . All disturbance variables  $U$  are standardized mutually independent normals. Note that, in reality, the treatment  $D$  has *no causal effect* on the outcome  $Y$ .

##### **Model 1:**

$$Z = U_z \tag{10}$$

$$X = U_x \tag{11}$$

$$D = X + Z + U_d \tag{12}$$

$$Y = X + Z + U_y \tag{13}$$

Also note that, in this model: (i) the unobserved confounder  $Z$  is independent of  $X$ ; and, (ii) the unobserved confounder  $Z$  is *exactly like*  $X$  in terms of its strength of association with the treatment and the outcome. The code below creates a sample of size 100 of this data generating process. We use the function `resid_maker()` to make sure the residuals are standardized and orthogonal, thus all properties that we describe here hold exactly even with finite sample size.

	<i>Dependent variable:</i>	
	Y	
	Restricted OLS	Full OLS
	(1)	(2)
D	0.500*** (0.088)	0.000 (0.102)
X	0.500*** (0.152)	1.000*** (0.144)
Z		1.000*** (0.144)
Observations	100	100
R <sup>2</sup>	0.500	0.667
Adjusted R <sup>2</sup>	0.490	0.656
Residual Std. Error	1.240 (df = 97)	1.020 (df = 96)
F Statistic	48.500*** (df = 2; 97)	64.000*** (df = 3; 96)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 3: First column: results of the restricted regression adjusting for  $X$  only. Second column: results of the full regression adjusting for  $X$  and  $Z$ .

```
R> n <- 100
R> X <- scale(rnorm(n))
R> Z <- resid_maker(n, X)
R> D <- X + Z + resid_maker(n, cbind(X, Z))
R> Y <- X + Z + resid_maker(n, cbind(X, Z, D))
```

In this example, the investigator knows she need to adjust for the confounder  $Z$  but, unfortunately, does not observe  $Z$ . Therefore, she is forced to fit the restricted linear model adjusting for  $X$  only.

```
R> model.ydx <- lm(Y ~ D + X)
```

Results from this regression are shown in the first column of Table 3, showing a large and statistically significant coefficient estimate  $X$ .

### Formal benchmarks

Let us suppose the investigator *correctly* knows that: (i)  $Z$  and  $X$  have the same strength of association with  $D$  and  $Y$ ; and, (ii)  $Z$  is independent of  $X$ . How can she leverage this information to understand how much bias a confounder  $Z$  “not unlike”  $X$  could cause? As we have seen in Section 2.3, using Equation 8 it is possible to obtain valid bounds on the amount

of confounding caused by an unobserved  $Z$  as strongly associated with the treatment  $D$  and with the outcome  $Y$  as the observed covariate  $X$ .

Separately from the main `sensemakr()` function, these bounds can be computed with the function `ovb_bounds()`. In this function one needs to specify the linear model being used (`model = model.ydx`), the treatment of interest (`treatment = "D"`), the observed variable used for benchmarking (`benchmark_covariates = "X"`), and how many times stronger  $Z$  is in explaining treatment (`kd = 1`) and outcome (`ky = 1`) variation, as compared to the benchmark variable  $X$ .

```
R> formal_bound <- ovb_bounds(model = model.ydx,
R+                             treatment = "D",
R+                             benchmark_covariates = "X",
R+                             kd = 1,
R+                             ky = 1)
```

We can now inspect the output of `ovb_bounds()`.

```
R> formal_bound[1:6]

bound_label r2dz.x r2yz.dx treatment adjusted_estimate adjusted_se
1          1x X    0.5    0.333          D              0        0.102
```

As we can see, the results of the bounding procedure correctly shows us that, an unobserved confounder  $Z$  that is truly “not unlike  $X$ ” would: (1) explain 50% of the residual variation of the treatment and 33% of the residual variation of the outcome; (2) bring the point estimate exactly to zero; and, (3) bring the standard error to 0.102. This is precisely what one obtains when running the full regression model adjusting for *both*  $X$  and  $Z$ , as shown in the second column of Table 3.

### *Informal benchmarks*

We now demonstrate an “informal benchmark” to show its dangers. Computing the bias due to the omission of  $Z$  requires two sensitivity parameters: its partial  $R^2$  with the treatment  $D$  and its partial  $R^2$  with the outcome  $Y$ . Informal approaches follow from the intuition that we can simply take the observed partial  $R^2$  of  $X$  with  $D$  and  $Y$  found directly from regressions for the treatment and the outcome, respectively. Unfortunately, as formalized in [Cinelli and Hazlett \(2020a\)](#), these observed associations are themselves affected by the omission of the omitted variable, making naive comparisons potentially misleading.

What happens if we nevertheless attempt to use those observed statistics for benchmarking? To compute the informal benchmarks, we first need to obtain the observed partial  $R^2$  of  $X$  with the outcome  $Y$ . This can be done using the `partial_r2()` function of `sensemakr` in the `model.ydx` regression.

```
R> r2yx.d <- partial_r2(model.ydx, covariates = "X")
```

We next need to obtain the partial  $R^2$  of  $X$  with the treatment  $D$ . For that, we need to fit a new regression of the treatment  $D$  on the observed covariate  $X$  here denoted by `model.dx`.

```
R> model.dx <- lm(D ~ X)
R> r2dx <- partial_r2(model.dx, covariates = "X")
```

We then determine what would be the implied adjusted estimate due to an unobserved confounder  $Z$  with this pair of partial  $R^2$  values. This can be computed using the `adjusted_estimate()` function.

```
R> informal_adjusted_estimate <- adjusted_estimate(model = model.ydx,
R+                                           treatment = "D",
R+                                           r2dz.x = r2dx,
R+                                           r2yz.dx = r2yx.d)
```

Let us now compare those informal benchmarks with the formal bounds. To prepare, we first plot sensitivity contours with the function `ovb_contour_plot()`. Next, we add the informal benchmark to the contours, using the numeric method of the function `add_bound_to_contour()`. Finally, we use `add_bound_to_contour()` again to add the previously computed formal bounds.

```
R> # draws sensitivity contours
R> ovb_contour_plot(model = model.ydx,
R+                 treatment = "D",
R+                 lim = .6)
R>
R> # adds informal benchmark
R> add_bound_to_contour(r2dz.x = r2dx,
R+                     r2yz.dx = r2yx.d,
R+                     bound_value = informal_adjusted_estimate,
R+                     bound_label = "Informal benchmark")
R>
R> # adds formal bound
R> add_bound_to_contour(bounds = formal_bound,
R+                     bound_label = "Formal bound")
```

Note how the results from informal benchmarking are misleading: even though  $Z$  and  $X$  are exactly alike in their roles, the informal benchmark point is still far from zero, which would suggest that an unobserved confounder  $Z$  “not unlike”  $X$  is unable to explain away the observed effect, when in fact we know it would, as shown in Table 3. This incorrect conclusion occurs *despite* the investigator *correctly* assuming both that: (i)  $Z$  and  $X$  have the same strength of association with  $D$  and  $Y$ ; and, (ii)  $Z$  is independent of  $X$ . Therefore, we do not recommend using informal benchmarks for sensitivity analysis, and suggest empirical researchers to use formal approaches such as the ones provided with `ovb_bounds()`. For further details and discussion, we point readers to Sections 4.4 and 6.1 of [Cinelli and Hazlett \(2020a\)](#).

## 4.2. Assessing the sensitivity of existing regression results

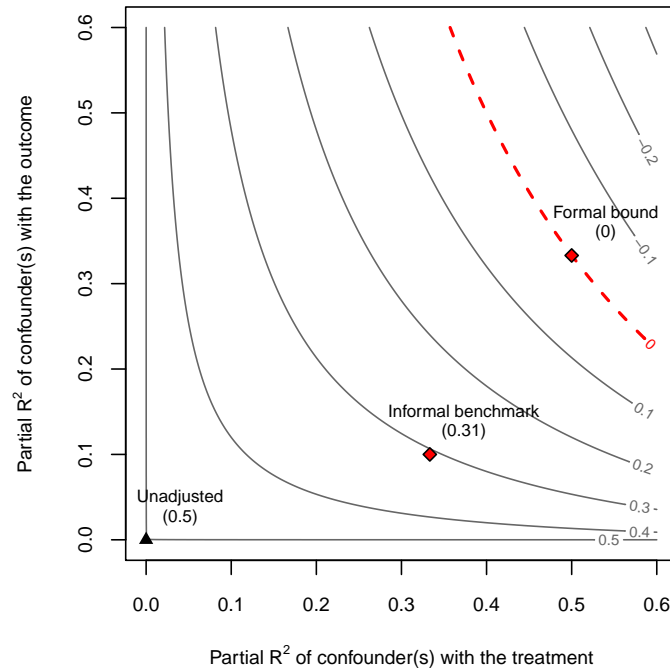


Figure 3: Informal benchmarking *versus* proper bounds.

We conclude this section by demonstrating how to replicate Section 3 using only the statistics found in the regression table along with the individual functions available in the package.

### *Sensitivity statistics*

The robustness value and partial  $R^2$  are key sensitivity statistics, useful for standardized sensitivity analyses reporting. Beyond the main `sensemakr()` function, these statistics can be computed directly by the user with the functions `robustness_value()` and `partial_r2()`. With a fitted `lm` model in hand, the most convenient way to compute the RV and partial  $R^2$  is by employing the `lm` methods for these functions, as in

```
R> robustness_value(model = darfur.model, covariates = "directlyharmed")
R> partial_r2(model = darfur.model, covariates = "directlyharmed")
```

However, when one does not have access to the data in order to run this model, simple summary statistics such as: (i) the point estimate for the `directlyharmed` (0.097); (ii) its estimated standard error (0.023); and, (iii) the degrees of freedom of the regression (783) are sufficient to compute the RV and partial  $R^2$ .

```
R> robustness_value(t_statistic = 0.097/0.023, dof = 783)
R> partial_r2(t_statistic = 0.097/0.023, dof = 783)
```

The convenience function `sensitivity_stats()` also computes all sensitivity statistics for a regression coefficient of interest and returns them in a `data.frame`.

### *Plotting functions*

All plotting functions can be called directly with `lm` objects or numerical data. For example, the code below uses the function `ovb_contour_plot()` to replicate Figure 1 (without the bounds) using only the summary statistics of Table 1.

```
R> ovb_contour_plot(estimate = 0.097, se = 0.023, dof = 783)
R> ovb_contour_plot(estimate = 0.097, se = 0.023, dof = 783,
R>                    sensitivity.of = "t-value")
```

The extreme scenario plots (as in Figure 2) can also be reproduced from summary statistics using the function `ovb_extreme_plot()`,

```
R> ovb_extreme_plot(estimate = 0.097, se = 0.023, dof = 783)
```

All plotting functions return (invisibly) the data needed to reproduce them, allowing users to create their own plots if they prefer.

### *Adjusted estimates, standard errors and t-values*

These functions allow users to compute the adjusted estimates given different postulated degrees of confounding. For instance, suppose a researcher has reasons to believe a confounder explains 10% of the residual variance of the treatment and 15% of the residual variance of the outcome. If the underlying data are not available, the investigator can still compute the adjusted estimate and t-value that one would have obtained in the full regression adjusting for such confounder.

```
R> adjusted_estimate(estimate = 0.097, se = 0.023, dof = 783,
R+                    r2dz.x = .1, r2yz.dx = 0.15)
```

```
[1] 0.0139
```

```
R> adjusted_t(estimate = 0.097, se = 0.023, dof = 783,
R+             r2dz.x = .1, r2yz.dx = 0.15)
```

```
[1] 0.622
```

The computation shows that this confounder is not strong enough to bring the estimate to zero, but it is sufficient to bring the t-value below the usual 5% significance threshold of 1.96.

### *Computing bounds from summary statistics*

Finally, we show how users can compute bounds on the strength of confounding using only summary statistics, if the paper also provides a *treatment* regression table, i.e., a regression of the treatment on the observed covariates. Such regressions are sometimes shown in published works as part of efforts to describe the “determinants” of the treatment, or as “balance tests” in which the investigator assesses whether observed covariates predict treatment assignment. For the Darfur example, this regression is shown in Table 4.

Using the results of Tables 1 and 4 we can compute the bounds on confounding 1, 2 and 3 times as strong as `female`, as we have done before. First we compute the partial  $R^2$  of `female` with the treatment and the outcome

<i>Dependent variable:</i>	
directlyharmed	
female	−0.097*** (0.036)
Observations	1,276
R <sup>2</sup>	0.426
Adjusted R <sup>2</sup>	0.066
Residual Std. Error	0.476 (df = 784)
F Statistic	1.180** (df = 491; 784)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Table 4: Treatment regression for the Darfur example. Due to space, only the results for `female` are shown, which will be used for benchmarking.

```
R> r2yxj.dx <- partial_r2(t_statistic = -0.232/0.024, dof = 783)
R> r2dxj.x <- partial_r2(t_statistic = -0.097/0.036, dof = 783)
```

Next, we compute the bounds on the partial  $R^2$  of the unobserved confounder using the `ovb_partial_r2_bound()` function.

```
R> bounds <- ovb_partial_r2_bound(r2dxj.x = r2dxj.x,
R+                               r2yxj.dx = r2yxj.dx,
R+                               kd = 1:3,
R+                               ky = 1:3,
R+                               bound_label = paste(1:3, "x", "female"))
```

Finally, the `adjusted_estimate()` function computes the estimates implied by such hypothetical confounders.

```
R> bound.values <- adjusted_estimate(estimate = 0.0973,
R+                                   se = 0.0232,
R+                                   dof = 783,
R+                                   r2dz.x = bounds$r2dz.x,
R+                                   r2yz.dx = bounds$r2yz.dx)
```

This information along with the numeric methods for the plot functions, allow us to reproduce the contour plots of Figure 1 using only summary statistics. Note that, since we are performing all calculations manually, appropriate limits of the plot area need to be set by the user.

```
R> ovb_contour_plot(estimate = 0.0973, se = 0.0232, dof = 783, lim = 0.45)
R> add_bound_to_contour(bounds, bound_value = bound.values)
```



## 5. sensemakr for Stata

For Stata users, we have also developed a homonymous package **sensemakr**.

**sensemakr** for `\proglang{Stata}` is available for download on SCC...

It also comes with the **darfur** data (?).

In this section, we demonstrate how to replicate the analysis of Section 3 using the `\proglang{Stata}` implementation of **sensemakr**. The main function of the **Stata** package is **sensemakr**...

## 6. Final Remarks

In this article we have demonstrated how to perform sensitivity analysis to unobserved confounding using the R and Stata packages **sensemakr**.

These packages can help making sensitivity analysis routine and standardized.

## References

- Blackwell M (2013). “A selection bias approach to sensitivity analysis for causal effects.” *Political Analysis*, **22**(2), 169–182.
- Brumback BA, Hernán MA, Haneuse SJ, Robins JM (2004). “Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures.” *Statistics in medicine*, **23**(5), 749–767.
- Carnegie NB, Harada M, Hill JL (2016). “Assessing sensitivity to unmeasured confounding using a simulated potential confounder.” *Journal of Research on Educational Effectiveness*, **9**(3), 395–420.
- Cinelli C, Ferwerda J, Hazlett C (2020). “sensemakr for Stata: sensitivity analysis tools for OLS.” *Stata Package v1.0*.
- Cinelli C, Hazlett C (2020a). “Making Sense of Sensitivity: Extending Omitted Variable Bias.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **82**(1), 39–67. doi:10.1111/rssb.12348.
- Cinelli C, Hazlett C (2020b). “An omitted variable bias framework for sensitivity analysis of instrumental variables.” *Working Paper*.
- Cinelli C, Hazlett C (2020c). “sensemakr: sensitivity analysis tools for OLS.” *R package Version 0.2*, **2**.
- Cinelli C, Kumor D, Chen B, Pearl J, Bareinboim E (2019). “Sensitivity Analysis of Linear Structural Causal Models.” *International Conference on Machine Learning*.
- Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959). “Smoking and lung cancer: recent evidence and a discussion of some questions.” *journal of National Cancer Institute*, (23), 173–203.

- Dorie V, Harada M, Carnegie NB, Hill J (2016). “A flexible, interpretable framework for assessing sensitivity to unmeasured confounding.” *Statistics in medicine*, **35**(20), 3453–3470.
- Frank KA (2000). “Impact of a confounding variable on a regression coefficient.” *Sociological Methods & Research*, **29**(2), 147–194.
- Frank KA, Maroulis SJ, Duong MQ, Kelcey BM (2013). “What would it take to change an inference? Using Rubin’s causal model to interpret the robustness of causal inferences.” *Educational Evaluation and Policy Analysis*, **35**(4), 437–460.
- Frank KA, Sykes G, Anagnostopoulos D, Cannata M, Chard L, Krause A, McCrory R (2008). “Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters?” *Educational Evaluation and Policy Analysis*, **30**(1), 3–30.
- Franks A, D’Amour A, Feller A (2019). “Flexible sensitivity analysis for observational studies without observable implications.” *Journal of the American Statistical Association*, (just-accepted), 1–38.
- Hazlett C (2019). “Angry or Weary? How Violence Impacts Attitudes toward Peace among Darfuriian Refugees.” *Journal of Conflict Resolution*, p. 0022002719879217.
- Hong G, Qin X, Yang F (2018). “Weighting-Based Sensitivity Analysis in Causal Mediation Studies.” *Journal of Educational and Behavioral Statistics*, **43**(1), 32–56.
- Hosman CA, Hansen BB, Holland PW (2010). “The Sensitivity of Linear Regression Coefficients’ Confidence Limits to the Omission of a Confounder.” *The Annals of Applied Statistics*, pp. 849–870.
- Imai K, Keele L, Yamamoto T, *et al.* (2010). “Identification, inference and sensitivity analysis for causal mediation effects.” *Statistical science*, **25**(1), 51–71.
- Imbens GW (2003). “Sensitivity to exogeneity assumptions in program evaluation.” *The American Economic Review*, **93**(2), 126–132.
- Middleton JA, Scott MA, Diakow R, Hill JL (2016). “Bias amplification and bias unmasking.” *Political Analysis*, **24**(3), 307–323.
- Oster E (2017). “Unobservable selection and coefficient stability: Theory and evidence.” *Journal of Business & Economic Statistics*, pp. 1–18.
- Robins JM (1999). “Association, causation, and marginal structural models.” *Synthese*, **121**(1), 151–179.
- Rosenbaum PR (2002). “Observational studies.” In *Observational studies*, pp. 1–17. Springer.
- Rosenbaum PR, Rubin DB (1983). “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome.” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 212–218.

Vanderweele TJ, Arah OA (2011). “Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders.” *Epidemiology (Cambridge, Mass.)*, **22**(1), 42–52. ISSN 1531-5487. doi:[10.1097/ede.0b013e3181f74493](https://doi.org/10.1097/ede.0b013e3181f74493). URL <http://dx.doi.org/10.1097/ede.0b013e3181f74493>.

**Affiliation:**

Carlos Cinelli

University of California, Los Angeles

Department of Statistics, 8125 Math Sciences Building, Los Angeles, CA 90095, USA.

E-mail: [carloscinelli@ucla.edu](mailto:carloscinelli@ucla.edu)

URL: <http://carloscinelli.com>

Jeremy Ferwerda

Dartmouth College

Department of Government, Hanover, NH 03755

E-mail: [jeremy.a.ferwerda@dartmouth.edu](mailto:jeremy.a.ferwerda@dartmouth.edu)

URL: <http://jeremyferwerda.com/>

Chad Hazlett

University of California, Los Angeles

Department of Statistics, 8125 Math Sciences Building, Los Angeles, CA 90095, USA.

E-mail: [chazlett@ucla.edu](mailto:chazlett@ucla.edu)

URL: <http://chadhazlett.com>