CHAPTER 22

Sensitivity Analysis and Bounds

22.1 INTRODUCTION

Part IV of this text focused on estimation and inference under regular assignment mechanisms, that is, ones that are individualistic with probabilistic assignment, as well as unconfounded. In Part V we study methods that confront the unconfoundedness assumption. In Chapter 21 we discussed methods to assess the plausibility of this assumption by combining it with additional assumptions. In the current chapter we relax the unconfoundedness assumption without replacing it with additional assumptions, and so do not focus on obtaining point estimates of the causal estimands of interest. Instead we end up with ranges of plausible values for these estimands, with the width of these ranges corresponding to the extent to which we allow the unconfoundedness assumption to be violated.

We consider two approaches that have much in common. The first, developed by Manski in a series of studies (e.g., Manski, 1990, 1996, 2003, 2013), allows for arbitrarily large violations of the unconfoundedness assumption. This *bounds* or *partial identification* approach, as it is called, leads to sharp results, but at the same time will be seen to limit severely the types of inferences about causal effects that can be drawn from observational data. The second approach, following work in this area by Rosenbaum and Rubin (1983) and Rosenbaum (1995), with important antecedents in the work by Cornfield et al. (1959), works from the other extreme in the sense that unconfoundedness is the starting point, and only limited violations from it need to be considered. If we allow for large violations in the Rosenbaum-Rubin approach, it will often lead to essentially the same results as the Manski bounds approach, but with limited violations of the unconfoundedness assumption, the sensitivity approach results in narrower ranges for the estimands than the partial identification approach.

The key to any sensitivity analysis will be how to assess the magnitude of violations from unconfoundedness. The setup in the current chapter assumes that unconfoundedness is satisfied conditional on an additional, unobserved covariate. If, conditional on the other, observed, covariates, this unobserved covariate is independent of the potential outcomes, or if, again conditional on the observed covariates, it is independent of treatment assignment, unconfoundedness holds even without conditioning on this additional covariate. If, however, this additional, unobserved covariate is associated both

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

with the potential outcomes and with the treatment indicator, biases will result from estimates based on the assumption of unconfoundedness. The magnitude of the bias depends on the strength of the associations between the unobserved covariate and the potential outcomes and treatment indicator.

In the Rosenbaum-Rubin sensitivity approach we consider the range of implied treatment effects as a function of the magnitude of the associations between the unobserved covariate and the potential outcomes and treatment indicator. To assess what reasonable magnitudes are for those associations, we compare them to the associations between observed covariates and the potential outcomes and treatment indicators in the current data, or in cases where other more extensive data are available, to those data.

We also consider a second approach to sensitivity analyses developed by Rosenbaum (1995). Here the sensitivity analyses only require the researcher to specify the magnitude of the association between the unobserved components and the treatment assignment, taking a Manski-style attitude to the associations between the hidden covariate and the potential outcomes. Without making assumptions about associations with the potential outcomes we again obtain ranges of average treatment effects consistent with the evidence in the current study.

Throughout this chapter we take a super-population approach where the sample is viewed as a random sample from an infinite population, with the random sampling generating a distribution for the potential outcomes. In Section 22.2 we describe the subset of the lottery data that will be used to illustrate the sensitivity analyses. Next, in Section 22.3, we study the Manski bounds approach. In Section 22.4 we study the Rosenbaum-Rubin sensitivity approach for the case with binary outcomes. Next, in Section 22.5 we discuss Rosenbaum's approach. Section 22.6 concludes.

22.2 THE IMBENS-RUBIN-SACERDOTE LOTTERY DATA

Here we use again the lottery data originally collected by Imbens, Rubin, and Sacerdote (2001) that we used previously in Chapters 14, 17, 19, and 21. In Chapter 14 we assessed the overlap in covariate distributions for the lottery data and found that overlap was substantial, although there were subsets of covariate values with little overlap. In Chapter 17 we used the methods from Chapter 16 to trim the sample, originally consisting of 496 units, which led to the creation of a sample containing information on $N = 323$ individuals, of whom $N_{\rm c} = 172$ are losers and $N_{\rm t} = 151$ are winners, which comprise the sample that is the basis for the analyses in this chapter. The outcome that we are studying is the indicator for having positive earnings during the six-year period (essentially being employed full time during each of these six years) following the lottery.

Assuming unconfoundedness, and using the subclassification estimator developed in Chapter 17, the point estimate of the average effect of winning the lottery on the outcome is -0.134, with an estimated standard error equal to 0.049.

22.3 BOUNDS

We start by focusing on a simple case with no covariates. For unit i, there are two potential outcomes, $Y_i(0)$ and $Y_i(1)$. For illustrative purposes, we consider the average effect

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

of the treatment in the super-population,

$$\tau_{\text{sp}} = \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0)\right].$$

In this section we restrict the discussion to the case with binary outcomes, $Y_i(0)$ and $Y_i(1) \in \{0, 1\}$ (some period of non-employment during the six years post lottery versus full-time employment during this period) to allow a sharper focus on the key conceptual issues.

We observe for unit i the treatment received, W_i, and the realized outcome, $Y_i^{\text{obs}} = Y_i(W_i)$. In the case without covariates, super-population unconfoundedness simply corresponds to independence of the treatment indicator and the potential outcomes:

$$W_i \perp\!\!\!\perp \big(Y_i(0), Y_i(1)\big).$$

Under random assignment we can unbiasedly estimate the average treatment effect as the difference in average observed outcomes by treatment status, which for the lottery data leads to:

$$\hat{\tau}^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} = 0.4106 - 0.5349 = -0.1243.$$

Using Neyman's approach (see Chapter 6), it follows that, if assignment were completely random, $\hat{\tau}^{\text{dif}}$ would be unbiased for both the finite-sample average treatment effect τ_{fs} and for the super-population average treatment effect τ_{sp}, with associated standard sampling variance estimate $\hat{\mathbb{V}}^{\text{neyman}} = 0.055^2$. We also calculate the exact Fisher p-value assuming complete randomization, using the difference in average outcomes for treated and control units as the statistic, leading to a p-value of 0.034.

Now suppose we do not wish to assume unconfoundedness and, moreover, we do not wish to make any alternative assumptions (but we maintain the stability assumption, SUTVA). What can we learn about τ_{sp} in the absence of this assumption? Manski's approach to this problem is as follows. Suppose we observe for all units in the super-population the treatment indicator W_i and the realized outcome Y_i^{obs}, $Y_i^{\text{obs}} = 1$ indicating employment every year versus $Y_i^{\text{obs}} = 0$ if individual i was unemployed for at least one year during the six-year post-lottery period. We can obtain method-of-moments estimates for, or using the terminology from the econometric literature, in large samples we can *identify*, three quantities. First, in the super-population share of treated units, $p = \mathbb{E}[W_i] = \Pr(W_i = 1)$, which, in this case without covariates, is also the propensity score for each unit. Second, we can similarly estimate the population distribution of $Y_i(0)$ conditional on $W_i = 0$. Because $Y_i(0)$ is binary, this distribution can be summarized by the scalar $\mu_{\text{c},0} = \Pr(Y_i(0) = 1|W_i = 0) = \mathbb{E}[Y_i(0)|W_i = 0]$. Finally, we can estimate the population distribution of $Y_i(1)$ given $W_i = 1$, $\mu_{\text{t},1} = \Pr(Y_i(1) = 1|W_i = 1) = \mathbb{E}[Y_i(1)|W_i = 1]$. In addition, define the super-population quantities

$$\mu_{\text{c},1} = \mathbb{E}\left[\,Y_i(0)|\,W_i = 1\right], \quad \text{and} \quad \mu_{\text{t},0} = \mathbb{E}\left[\,Y_i(1)|\,W_i = 0\right].$$

Note that if super-population unconfoundedness holds, then $\mu_{\text{c},1}$ and $\mu_{\text{t},0}$ are equal to

$$\mu_{\text{c},1} = \mu_{\text{c},0} = \mathbb{E}[Y_i(0)] = \mathbb{E}[Y_i^{\text{obs}}|W_i = 0],$$

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

and

$$\mu_{\mathrm{t},0} = \mu_{\mathrm{t},1} = \mathbb{E}[Y_i(1)] = \mathbb{E}[Y_i^{\mathrm{obs}} | W_i = 1],$$

respectively, so that under super-population unconfoundedness

$$\begin{aligned}\tau_{\mathrm{sp}} &= \mu_{\mathrm{t},1} - \mu_{\mathrm{c},0} = \mathbb{E}\left[\, Y_i(1) |\, W_i = 1\right] - \mathbb{E}\left[\, Y_i(0) |\, W_i = 0\right] \\ &= \mathbb{E}[Y_i^{\mathrm{obs}} | W_i = 1] - \mathbb{E}[Y_i^{\mathrm{obs}} | W_i = 0],\end{aligned}$$

and $\hat{\tau}^{\mathrm{dif}}$ is unbiased for τ_{sp}. Without the unconfoundedness assumption, however, we cannot infer τ_{sp} from only these three quantities, p, $\mu_{\mathrm{c},0}$, and $\mu_{\mathrm{t},1}$.

In general, without assuming unconfoundedness, we can rewrite τ_{sp} as the difference in the average of the potential outcomes,

$$\tau_{\mathrm{sp}} = \mu_{\mathrm{t}} - \mu_{\mathrm{c}},$$

where

$$\mu_{\mathrm{t}} = \mathbb{E}\left[Y_i(1)\right] = p \cdot \mu_{\mathrm{t},1} + (1-p) \cdot \mu_{\mathrm{t},0},$$

and

$$\mu_{\mathrm{c}} = \mathbb{E}\left[Y_i(0)\right] = p \cdot \mu_{\mathrm{c},1} + (1-p) \cdot \mu_{\mathrm{c},0}.$$

Without unconfoundedness (and without making any additional assumptions to replace it), the data are not informative about $\mu_{\mathrm{t},0}$ or $\mu_{\mathrm{c},1}$ beyond the obvious fact that, because the outcomes are binary, these quantities must lie inside the interval $[0, 1]$. These *natural bounds* on $\mu_{\mathrm{t},0}$ and $\mu_{\mathrm{c},1}$ imply bounds on μ_{t} and μ_{c}:

$$\mu_{\mathrm{c}} \in \Big[(1-p) \cdot \mu_{\mathrm{c},0}, (1-p) \cdot \mu_{\mathrm{c},0} + p\Big],$$

and

$$\mu_{\mathrm{t}} \in \Big[p \cdot \mu_{\mathrm{t},1}, p \cdot \mu_{\mathrm{t},1} + (1-p)\Big].$$

These ranges on μ_{t} and μ_{c} in turn imply bounds on the estimand, the population average effect τ_{sp}:

$$\tau_{\mathrm{sp}} \in \Big[p \cdot \mu_{\mathrm{t},1} - p - (1-p) \cdot \mu_{\mathrm{c},0}, p \cdot \mu_{\mathrm{t},1} + (1-p) - (1-p) \cdot \mu_{\mathrm{c},0}\Big]. \tag{22.1}$$

These bounds on the average treatment effect are *sharp*, in the sense that any value of τ_{sp} inside these bounds is consistent with the data if we are not assuming unconfoundedness. In other words, we cannot rule out, even in an infinitely large sample, any value inside these bounds. If we wish to obtain sharper inferences for τ_{sp}, we need to make stronger assumptions about the distribution of the potential outcomes, the assignment mechanism, or both. It is useful to see precisely why the bounds are sharp. Consider the upper bound in (22.1), $p \cdot \mu_{\mathrm{t},1} + (1-p) - (1-p) \cdot \mu_{\mathrm{c},0}$. What is the joint distribution of the potential outcomes and the assignment mechanism that would lead to this value for the average

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

treatment effect? In order for τ to be equal to this upper bound, it must be the case that $\mu_{t,0} = 1$ (i.e., all the units who received the control treatment would have positive earnings given the active treatment), and $\mu_{c,1} = 0$ (i.e., all the units receiving the active treatment would have zero earnings had they received the control treatment). Although such a scenario appears extreme, there is nothing in the data that formally rules out this possibility.

In this specific setting, the bounds are arguably not very informative. Note that without any data, we can infer from the fact that the outcomes are binary that the average effect τ_{sp} must lie in the interval $[-1, 1]$, with the width of that interval equal to two. The data, with everyone exposed to treatment or control, but without the unconfoundedness assumption, can narrow this range to Equation (22.1). Inspection of these bounds shows that they are of the form $[-c, 1-c]$ for some $c \in [0, 1]$. Thus, in this case, the bounds always have range one, and always include zero (corresponding to the Fisher null hypothesis of no effect of the treatment for any unit), irrespective of the data. The fact that the bounds must include zero follows immediately from the fact that nothing in our setup so far rules out the possibility that the treatment effect is zero for all units. The fact that the width of this bounding interval is always one follows from the fact that the width of the interval without the data is two, in combination with the fact that exactly half the potential outcomes are missing.

For the IRS lottery data, the fraction treated is $N_t/N = 0.4675$, and the fraction of individuals with positive earnings in the control and treatment groups are $\overline{Y}_c^{obs} = 0.5349$ and $\overline{Y}_t^{obs} = 0.4106$, respectively. Replacing p, $\mu_{t,1}$ and $\mu_{c,0}$ by N_t/N, $\overline{Y}_t^{obs}$, and $\overline{Y}_c^{obs}$, respectively, in Equation (22.1) leads to a lower and upper bound for the super-population average treatment effect, without additional assumptions, equal to:

$$\begin{aligned}\tau_{sp} &\in \Big[p \cdot \mu_{t,1} - p - (1-p) \cdot \mu_{c,0}, p \cdot \mu_{t,1} + (1-p) - (1-p) \cdot \mu_{c,0}\Big] \\ &= [-0.56, 0.44].\end{aligned}$$

22.4 BINARY OUTCOMES: THE ROSENBAUM-RUBIN SENSITIVITY ANALYSIS

Now let us study the same setting from a different perspective, the sensitivity analysis approach developed by Rosenbaum and Rubin (1983). Rosenbaum and Rubin start with the assumption that super-population unconfoundedness holds given an unobserved scalar covariate. Let us denote this unobserved covariate by U_i. Super-population unconfoundedness given this unobserved covariate, in the absence of observed covariates, requires that

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \;\big|\; U_i. \tag{22.2}$$

It is convenient, at least initially, to model U_i as binary with

$$q = \Pr(U_i = 1) = 1 - \Pr(U_i = 0).$$

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

Now let us build parametric models for the relations between the unobserved covariate U_i and both the treatment indicator and both potential outcomes. In principle we would like a model for

$$f(W_i, Y_i(0), Y_i(1)|U_i).$$

By Equation (22.2) W_i is independent of $(Y_i(0), Y_i(1))$ given U_i, so we can write this as

$$f(W_i|U_i) \cdot f(Y_i(0), Y_i(1)|U_i).$$

As discussed in Chapters 6 and 8, the data are not informative about the dependence structure between $Y_i(0)$ and $Y_i(1)$, so here, for simplicity, we model them as independent conditional on U_i. Thus, we need to specify models for $f(W_i|U_i)$, $f(Y_i(0)|U_i)$, and $f(Y_i(1)|U_i)$. We use the following specifications, taking into account the fact that $Y_i(0)$ and $Y_i(1)$ are binary:

$$\Pr(W_i = 1|U_i = u) = \frac{\exp(\gamma_0 + \gamma_1 \cdot u)}{1 + \exp(\gamma_0 + \gamma_1 \cdot u)},$$

$$\Pr(Y_i(1) = 1|U_i = u) = \frac{\exp(\alpha_0 + \alpha_1 \cdot u)}{1 + \exp(\alpha_0 + \alpha_1 \cdot u)},$$

and

$$\Pr(Y_i(0) = 1|U_i = u) = \frac{\exp(\beta_0 + \beta_1 \cdot u)}{1 + \exp(\beta_0 + \beta_1 \cdot u)},$$

where dependence on the parameters is notationally suppressed to avoid clutter.

There are seven scalar components of the parameter $\theta = (q, \gamma_1, \alpha_1, \beta_1, \gamma_0, \alpha_0, \beta_0)$, which we partition into two subvectors. The first, $\theta_s = (q, \gamma_1, \alpha_1, \beta_1)$, comprises the *sensitivity* parameters, which we do not attempt to estimate. Instead we postulate (ranges of) values for them *a priori*. We discuss later how we select the particular values, or rather the range of values for these parameters, but now we discuss how to proceed conditional on postulated values for these parameters. Conditional on values for the sensitivity parameters $(q, \gamma_1, \alpha_1, \beta_1)$, we estimate the remaining parameters, that is the *estimable parameters* $\theta_e = (\gamma_0, \alpha_0, \beta_0)$, from the data and infer the average treatment effect τ_{sp}. The approach has in common with the bounds approach that even in large samples we cannot reject any combination of values for $(q, \gamma_1, \alpha_1, \beta_1)$: the data do not lead to unbiased method-of-moment estimates of these parameters even in infinite samples, or, in the econometric terminology, these parameters are *not identified*.

Let us look at this argument in more detail. The data allow for unbiased estimates of $p = \mathbb{E}[W_i]$, $\mu_{t,1} = \mathbb{E}[Y_i^{obs}|W_i = 1] = \mathbb{E}[Y_i(1)|W_i = 1]$ and $\mu_{c,0} = \mathbb{E}[Y_i^{obs}|W_i = 0] = \mathbb{E}[Y_i(0)|W_i = 0]$. Let us take those parameters as known, and ignore for the moment the sampling variation in their estimates. These three estimable quantities relate to the parameters $(q, \gamma_1, \alpha_1, \beta_1)$ and $(\gamma_0, \alpha_0, \beta_0)$ through the three equalities

$$p = q \cdot \frac{\exp(\gamma_0 + \gamma_1)}{1 + \exp(\gamma_0 + \gamma_1)} + (1 - q) \cdot \frac{\exp(\gamma_0)}{1 + \exp(\gamma_0)}, \tag{22.3}$$

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

$$\begin{aligned}\mu_{\mathrm{t},1} &= \Pr(U_i = 1 | W_i = 1) \cdot \mathbb{E}[Y_i(1) | W_i = 1, U_i = 1] \\ &\quad + (1 - \Pr(U_i = 1 | W_i = 1)) \cdot \mathbb{E}[Y_i(1) | W_i = 1, U_i = 0] \\ &= \frac{q \cdot \frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)}}{q \cdot \frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)} + (1-q) \cdot \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}} \cdot \frac{\exp(\alpha_0+\alpha_1)}{1+\exp(\alpha_0+\alpha_1)} \\ &\quad + \frac{(1-q) \cdot \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}}{q \cdot \frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)} + (1-q) \cdot \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}} \cdot \frac{\exp(\alpha_0)}{1+\exp(\alpha_0)},\end{aligned} \tag{22.4}$$

and

$$\begin{aligned}\mu_{\mathrm{c},0} &= \frac{q \cdot \frac{1}{1+\exp(\gamma_0+\gamma_1)}}{q \cdot \frac{1}{1+\exp(\gamma_0+\gamma_1)} + (1-q) \cdot \frac{1}{1+\exp(\gamma_0)}} \cdot \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} \\ &\quad + \frac{(1-q) \cdot \frac{1}{1+\exp(\gamma_0)}}{q \cdot \frac{1}{1+\exp(\gamma_0+\gamma_1)} + (1-q) \cdot \frac{1}{1+\exp(\gamma_0)}} \cdot \frac{\exp(\beta_0)}{1+\exp(\beta_0)}.\end{aligned} \tag{22.5}$$

It is straightforward to see that for all values of $(q, \gamma_1, \alpha_1, \beta_1)$, and for all distributions of the observed data (captured by the values for the triple $(p, \mu_{\mathrm{t},1}, \mu_{\mathrm{c},0})$), we can find values for the triple $(\gamma_0, \alpha_0, \beta_0)$ such that all three of these equalities hold. Moreover, these values for the estimable parameters $(\gamma_0, \alpha_0, \beta_0)$ are unique for all values of $\mu_{\mathrm{c},0}, \mu_{\mathrm{t},1}$, and p, and for all values of the sensitivity parameters. For example, If $\gamma_0 \to -\infty$, the right-hand side of the first equality goes to zero, and if $\gamma_0 \to \infty$, the right-hand side goes to one. Because the right-hand side is strictly increasing in γ_0, there must be a unique value such that (22.3) holds for any $p \in (0, 1)$. Let us write these implied values for $(\gamma_0, \alpha_0, \beta_0)$ as functions of the data and $(q, \gamma_1, \alpha_1, \beta_1)$:

$$\gamma_0(q, \gamma_1, \alpha_1, \beta_1 | \text{data}), \qquad \alpha_0(q, \gamma_1, \alpha_1, \beta_1 | \text{data}),$$

and

$$\beta_0(q, \gamma_1, \alpha_1, \beta_1 | \text{data}),$$

where, ignoring sampling variation, the data consist of the triple

$$\text{data} = (p, \mu_{\mathrm{t},1}, \mu_{\mathrm{c},0}).$$

Given the postulated values for $(q, \gamma_1, \alpha_1, \beta_1)$, and given the values for $(\gamma_0, \alpha_0, \beta_0)$ that are implied by the combination of the data and the postulated values for $(q, \gamma_1, \alpha_1, \beta_1)$, there are implied values for $\mu_{\mathrm{t},0}$ and $\mu_{\mathrm{t},1}$. In terms of $\theta = (q, \gamma_1, \alpha_1, \beta_1, \gamma_0, \alpha_0, \beta_0)$, we can write

$$\begin{aligned}\mu_{\mathrm{t},0}(q, \gamma_1, \alpha_1, \beta_1, \gamma_0, \alpha_0, \beta_0) &= \mathbb{E}[Y_i(1) | W_i = 0] \\ &= \frac{q \cdot \frac{1}{1+\exp(\gamma_0+\gamma_1)}}{q \cdot \frac{1}{1+\exp(\gamma_0+\gamma_1)} + (1-q) \cdot \frac{1}{1+\exp(\gamma_0)}} \cdot \frac{\exp(\alpha_0+\alpha_1)}{1+\exp(\alpha_0+\alpha_1)} \\ &\quad + \frac{(1-q) \cdot \frac{1}{1+\exp(\gamma_0)}}{q \cdot \frac{1}{1+\exp(\gamma_0+\gamma_1)} + (1-q) \cdot \frac{1}{1+\exp(\gamma_0)}} \cdot \frac{\exp(\alpha_0)}{1+\exp(\alpha_0)},\end{aligned}$$

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

and

$$\begin{aligned}\mu_{\mathrm{c},1}(q,\gamma_1,\alpha_1,\beta_1,\gamma_0,\alpha_0,\beta_0) &= \mathbb{E}[Y_i(0)|W_i=1]\\ &= \frac{q\cdot\frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)}}{q\cdot\frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)}+(1-q)\cdot\frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}}\cdot\frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)}\\ &+\frac{(1-q)\cdot\frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}}{q\cdot\frac{\exp(\gamma_0+\gamma_1)}{1+\exp(\gamma_0+\gamma_1)}+(1-q)\cdot\frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}}\cdot\frac{\exp(\beta_0)}{1+\exp(\beta_0)},\end{aligned}$$

where the conditioning on parameters is notationally suppressed. Then, finally, we can write the average treatment effect τ_{sp} as

$$\tau_{\mathrm{sp}} = \mu_{\mathrm{t}} - \mu_{\mathrm{c}} = p\cdot(\mu_{\mathrm{t},1}-\mu_{\mathrm{c},1}) + (1-p)\cdot(\mu_{\mathrm{t},0}-\mu_{\mathrm{c},0}).$$

In summary, given the (super-population) data $= (p, \mu_{\mathrm{t},1}, \mu_{\mathrm{c},0})$, there is a function that gives τ_{sp} as a function of $(p, \mu_{\mathrm{t},1}, \mu_{\mathrm{c},0})$ and the sensitivity parameters:

$$\tau_{\mathrm{sp}} = \tau(q,\gamma_1,\alpha_1,\beta_1|\mathrm{data}) = \tau(q,\gamma_1,\alpha_1,\beta_1|p,\mu_{\mathrm{t},1},\mu_{\mathrm{c},0}). \tag{22.6}$$

It is this function in which we are interested. Given the data, that is, for fixed values for $(p, \mu_{\mathrm{t},1}, \mu_{\mathrm{c},0})$, we wish to inspect how sensitive the average treatment effect τ_{sp} is to assumptions about the sensitivity parameters $(q, \gamma_1, \alpha_1, \beta_1)$.

There are two special sets of values for the sensitivity parameters. First, if we fix $\gamma_1 = 0$, then we are back assuming unconfoundedness (or a completely randomized experiment in this case without covariates), and

$$\tau_{\mathrm{sp}} = \mu_{\mathrm{t},1} - \mu_{\mathrm{c},0}.$$

The same holds if we fix both $\alpha_1 = \beta_1 = 0$. Note that it is not necessary that both $\gamma_1 = 0$, *and* $\alpha_1 = \beta_1 = 0$, for there to be no bias associated with estimates based on unconfoundedness ignoring U_i. It is sufficient if (*a*) the unobserved covariate does not affect assignment ($\gamma_1 = 0$), or (*b*) the unobserved covariate is not associated with either potential outcome ($\alpha_1 = \beta_1 = 0$).

Second, suppose we fix $q = p$, and let $\gamma_1 \to \infty$. In that case W_i and U_i become perfectly correlated. If we also let $\alpha_1 \to -\infty$ and $\beta_1 \to -\infty$, then

$$\tau_{\mathrm{sp}} \to p\cdot\mu_{\mathrm{t},1} + (1-p) - (1-p)\cdot\mu_{\mathrm{c},0},$$

which equals to the upper limit in the Manski bounds, showing that the setup with unconfoundedness given an unobserved binary covariate is not technically restrictive in this sense. Similarly, if we again fix $q = p$, and let $\gamma_1 \to \infty$, $\alpha_1 \to \infty$, and $\beta_1 \to \infty$, then

$$\tau_{\mathrm{sp}} \to p\cdot\mu_{\mathrm{t},1} - p - (1-p)\cdot\mu_{\mathrm{c},0},$$

which is equal to the lower limit in the Manski bounds. This demonstrates that, in this setting, the bounds analysis can be viewed as an extreme version of a sensitivity analysis, or

taking the opposite perspective, the sensitivity analysis can be viewed as a generalization of the bounds analysis.

Outside of these special values, the key question concerns the set of reasonable values for the sensitivity parameters $\theta_s = (q, \gamma_1, \alpha_1, \beta_1)$. Given a set of reasonable values Θ_s for θ_s, we calculate a lower and upper bound of the average treatment effect τ_{sp} over that set,

$$\tau_{\text{low}} = \inf_{(q,\gamma_1,\alpha_1,\beta_1)\in\Theta_s} \tau(q, \gamma_1, \alpha_1, \beta_1 | p, \mu_{\text{t},1}, \mu_{\text{c},0}),$$

and

$$\tau_{\text{high}} = \sup_{(q,\gamma_1,\alpha_1,\beta_1)\in\Theta_s} \tau(q, \gamma_1, \alpha_1, \beta_1 | p, \mu_{\text{t},1}, \mu_{\text{c},0}),$$

leading to the range

$$\tau_{\text{sp}} \in \left[\tau_{\text{low}}, \tau_{\text{high}}\right].$$

We generally do not have any substantive judgment regarding q, and one could simply investigate all values for q. Often results are not sensitive to intermediate values for q, and q can be taken to be equal to $\mathbb{E}[W_i] = p$. The remaining parameters are more interesting. The sensitivity parameter γ_1 represents the effect on the log odds ratio of receiving the treatment of a change from $U_i = 0$ to $U_i = 1$. In cases where the researcher has specific variables in mind that could bias the results based on assuming unconfoundedness, this can be a meaningful, interpretable parameter. In specific cases, one could be able to make an informed judgment about reasonable values for this parameter. Note that the Manski bounds on τ_{sp} implicitly allow U_i to be perfectly correlated with the receipt of treatment W_i, corresponding to $\gamma_1 \to \infty$. In settings where researchers have attempted to record all relevant determinants of treatment assignment, such a correlation may be viewed as logically too extreme. On the other hand, it may be difficult to specify a number that would meet with widespread agreement as a bound for α_1, and our preferred strategy is therefore to report the sensitivity of τ_{sp} to changes in these parameters.

One specific strategy, in cases where covariates are available, is to consider the association between the observed covariates and both the treatment assignment and the potential outcomes, assuming unconfoundedness, and use the estimated associations as indicative of ranges of reasonable values for the association between the unobserved covariate and the treatment indicator and the potential outcomes.

We illustrate this strategy with the lottery data where we observe eighteen covariates. For each covariate we estimate two logistic regression models. Denote the k^{th} covariate, after normalizing by its standard deviation, by X_{ik}. We estimate a model for the treatment indicator conditional on the covariate,

$$\Pr(W_i = 1 | X_{ik}) = \frac{\exp(\delta_{k0} + \delta_{k1} \cdot X_{ik})}{1 + \exp(\delta_{k0} + \delta_{k1} \cdot X_{ik})},$$

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

and another model for the outcome conditional on the covariate and the treatment indicator,

$$\Pr(Y_i^{\text{obs}} = 1|W_i, X_{ik}) = \frac{\exp(\zeta_{k0} + \zeta_{k1} \cdot X_{ik} + \zeta_{k2} \cdot W_i)}{1 + \exp(\zeta_{k0} + \zeta_{k1} \cdot X_{ik} + \zeta_{k2} \cdot W_i)},$$

again with dependence on parameters notationally suppressed.

Estimating these two models for each covariate X_{ik} leads to eighteen estimates $\hat{\delta}_{k1}$ and $\hat{\zeta}_{k2}$, $k = 1, \ldots, 18$. The largest values, in absolute value, were $|\hat{\delta}_{2,1}| = 0.56$ (the association between the number of tickets bought and winning the lottery) and $|\hat{\zeta}_{18,1}| = 1.61$ (the association between the indicator for positive earnings in the year prior to winning the lottery and post-lottery employment). We use these two values to anchor the sensitivity parameters. The idea is to limit the association between the unobserved binary covariate U_i and the treatment indicator and potential outcomes by assuming that they are bounded by the strongest marginal associations of the observed covariates. If one has made a good-faith effort to collect all relevant covariates, it may be difficult to see how one would miss covariates more important than any of those observed, at least unless there are specific reasons, such as confidentiality concerns. If $q = 1/2$, the standard deviation of U_i is 1/2, so we implement the sensitivity analysis by letting γ_1 range over the interval $[-0.56/(1/2), 0.56/(1/2)] = [-1.12, 1.12]$ and α_1 and β_1 over the interval $[-1.61/(1/2), 1.61/(1/2)] = [-3.22, 3.22]$ (multiplying the maximum of the coefficients by a factor two, equal to the ratio of the standard deviation of the normalized covariates and the maximum standard deviation of U, to take account of the normalization of the covariates). We let q range over the interval $[0, 1]$ because there is no substantive argument to restrict its range. Choosing values for the sensitivity parameters in this range leads to values for the average treatment effect in the interval

$$\tau_{\text{sp}} \in \big[-0.28, 0.05\big] \,\big|\, \{q \in [0, 1], \gamma_1 \in [-1.12, 1.12], \alpha_1 \in [-3.22, 3.22], \beta_1 \in [-3.22, 3.22]\}.$$

Substantively this range suggests that the unobserved covariate would have to be fairly strongly associated with both treatment and potential outcomes to change the conclusion in the lottery example that the treatment has a positive and substantial effect on employment. We do not know whether such a covariate exists, but it would have to be somewhat stronger than any of the covariates the researchers managed to collect in terms of its association with the outcome and the treatment indicator. Note that in these calculations we allow γ_1 to be as large as the effect of any observed covariate on the log odds ratio for receiving the treatment, and simultaneously allow α_1 and β_1 to be as large as the effect of any observed covariates on the log odds ratio for the potential outcome. No single covariate in the sample had such strong effects on both simultaneously. In fact, for each covariate separately, the range of values for the average treatment effect τ associated with letting $q \in [0, 1]$, $\gamma_1 \in [-2 \cdot |\hat{\delta}_{k1}|, 2 \cdot |\hat{\delta}_{k1}|]$, and $\alpha_1, \beta_1 \in [-2 \cdot |\hat{\zeta}_{k1}|, 2 \cdot |\hat{\zeta}_{k1}|]$, for some $k = 1, \ldots, 18$, the widest range we find for the average treatment effect is $[-0.18, -0.07]$, with these values corresponding to the 12^{th} covariate, earnings in the year prior to winning the lottery, with $\hat{\delta}_{12,1} = -0.1891$ and $\hat{\zeta}_{12,1} = 1.3257$.

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

Another approach for assessing the sensitivity that does not directly require us to postulate reasonable values for the sensitivity parameters is to explore the magnitude necessary for $(\gamma_1, \alpha_1, \beta_1)$ in order to change the sign for the estimated average treatment effect found under unconfoundedness. There are trade-offs between the parameters, because with a larger value for γ_1, the required values for α_1 and β_1 will not be quite as large. For example, and this is also useful for a subsequent comparison with the Rosenbaum sensitivity analysis, it is also interesting to look at the range of values for the average treatment effect given that $|\gamma_1| \leq 0.52$, with α_1 and β_1 essentially unrestricted. Then,

$$\tau_{\text{sp}} \in [-0.22, 0.48] \,|\{q \in [0,1], \gamma_1 \in [-0.52, 0.52], \alpha_1 \in (-\infty, \infty), \beta_1 \in (-\infty, \infty)\},$$

just on the margin where the sign of the average treatment effect switches.

22.5 BINARY OUTCOMES: THE ROSENBAUM SENSITIVITY ANALYSIS FOR P-VALUES

Rosenbaum (1995) is interested in calculating Fisher p-values under the sharp null hypothesis of no treatment effects and wishes to assess how sensitive the conclusions under unconfoundedness are to that assumption. In principle applying these methods requires knowledge of the propensity score. Although we do not know the propensity score in observational studies, under unconfoundedness we can estimate the propensity score for each unit. Let these estimated propensity score values be denoted by $\hat{e}_i$. Given these values, we can use Fischer's exact p-value approach to obtain p-values for the null hypothesis of no effect whatsoever of the treatment. In the IRS lottery data, still without covariates, using the difference in average ranks as the statistic, the p-value is 0.034. Assuming random assignment, we can be very confident that the treatment has some effect on employment.

Now suppose unconfoundedness does not hold. In that case it is no longer the case that the estimated probability of the treatment is $\hat{e}_i$, where $\hat{e}_i$ was estimated under the assumption of unconfoundedness. Let us denote the actual treatment probability by p_i. Rosenbaum then limits the difference between the actual probability p_i and the estimated probability under unconfoundedness $\hat{e}_i$. Specifically, he assumes that the difference in log odds ratios, under the assumption of unconfoundedness, and based on the true assignment probabilities, is bounded by a pre-specified constant Γ:

$$\left| \ln\left(\frac{\hat{e}_i}{1-\hat{e}_i}\right) - \ln\left(\frac{p_i}{1-p_i}\right) \right| \leq \Gamma, \tag{22.7}$$

for all $i = 1, \ldots, N$. We can relate this to the analysis in the previous subsection by specifying a model for the treatment assignment as a function of an unobserved binary covariate u:

$$p_i = \Pr(W_i = 1 | U_i = u) = \frac{\exp(\gamma_0 + \gamma_1 \cdot u)}{1 + \exp(\gamma_0 + \gamma_1 \cdot u)}, \tag{22.8}$$

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

so that the logarithm of the true odds ratio is $\ln(p_i/(1-p_i)) = \gamma_0 + \gamma_1 \cdot u$. If we approximate the average propensity score, averaged over the distribution of U_i, by the propensity score at the average value of U_i, $q = \mathbb{E}[U_i]$, so that $e_i = \exp(\gamma_0 + \gamma_1 \cdot q)/(1 + \exp(\gamma_0 + \gamma_1 \cdot q))$, the implied logarithm of the odds ratio is $\ln(e_i/(1-e_i)) = \gamma_0 + \gamma_1 \cdot q$. The difference between the log odds ratio for the average propensity score under unconfoundedness and the log odds ratio for the true treatment probability is then $\ln(e_i/(1-e_i)) - \ln(p_i/(1-p_i)) = \gamma_1 \cdot (q - U_i)$. The Rosenbaum restriction implies we should consider all possible values for (q, γ_1) such that

$$q \cdot |\gamma_1| < \Gamma, \quad \text{and} \quad (1-q) \cdot |\gamma_1| < \Gamma.$$

We can simplify the problem in this context by allowing for all possible values for q in the interval $[0, 1]$, and all possible values for γ_1 such that $|\gamma_1| < \Gamma$, thus requiring the difference in log odds ratios for units with $U_i = 1$ and units with $U_i = 0$ to be restricted to

$$\left| \ln\left(\frac{\Pr(W_i = 1|U_i = 1)}{1 - \Pr(W_i = 1|U_i = 1)} \right) - \ln\left(\frac{\Pr(W_i = 1|U_i = 0)}{1 - \Pr(W_i = 1|U_i = 0)} \right) \right| = \gamma_1 \leq \Gamma.$$

The question we now address is, given that we restrict γ_1 but place no restrictions on q, what is the evidence in the data against the null hypothesis of no effect whatsoever of the treatment? It is immediately clear that without any restriction on γ_1 there is no evidence against the null hypothesis that there is no effect of the treatment: if we let $\gamma_1 \to \infty$, then W_i and U_i are perfectly correlated.

Let us consider a particular statistic. In this case with a binary outcome, the natural statistic is the difference in means, $T^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}$. The value of this statistic for the lottery data is -0.12, with an exact Fisher p-value for the null hypothesis of no effects, calculated under complete random assignment, equal to 0.034. To make the comparison with the Rosenbaum sensitivity analysis easier, it is useful to change the assignment mechanism slightly; from a completely randomized experiment to a Bernoulli experiment with assignment probability 0.47, the p-value changes to 0.026. Now, pick a particular value for (q, γ_1). Given these values, the probability of receiving the treatment for unit i can be either

$$p_{\text{low}} = \frac{\exp(\gamma_0)}{1 + \exp(\gamma_0)}, \quad \text{or} \quad p_{\text{high}} = \frac{\exp(\gamma_0 + \gamma_1)}{1 + \exp(\gamma_0 + \gamma_1)},$$

with the first probability corresponding to the case where unit i has $U_i = 0$, and the second corresponding to the case where unit i has $U_i = 1$. Now suppose we assign each unit a value for U_i, and thus implicitly assign the unit a value for the assignment probability. Denote this assignment probability for unit i by p_i. Given that assignment probability, we can calculate the p-value for any statistic under its randomization distribution. The statistic we focus on is the difference in average outcomes by treatment status, $T^{\text{dif}} = \bar{Y}_{\text{t}}^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}}$. The fact that the assignment probabilities are not all equal does not create any problems when calculating or simulating the p-values.

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

The question now is what the most extreme (and in particular what the largest) p-value is we can find by assigning the unobserved covariate U_i to each unit for a given value of γ_1, allowing q to range over the interval $[0, 1]$. We can again turn to the associations between covariates and the treatment indicator to find a possibly reasonable value for γ_1. The largest value we found for δ_{k1}, which captures the relationship between an observed covariate (normalized to have unit variance) and the treatment indicator was approximately 0.56. (Recall that this corresponds to the number of lottery tickets bought.) This suggests that limiting γ_1 to be less than or equal to $2 \cdot 0.56 = 1.12$ (where the factor 2 captures the fact that the standard deviation of the binary covariate U is bounded by 1/2) may present a reasonable range of values for γ_1. This changes the p-value from 0.026 to 0.99, suggesting that such an association between the treatment indicator and the unobserved covariate eliminates any evidence of a negative effect of the treatment. Instead, using the δ_k for earnings, 0.19, to bound γ to less than 0.38 in absolute value leads to p-value of 0.27. Finally, using an upper bound on γ_1 equal to 0.52 leads to a p-value equal to 0.50.

22.5.1 The Rosenbaum Sensitivity Analysis for Average Treatment Effects

It is instructive, for the purpose of understanding the similarities and differences between the two approaches to sensitivity analyses, to modify Rosenbaum's approach to derive a range of feasible values for the average treatment effect. Instead of looking at the p-values associated with a pair of values for (q, γ_1), we again look at a range of values for the average treatment effect. Using the derivations from Section 22.4, we look at the range of values for τ if we allow $q \in [0, 1]$, $\gamma_1 \in [-\Gamma, \Gamma]$. In addition, we allow $\alpha_1 \in (-\infty, \infty)$ and $\beta_1 \in (-\infty, \infty)$, which reveals how the Rosenbaum sensitivity approach differs from the Cornfield-Rosenbaum-Rubin method for assessing sensitivity. In the latter we restrict α_1 and β_1, in addition to γ_1, whereas the former approach only restricts γ_1. This modification obviously leads to a wider range of possible values for τ. Restricting γ_1 to be less than 1.12 in absolute value, without restricting α_1 or β_1, leads to a range of possible values for the average effect of the treatment equal to

$$\tau_{\text{sp}} \in [-0.62, 0.48] \,|\{q \in [0, 1], \gamma_1 \in [-1.12, 1.12], \alpha_1 \in (-\infty, \infty), \beta_1 \in (-\infty, \infty)\},$$

considerably wider than the values we found before when we also restricted α_1 and β_1.

It is also interesting to restrict γ_1 to be less than 0.52 in absolute value, without restricting α_1 or β_1. This leads to a range of possible values for the average effect of the treatment equal to

$$\tau_{\text{sp}} \in [-0.22, 0.48] \,|\{q \in [0, 1], \gamma_1 \in [-0.52, 0.52], \alpha_1 \in (-\infty, \infty), \beta_1 \in (-\infty, \infty)\}.$$

Now the set of estimates (ignoring sampling uncertainty) has zero as its upper limit. This corresponds to the case where the upper bound on the p-values is equal to 0.50.

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press

22.6 CONCLUSION

In this chapter we present methods for assessing the sensitivity of results obtained under unconfoundedness. The unconfoundedness assumption can be controversial, and the analyses discussed here allow the researcher to quantify how much the estimates and p-values rely on the full force of this assumption. Finding that particular results are, or are not, sensitive to this assumption helps evaluate the results of any analysis under unconfoundedness.

However, in our limited experience, the application and value of such sensitivity analyses depend rather critically on the context of the study and general scientific knowledge that the investigators can bring to bear on the problem at hand.

NOTES

The key papers underlying the first sensitivity analysis in this chapter are Cornfield et al. (1959) and Rosenbaum and Rubin (1983a). Rosenbaum and Rubin focus on the case with a binary outcome, where, in their example, the sample is divided into five subclasses or blocks, and use the analysis where the sensitivity parameters are restricted to the same values in each block. They directly limit the values of the sensitivity parameters γ_1, α_1, and β_1 to be less than or equal to three in absolute value. Rosenbaum (1995, 2002) developed the sensitivity analysis that restricts only the assignment probabilities. Imbens (2003) applies the Rosenbaum-Rubin sensitivity analysis and is the original source for the suggestion to anchor the thresholds to values based on the association between treatment and observed covariates and between the outcomes and observed values. For another recent application, see Ichino, Mealli, and Nannichini (2008).

Manski (1990, 1996, 2003, 2013) in a series of papers proposed calculating worst-case bounds of the type discussed in this chapter, with earlier results for special cases in Cochran (1977). Manski, Sandefur, McLanahan, and Powers (1992) present an early application.

https://doi.org/10.1017/CBO9781139025751.023 Published online by Cambridge University Press