CHAPTER 23

Instrumental Variables Analysis of Randomized Experiments with One-Sided Noncompliance

23.1 INTRODUCTION

In this chapter we discuss a second approach to analyzing causal effects when unconfoundedness of the treatment of interest is questionable. In Chapter 22 we also relaxed the unconfoundedness assumption, but there we did not make any additional assumptions. The resulting sensitivity and bounds analyses led to a range of estimated values for treatment effects, all of which were consistent with the observed data. Instead, in this chapter we consider alternatives to the standard unconfoundedness assumption that still allow us to obtain essentially unbiased point estimates of some treatment effects of interest, although typically *not* the overall average effect. In the settings we consider, there is, on substantive grounds, reason to believe that units receiving and units not receiving the treatment of interest are systematically different in characteristics associated with the potential outcomes. Such cases may arise if receipt of treatment is partly the result of deliberate choices by units, choices that take into account perceptions or expectations of the causal effects of the treatment based on information that the analyst may not observe. In order to allow for such violations of unconfoundedness, we rely on the presence of additional information and consider alternative assumptions regarding causal effects. More specifically, a key feature of the Instrumental Variables (IV) approach, the topic of the current chapter and the next two, is the presence of a secondary treatment, in the current setting the *assignment* to treatment instead of the *receipt* of treatment, where by "secondary" we do not mean temporily but secondary in terms of scientific interest. This secondary treatment is assumed to be unconfounded. In fact, in the randomized experiment setting of the current chapter, the assignment to treatment is unconfounded by design. This implies we can, using the methods from Part II of the book, unbiasedly estimate causal effects of the *assignment* to treatment. The problem is that these causal effects are not the causal effects of primary interest, which are the effects of the receipt of treatment. Assumptions that allow researchers to link these causal effects are at the core of the instrumental variables approach.

This chapter is the first of three chapters on instrumental variables approaches. For readers unfamiliar with this terminology, instrumental variables methods refer to a set of techniques originating in the econometrics literature, starting in the 1920s with work by Wright (1927), Tinbergen (1930), and later Haavelmo (1943). A central role in these

methods is played by a variable, the so-called *instrument* or *instrumental variable*, which is a variable known *a priori* almost certainly to have a causal effect on the treatment of primary interest, W_i. The key characteristic of this instrument, here denoted by Z_i, is the *a priori* assumed absence of a "direct" causal effect of the instrument on the outcome of interest Y_i, with any causal effect of Z_i on Y_i "passing through" a causal effect of the instrument on the treatment W_i, where these terms will become clear shortly. More generally, *principal stratification* refers to settings with *latent unconfoundedness* of the primary treatment, where, conditional on an only partially observed covariate, unconfoundedness holds. In the special case of instrumental variables, this latent unconfoundedness applies with the latent compliance status to assigned secondary treatment, more precisely defined later, playing the role of the partially unobserved covariate.

We start this instrumental variables discussion in the simplest setting of a completely randomized experiment with one-sided noncompliance. By noncompliance we refer to the situation where some units who are assigned to receive a particular treatment level do not comply with their assignment and instead receive an alternative treatment. In this chapter, compliance is assumed to be all or nothing: units cannot receive, or be exposed to, only part of the treatment. By one-sided, we mean that the noncompliance is asymmetric in the sense that only units assigned to receive the active treatment can potentially circumvent their assigned treatment and receive the control treatment. In contrast, all units assigned to receive the control treatment do, in fact, comply with this assignment. This type of noncompliance is common in settings with individual people as the units of analysis, where receipt of the active treatment requires individuals to take, or subject themselves to, a particular action, such as undergoing surgery or entering a job-training program. In such cases, it is often difficult, or even impossible, to compel individuals to undergo the active treatment if assigned to it, even if individuals give consent prior to the randomization. As a result, compliance among those assigned to the active treatment is often imperfect. In contrast, those assigned to receive the control treatment can often effectively be denied access to the active treatment, so the noncompliance is one-sided. In this setting, the assignment to treatment is the instrument Z_i, and the receipt of treatment is the treatment of primary interest W_i.

Many traditional formal statistical analyses of randomized experiments with noncompliance focus on the relation between the random assignment and the outcome of interest, discarding entirely any information about the treatment, in the current setting actually received, that is, ignoring W_i. Such an approach is generally referred to as an *intention-to-treat* (ITT) analysis. In our setting of a completely randomized experiment, ITT analyses are validated by the randomization of the assignment to treatment, without the need for additional assumptions beyond SUTVA. The main drawback of these ITT analyses is that they do not answer questions about causal effects of the receipt of treatment itself, only about causal effects of the assignment to treatment. Two other simple analyses, focusing directly on causal effects of the treatment of interest, but neither of which is generally valid, are sometimes conducted in such settings. First, *per protocol* analyses, where units that are observed not to comply with the treatment assigned are discarded (i.e., units with $Z_i \neq W_i$), and the data for all units who are observed to comply with their assigned treatment (i.e., units with $Z_i = W_i$) are analyzed as if they came from a randomized experiment with full compliance; that is, the analysis is as if W_i were randomized for units who appear to comply, discarding units who are observed to be

noncompliers. A second simple alternative is an *as-treated* analysis where data from all units are analyzed as if they had been randomly assigned to the treatment they actually received, ignoring information on assignment Z_i, and simply comparing treated units having $W_i = 1$ with control units having $W_i = 0$, as if W_i were randomized for all units. Both of these naive analyses are generally invalid as we discuss in Section 23.9.

In this chapter we focus on defining causal estimands and on the additional assumptions that allow us to go beyond the global effect of assignment that is the focus of ITT analyses, and estimate "local" average effects for the treatment of interest, that is, averages for subsets of units. Although we briefly mention some traditional econometric, moment-based, estimators for simple cases with no covariates, we leave the main discussion of our preferred model-based estimators and inference to Chapter 25.

In order to obtain alternatives to the assumption of unconfoundedness of the receipt of the treatment, we consider separately the nature of the noncompliance and the causal effects of the assignment to treatment for what we will call *compliers* and *noncompliers*. These groups are defined by their partly unobserved compliance behavior, and thus define *latent strata*. A key insight is that, although unconditionally receipt of treatment is confounded, within these latent strata the receipt of treatment is unconfounded. We then consider assumptions that rule out effects of assignment to the treatment on outcomes for certain groups but allow for general differences between units who comply and those who do not comply with their treatment assignment. Assessment of the plausibility of these assumptions relies heavily on subject-matter knowledge, in addition to the design of the assigned treatment.

In general there are two key assumptions justifying instrumental variables approaches. The first is that, although the receipt of the treatment is generally confounded when noncompliance occurs, the assignment to the treatment *is* unconfounded. As a result of unconfoundedness, we can estimate the effect of the assignment to treatment on both the outcome of interest, and on the receipt of treatment, that is, the two ITT effects. The unconfoundedness of assignment assumption is satisfied by design in the completely randomized experiment setting considered in this chapter, although in other applications of IV methods, this assumption can be controversial. The second key assumption is that the assignment to treatment has no effect on the final outcome of interest for those units whose receipt of treatment is unaffected by the assignment. For instance, for those who do not take the drug even when assigned to take it, the assignment itself is assumed to have no effect on the final outcome. We refer to this assumption as an *exclusion restriction*, because the instrument is excluded from affecting the outcome of interest for noncompliers. This assumption can be justified by design, for example, using double-blind experiments, where neither the unit nor the physician knows which treatment was assigned, thereby supporting the exclusion restriction. The key result in this chapter is that the exclusion assumption, when combined with the unconfoundedness assumption, enables us to estimate causal effects of the assignment to treatment on the principal outcome, Y_i, for the subpopulation of compliers, known as the *local average treatment effect* (LATE) or the *complier average causal effect* (CACE). The estimand, the average effect for compliers, is equal to the ratio of the ITT effect of Z_i on the outcome of interest, Y_i, and the ITT effect of Z_i on the receipt of treatment W_i. In other words, under the exclusion restrction, the ITT effect of assignment on the outcome of interest is due entirely to those units for whom receipt of treatment W_i is always identical to the

assignment to treatment Z_i, irrespective of their assignment. In many cases, it may then be reasonable to attribute the causal effect of *assignment* for the compliers to the causal effect of the *receipt* of treatment, the same way researchers often do, typically implicitly, in completely randomized experiments with full compliance.

We must emphasize from the outset that the assumptions underlying the instrumental variables approach, most importantly various forms of exclusion restrictions, are often controversial. When appropriate, these assumptions allow the researcher to make more interesting, and stronger, inferences than those obtained from ITT analyses. However, these assumptions are not always appropriate. Moreoever, unlike the unconfoundedness assumption, the validity of the exclusion restriction cannot be guaranteed solely by physical randomization, requiring in addition double blinding. Therefore, like SUTVA, its validity often relies on subject-matter knowledge.

The rest of the chapter is organized as follows. In Section 23.2 we describe the data set that will be used to illustrate the theoretical concepts introduced in this chapter. Next, in Section 23.3 we extend the potential outcomes notation to account for the instrumental variables setup. In the following section, Section 23.4, we analyze intention-to-treat effects. We define compliance behavior in Section 23.5. In Section 23.6 we discuss the instrumental variables estimand. In Section 23.7 we briefly discuss traditional moment-based estimators for the instrumental variables estimand. Then, in Section 23.8 we relate the discussion to traditional, linear-model-based instrumental variables methods. In Section 23.9 we discuss three naive methods for analyzing data from a randomized experiment with one-sided noncompliance. Section 23.10 concludes.

23.2 THE SOMMER-ZEGER VITAMIN A SUPPLEMENT DATA

We illustrate the methods discussed in this chapter using data previously analyzed by Sommer and Zeger (1991). Sommer and Zeger study the effect of vitamin A supplements on infant mortality in Indonesia. The vitamin supplements were randomly assigned to villages, but some of the individuals in villages assigned to the treatment group failed to receive them. None of the individuals assigned to the control group received the supplements, so noncompliance is one-sided. In this study, outcomes are observed for $N = 23{,}682$ infants. The observed outcome of interest, denoted by Y_i^{obs}, is a binary variable, indicating survival of an infant. Receipt of the vitamin supplements, which is considered the treatment of interest, is denoted by $W_i^{\text{obs}} \in \{0, 1\}$. In a slight departure from the notation in previous chapters, we add here the superscript "obs" to W_i for reasons that will become apparent later. Assignment to the supplements, the instrument, is denoted by $Z_i \in \{0, 1\}$. This assignment varies only at the village level. We ignore the clustering of the assignment at the village level because we do not have indicators for villages; this will tend to lead us to understate standard errors.

With all three observed variables binary, there are, in principal, eight different possible values for the triple $(Z_i, W_i^{\text{obs}}, Y_i^{\text{obs}})$. Because of the noncompliance, there may be units with $Z_i \neq W_i^{\text{obs}}$, but because $Z_i = 0$ implies $W_i^{\text{obs}} = 0$, there are only six values of the triple with positive counts in our sample. Table 23.1 contains the counts of the six observed values for the triple in the data set, with a total sample size of $N = 23{,}682$.

Table 23.1. ***Sommer–Zeger Vitamin Supplement Data***

Compliance Type	Assignment Z_i	Vitamin Supplements W_i^{obs}	Survival Y_i^{obs}	Number of Units ($N = 23{,}682$)
co or nc	0	0	0	74
co or nc	0	0	1	11,514
nc	1	0	0	34
nc	1	0	1	2385
co	1	1	0	12
co	1	1	1	9663

23.3 SETUP

First, let us expand the potential outcomes notation to fit the IV setting. Given the divergence between assignment to, and receipt of, treatment, the potential outcomes notation becomes more complex than in previous chapters. We maintain throughout this chapter the SUTVA assumption, that (*i*) there are no versions of the treatments, and (*ii*) there are no causal effects of one unit's treatment assignment on another unit's outcome. We focus on the case where the assignment Z_i takes on two values, $Z_i = 0$ if unit i is assigned to the control group, and $Z_i = 1$ if unit i is assigned to the treatment group. The treatment of primary interest (in the Sommer-Zeger application, the receipt of the vitamin supplements) is denoted by W_i^{obs}. Formally recognizing the role of this variable as an outcome, possibly affected by the assignment to treatment Z_i, we postulate the existence of two potential outcomes, $W_i(0)$ and $W_i(1)$, describing the treatment that would be received under each of the two values of the assignment Z_i. Thus $W_i(0)$ is the treatment unit i would receive if assigned to the control, $Z_i = 0$, and $W_i(1)$ is the treatment unit i would receive if assigned to the active treatment, $Z_i = 1$. Both $W_i(0)$ and $W_i(1)$ take values in $\{0, 1\}$. For unit i, the realized or observed treatment status, W_i^{obs}, equals

$$W_i^{\text{obs}} = W_i(Z_i) = \begin{cases} W_i(0) & \text{if } Z_i = 0, \\ W_i(1) & \text{if } Z_i = 1. \end{cases}$$

In contrast to earlier chapters, we use the superscript "obs" for the treatment here to distinguish the observed value of the primary treatment from the potential primary treatment, which is generally a function of the secondary treatment, Z_i.

For the outcome of interest we take into account that there are, in the noncompliance setting, two "treatments," assignment to treatment Z_i and receipt of treatment W_i. Each takes on two values, so to be general we postulate four potential outcomes, $Y_i(z, w)$, describing the outcome observed if unit i were assigned treatment z and actually received treatment w. For each unit, only two of these four potential outcomes can possibly be observed, $Y_i(0, W_i(0))$ and $Y_i(1, W_i(1))$. The remaining two, $Y_i(0, 1 - W_i(0))$ and $Y_i(1, 1 - W_i(1))$, cannot be observed irrespective of the assignment, and so we refer to the last two as *a priori counterfactuals*. The observed outcome for unit i in our sample, denoted

by Y_i^{obs}, is

$$Y_i^{\text{obs}} = Y_i(Z_i, W_i^{\text{obs}}) = Y_i(Z_i, W_i(Z_i)) = \begin{cases} Y_i(0,0), & \text{if } Z_i = 0, W_i^{\text{obs}} = 0, \\ Y_i(1,0), & \text{if } Z_i = 1, W_i^{\text{obs}} = 0, \\ Y_i(1,1), & \text{if } Z_i = 1, W_i^{\text{obs}} = 1. \end{cases}$$

Note that because the noncompliance is one-sided, there are no units for whom we observe $Y_i(0,1)$. As usual, we think of the population of interest as the N units for which we observe: the instrument Z_i, the treatment received W_i^{obs}, and the outcome Y_i^{obs}.

In this chapter we consider both (*a*) averages over observations by treatment received, and (*b*) averages by treatment assigned. It is therefore useful to have formal notation for these. For notational clarity, the subscripts 0 and 1 denote treatment assignment levels, and the subscripts c and t denote the level of receipt of treatment. Define the subsample sizes by treatment assignment:

$$N_0 = \sum_{i=1}^{N} (1 - Z_i), \quad N_1 = \sum_{i=1}^{N} Z_i,$$

sample sizes by treatment received:

$$N_{\text{c}} = \sum_{i=1}^{N} (1 - W_i^{\text{obs}}), \quad \text{and} \quad N_{\text{t}} = \sum_{i=1}^{N} W_i^{\text{obs}}$$

and sample sizes by both treatment assignment and receipt:

$$N_{0\text{c}} = \sum_{i=1}^{N} (1 - Z_i) \cdot (1 - W_i^{\text{obs}}), \quad N_{0\text{t}} = \sum_{i=1}^{N} (1 - Z_i) \cdot W_i^{\text{obs}},$$

$$N_{1\text{c}} = \sum_{i=1}^{N} Z_i \cdot (1 - W_i^{\text{obs}}), \quad \text{and} \quad N_{1\text{t}} = \sum_{i=1}^{N} Z_i \cdot W_i^{\text{obs}}.$$

Analogously, define the average outcomes and average treatment received by assignment subsample:

$$\overline{Y}_0^{\text{obs}} = \frac{1}{N_0} \sum_{i=1}^{N} (1 - Z_i) \cdot Y_i^{\text{obs}}, \quad \overline{Y}_1^{\text{obs}} = \frac{1}{N_1} \sum_{i=1}^{N} Z_i \cdot Y_i^{\text{obs}},$$

$$\overline{W}_0^{\text{obs}} = \frac{1}{N_0} \sum_{i=1}^{N} (1 - Z_i) \cdot W_i^{\text{obs}}, \quad \overline{W}_1^{\text{obs}} = \frac{1}{N_1} \sum_{i=1}^{N} Z_i \cdot W_i^{\text{obs}},$$

average outcomes by treatment received:

$$\overline{Y}_{\text{c}}^{\text{obs}} = \frac{1}{N_{\text{c}}} \sum_{i=1}^{N} (1 - W_i^{\text{obs}}) \cdot Y_i^{\text{obs}}, \quad \text{and} \quad \overline{Y}_{\text{t}}^{\text{obs}} = \frac{1}{N_{\text{t}}} \sum_{i=1}^{N} W_i^{\text{obs}} \cdot Y_i^{\text{obs}};$$

and, finally, average outcomes by both treatment assignment and treatment receipt:

$$\overline{Y}_{0c}^{obs} = \frac{1}{N_{0c}} \sum_{i=1}^{N} (1 - Z_i) \cdot (1 - W_i^{obs}) \cdot Y_i^{obs}, \quad \overline{Y}_{0t}^{obs} = \frac{1}{N_{0t}} \sum_{i=1}^{N} (1 - Z_i) \cdot W_i^{obs} \cdot Y_i^{obs},$$

$$\overline{Y}_{1c}^{obs} = \frac{1}{N_{1c}} \sum_{i=1}^{N} Z_i \cdot (1 - W_i^{obs}) \cdot Y_i^{obs}, \quad \text{and} \quad \overline{Y}_{1t}^{obs} = \frac{1}{N_{1t}} \sum_{i=1}^{N} Z_i \cdot W_i^{obs} \cdot Y_i^{obs}.$$

Some of the N_{zw} may be zero (in fact, N_{0t} is zero in the current chapter with one-sided compliance), and the corresponding $\overline{Y}_{zw}$ would not be defined in that case.

23.4 INTENTION-TO-TREAT EFFECTS

The first step in our discussion of the IV approach is to study intention-to-treat (ITT) estimands. As we mentioned in Section 23.1, ITT analyses entirely avoid the problem of noncompliance by focusing only on the relationship between the random assignment of Z_i and the outcome, because inference for such effects relies solely on the randomization of the assignment. In contrast to many conventional ITT analyses, we consider two versions of such analyses: analyzing, as "outcomes," both the receipt of treatment (receipt of vitamin A supplements) and the final outcome (survival).

23.4.1 ITT Estimands

Let us first consider the intention-to-treat effect on the receipt of treatment. The unit-level effect of the assignment on the receipt of treatment is

$$\text{ITT}_{\text{W},i} = W_i(1) - W_i(0).$$

The ITT effect on the receipt of treatment is the average of this over all units:

$$\text{ITT}_{\text{W}} = \frac{1}{N} \sum_{i=1}^{N} \text{ITT}_{\text{W},i} = \frac{1}{N} \sum_{i=1}^{N} \big(W_i(1) - W_i(0)\big). \tag{23.1}$$

Because noncompliance is one-sided, $W_i(0) = 0$ for all i, and the expression in Equation (23.1) simplifies to

$$\text{ITT}_{\text{W}} = \frac{1}{N} \sum_{i=1}^{N} W_i(1).$$

Next, let us consider the outcome of primary interest, Y_i. The unit-level intention-to-treat effect is equal to the difference in unit-level outcomes Y_i by assignment status Z_i:

$$\text{ITT}_{\text{Y},i} = Y_i(1, W_i(1)) - Y_i(0, W_i(0)),$$

for $i = 1, \ldots, N$. The average ITT effect on Y is therefore

$$\mathrm{ITT_Y} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{ITT}_{\mathrm{Y},i} = \frac{1}{N}\sum_{i=1}^{N} \big(Y_i(1, W_i(1)) - Y_i(0, W_i(0))\big).$$

The key assumption for identifying the ITT effects in the simple setting of the Sommer-Zeger data set is that the assignment is random. (More generally, we could allow for unconfounded treatment assignment.) Here we formulate that in terms of the extended potential outcome notation, by assuming the distribution of Z_i is free from dependence on all potential outcomes, including the two potential treatments $W_i(z)$ and four potential outcomes $Y_i(z, w)$:

Assumption 23.1 (Random Assignment of Z_i)

$$\Pr(Z_i = 1 \,|\, W_i(0), W_i(1), Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1)) = \Pr(Z_i = 1)\,.$$

From a super-population perspective, the assumption, is in the Dawid conditional-independence notation,

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

23.4.2 Estimating the ITT Effect for the Receipt of Treatment

Given Assumption 23.1, we can estimate $\mathrm{ITT_W}$ following Neyman's approach, outlined in Chapter 6. Complete randomization of the assignment implies that an unbiased estimator for the average causal effect $\mathrm{ITT_W}$ exists in the form of the average difference in treatment status by assignment status:

$$\widehat{\mathrm{ITT_W}} = \overline{W}_1^{\mathrm{obs}} - \overline{W}_0^{\mathrm{obs}} = \overline{W}_1^{\mathrm{obs}},$$

where we use the fact that $W_i(0) = 0$ for all units. Following the derivation presented in Chapter 6, the general form of the (conservative) estimator for the finite-sample sampling variance of $\widehat{\mathrm{ITT_W}}$, under the randomization distribution, is

$$\widehat{\mathbb{V}}(\widehat{\mathrm{ITT_W}}) = \frac{s^2_{W,0}}{N_0} + \frac{s^2_{W,1}}{N_1},$$

where $s^2_{W,0}$ and $s^2_{W,1}$ are the sample variances of $W_i(z)$ within each assignment arm. Because $W_i(0) = 0$, it follows that

$$s^2_{W,0} = \frac{1}{N_0 - 1}\sum_{i:Z_i=0}\left(W_i^{\mathrm{obs}} - \overline{W}_0^{\mathrm{obs}}\right)^2 = 0,$$

and we are concerned only with

$$s^2_{W,1} = \frac{1}{N_1 - 1}\sum_{i:Z_i=1}\left(W_i^{\mathrm{obs}} - \overline{W}_1^{\mathrm{obs}}\right)^2 = \frac{N_1}{N_1 - 1}\cdot \overline{W}_1^{\mathrm{obs}} \cdot (1 - \overline{W}_1^{\mathrm{obs}}).$$

Hence the estimator for the sampling variance of $\widehat{\text{ITT}_\text{W}}$ reduces to

$$\widehat{\mathbb{V}}(\widehat{\text{ITT}_\text{W}}) = \frac{1}{N_1 - 1} \cdot \overline{W}_1^{\text{obs}} \cdot (1 - \overline{W}_1^{\text{obs}}).$$

Recall, from the discussion of randomized experiments in Chapter 6, that this is also a valid estimator for the sampling variance of $\widehat{\text{ITT}_\text{W}}$ when it is viewed as an estimator of the super-population average treatment effect. Using a normal approximation to the sampling distribution, we can construct a randomization-distribution-based, large-sample, 95% confidence interval for ITT_W as

$$\text{CI}^{0.95}(\text{ITT}_\text{W}) = \left(\widehat{\text{ITT}_\text{W}} - 1.96 \cdot \sqrt{\widehat{\mathbb{V}}(\widehat{\text{ITT}_\text{W}})}, \widehat{\text{ITT}_\text{W}} + 1.96 \cdot \sqrt{\widehat{\mathbb{V}}(\widehat{\text{ITT}_\text{W}})}\right).$$

Let us illustrate this using the Sommer-Zeger vitamin A data. For these data we find

$$\overline{W}_1^{\text{obs}} = 0.8000, \quad \text{and} \quad s^2_{W,1} = 0.4000^2.$$

Given that $N_1 = 12{,}094$ individuals were assigned to receive the vitamin supplements, it follows that

$$\widehat{\text{ITT}_\text{W}} = 0.8000, \quad \text{and} \quad \widehat{\mathbb{V}}\left(\widehat{\text{ITT}_\text{W}}\right) = 0.0036^2,$$

leading to a 95% large-sample confidence interval for ITT_W equal to

$$\text{CI}^{0.95}(\text{ITT}_\text{W}) = \big(0.7929, 0.8071\big).$$

Thus, we obtain for the Sommer-Zeger data a precise estimate of the ITT effect of assignment to treatment on the receipt of treatment (with the caveat that we ignore the clustered randomization).

23.4.3 Estimating the ITT Effect for the Outcome of Interest

Next let us consider the outcome of primary interest, Y_i. Because the assignment Z_i is unconfounded, we can unbiasedly estimate the conventional intention-to-treat estimand, ITT_Y. Using the analysis for randomized experiments from Chapter 6, an unbiased estimator for this effect can be obtained by differencing the average outcomes for those assigned to the treatment and those assigned to the control:

$$\widehat{\text{ITT}_\text{Y}} = \overline{Y}_1^{\text{obs}} - \overline{Y}_0^{\text{obs}},$$

where $\overline{Y}_1^{\text{obs}}$ and $\overline{Y}_0^{\text{obs}}$ are as defined in Section 23.3. The sampling variance for this estimator can also be estimated using the methods from Chapter 6:

$$\widehat{\mathbb{V}}(\widehat{\text{ITT}_\text{Y}}) = \frac{s^2_{Y,1}}{N_1} + \frac{s^2_{Y,0}}{N_0},$$

where

$$s^2_{Y,0} = \frac{1}{N_0 - 1} \sum_{i:Z_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_0^{\text{obs}} \right)^2, \quad \text{and} \quad s^2_{Y,1} = \frac{1}{N_1 - 1} \sum_{i:Z_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_1^{\text{obs}} \right)^2.$$

Let us return again to the vitamin A supplement data. Using the survival indicator Y_i^{obs} as the outcome, we find:

$$\overline{Y}_0^{\text{obs}} = 0.9956, \quad \overline{Y}_1^{\text{obs}} = 0.9962, \quad s^2_{Y,0} = 0.0797^2, \quad \text{and} \quad s^2_{Y,1} = 0.0616^2.$$

Given that $N_1 = 12,094$ individuals were assigned to receive the supplements, and $N_0 = 11,588$ were assigned to receive no supplements, it follows that

$$\widehat{\text{ITT}_\text{Y}} = 0.0026, \quad \text{and} \quad \widehat{\mathbb{V}}\left(\widehat{\text{ITT}_\text{Y}}\right) = 0.0009^2,$$

leading to a large-sample 95% confidence interval for ITT_Y:

$$\text{CI}^{0.95}(\text{ITT}_\text{Y}) = \big(0.0008, 0.0044\big).$$

We conclude that the estimated ITT effect of assignment to supplements on survival is positive and statistically different from zero at conventional significance levels. If all we were interested in is these ITT effects, we could stop here. In many cases, however, there is also interest in the causal effect of *taking* the supplements as opposed to the causal effect of *being assigned to take* them. Part of the motivation is that one may believe that the causal effect of actually taking the treatment has more external validity, that is, is more likely to generalize to other settings and populations, than the causal effect of being assigned to take them. The argument for this is that the ITT effect combines partly the biological effect of taking the supplements, and the psychological effect of assignment to take the supplements on actually taking them. When this is true, the causal effect of taking the supplements may be more relevant than the causal effect of assigning individuals to take the supplements for policy makers who are considering making them available in other parts of the country or on a wider scale, with more or less encouragement to them than in the current experiment. This point is particularly compelling when the reasons for the noncompliance are idiosyncratic to the setting in which the experiment was conducted, so that in different settings, compliance may be substantially different.

23.5 COMPLIANCE STATUS

A crucial role in the analyses discussed in this chapter is played by the compliance behavior of the units. Here we continue our analysis of the IV approach with a detailed discussion of this behavior, captured by the pair of potential outcomes $(W_i(0), W_i(1))$. A key feature of our approach is that we view the compliance behavior in this study when assigned not to take $(W_i(0))$ and when assigned to take $(W_i(1))$ as reflecting partially observed characteristics of each unit.

Table 23.2. ***Possibly Compliance Status by Observed Assignment and Receipt of Treatment for the Sommer-Zeger Vitamin Supplement Data***

		Assignment Z_i	
		0	1
Receipt of treatment W_i^{obs}	0	nc or co	nc
	1	–	co

Note: One-sided noncompliance rules out the $Z_i = 0$ $W_i^{\text{obs}} = 1$ cell.

23.5.1 Compliers and Noncompliers

Let us return to the two potential outcomes for the treatment received, $W_i(0)$ and $W_i(1)$. By the assumption that noncompliance is one-sided, it follows that all units assigned to the control in fact receive the control, thus $W_i(0) = 0$ for all i. In contrast, $W_i(1)$, the treatment unit i would receive if assigned to the active treatment, can equal either 0 or 1. Units with $W_i(1) = 1$ will be observed to comply with their assignment, irrespective of what that assignment is, whereas those with $W_i(1) = 0$ will be observed not to comply if assigned to $Z_i = 1$. We therefore label the former group *compliers* and the latter group *noncompliers*. In a randomized experiment with full compliance, $W_i(z)$ would be equal to z for all units, and as a result, all units would be compliers. Note that this definition of compliance status is based solely on a unit's behavior given assignment to the active treatment in this experiment. Because all units assigned to the control can be prevented from receiving the active treatment, all units will be observed to comply when assigned $Z_i = 0$. Thus we can only distinguish, by observation, compliers from noncompliers in the subgroup assigned to the treatment. For the purposes of our discussion, compliance status will be denoted by a group indicator $G_i \in \{\text{co}, \text{nc}\}$, with $G_i = \text{co}$ for compliers and $G_i = \text{nc}$ for noncompliers:

$$G_i = \begin{cases} \text{co} & \text{if } W_i(1) = 1, \\ \text{nc} & \text{if } W_i(1) = 0. \end{cases}$$

Table 23.2 illustrates the compliance status and its relation to the observed assignment Z_i and the observed receipt of the treatment W_i^{obs}. The "–" entry, corresponding to $Z_i = 0$ and $W_i^{\text{obs}} = 1$, indicates that by the fact that noncompliance is one-sided, there are no units with $Z_i = 0$ and $W_i^{\text{obs}} = 1$.

When we consider two-sided noncompliance in the next chapter, we generalize these ideas to allow for the possibility that some of those assigned to the control group in fact can receive the active treatment, and thus allow $W_i(0)$ to differ from zero.

Let N_{co} and N_{nc} denote the number of units of each type in the sample:

$$N_{\text{co}} = \sum_{i=1}^{N} \mathbf{1}_{G_i = \text{co}}, \quad \text{and} \quad N_{\text{nc}} = \sum_{i=1}^{N} \mathbf{1}_{G_i = \text{nc}} = N - N_{\text{co}},$$

and let π_{co} and π_{nc} denote the sample fractions of compliers and noncompliers:

$$\pi_{\text{co}} = \frac{N_{\text{co}}}{N}, \quad \text{and} \quad \pi_{\text{nt}} = \frac{N_{\text{nc}}}{N} = 1 - \pi_{\text{co}}.$$

In the potential outcomes notation, it becomes clear that the compliance status in this experiment is a latent characteristic of an individual unit. It is a characteristic in the sense that compliance status is not affected by outside manipulation (specifically, it is not affected by the assignment to treatment Z_i); it is latent because we cannot observe its value for all units: that is, for those units assigned to the control group, we do not observe their compliance status. In contrast, for units assigned to receive the active treatment, we do observe whether they are compliers or noncompliers (although this will change when we allow for two-sided noncompliance in the next chapter). Hence the three key features of this latent compliance status are: (*i*) it is a function of the two (secondary) potential outcomes, which describe the receipt of treatment for different values of the assignment Z_i; (*ii*) the value of the characteristic is not affected by the assignment to treatment, although which value is observed is affected by the assignment; and (*iii*) it cannot always be entirely inferred from the observed values for assignment and treatment, Z_i and W_i^{obs}. This last feature is illustrated in Table 23.2 by the fact that the $(Z_i = 0, W_i^{\text{obs}} = 0)$ cell contains a mixture of compliers and noncompliers.

23.5.2 The ITT Effect on the Treatment Received by Compliance Status

First let us consider the population ITT effect on the secondary outcome, treatment received, separately by compliance status. For noncompliers, $W_i(z) = 0$ for $z = 0, 1$. Hence

$$\text{ITT}_{\text{W,nc}} = \frac{1}{N_{\text{nc}}} \sum_{i:G_i=\text{nc}} \big(W_i(1) - W_i(0)\big) = \frac{1}{N_{\text{nc}}} \sum_{i:G_i=\text{nc}} W_i(1) = 0.$$

For compliers, $W_i(z) = z$ for $z = 0, 1$. Hence

$$\text{ITT}_{\text{W,co}} = \frac{1}{N_{\text{co}}} \sum_{i:G_i=\text{co}} \big(W_i(1) - W_i(0)\big) = \frac{1}{N_{\text{co}}} \sum_{i:G_i=\text{co}} W_i(1) = 1.$$

The overall ITT effect on treatment received is a weighted average of the within-compliance subpopulation ITT effects:

$$\text{ITT}_{\text{W}} = \pi_{\text{nc}} \cdot \text{ITT}_{\text{W,nc}} + \pi_{\text{co}} \cdot \text{ITT}_{\text{W,co}} = \pi_{\text{co}},$$

and $\pi_{\text{nc}} = 1 - \text{ITT}_{\text{W}}$. In words, the ITT effect on treatment received is equal to the population fraction of compliers. Note that this does not rely on any assumptions. It simply follows from the definition of compliance behavior and the existence of the potential outcomes.

23.5.3 The ITT Effect on the Primary Outcome by Compliance Status

The next step is to decompose the intention-to-treat effect for the primary outcome, ITT_{Y}, into a weighted average of the intention-to-treat effects by compliance status. Define

$$\text{ITT}_{\text{Y,co}} = \frac{1}{N_{\text{co}}} \sum_{i:G_i=\text{co}} \big(Y_i(1, W_i(1)) - Y_i(0, W_i(0))\big),$$

and

$$\mathrm{ITT}_{\mathrm{Y,nc}} = \frac{1}{N_{\mathrm{nc}}} \sum_{i:G_i=\mathrm{nc}} \big(Y_i(1, W_i(1)) - Y_i(0, W_i(0))\big),$$

so that we can write

$$\begin{aligned} \mathrm{ITT}_{\mathrm{Y}} &= \mathrm{ITT}_{\mathrm{Y,co}} \cdot \pi_{\mathrm{co}} + \mathrm{ITT}_{\mathrm{Y,nc}} \cdot \pi_{\mathrm{nc}} \\ &= \mathrm{ITT}_{\mathrm{Y,co}} \cdot \mathrm{ITT}_{\mathrm{W}} + \mathrm{ITT}_{\mathrm{Y,nc}} \cdot (1 - \mathrm{ITT}_{\mathrm{W}}). \end{aligned} \tag{23.2}$$

Let us consider directly the $\mathrm{ITT}_{\mathrm{Y}}$ effects by compliance type. The average ITT effect for noncompliers is

$$\mathrm{ITT}_{\mathrm{Y,nc}} = \frac{1}{N_{\mathrm{nc}}} \sum_{i:G_i=n} \big(Y_i(1,0) - Y_i(0,0)\big).$$

Note, however, that this ITT effect for noncompliers is not informative about the effect of the primary treatment: it compares two potential outcomes for a group of units, all of which always receive the control treatment.

For compliers the ITT effect is generally more interesting for the causal effects of the receipt of treatment. The average $\mathrm{ITT}_{\mathrm{Y}}$ effect for compliers is

$$\mathrm{ITT}_{\mathrm{Y,co}} = \frac{1}{N_{\mathrm{co}}} \sum_{i:G_i=c} \big(Y_i(1,1) - Y_i(0,0)\big).$$

This ITT effect is at least potentially informative about the effect of the primary treatment, because it is based on a comparison of potential outcomes when receiving the active treatment and when not receiving the active treatment for the subpopulation of compliers.

The two ITT effects on Y by complier status, $\mathrm{ITT}_{\mathrm{Y,co}}$ and $\mathrm{ITT}_{\mathrm{Y,nc}}$, cannot be estimated directly from the observable data, because we cannot infer the latent compliance status for units assigned to the control group. Nevertheless, because receipt of treatment, W_i^{obs}, is unconfounded conditional on compliance status given randomization of the assignment, we can still distentangle the ITT effects by compliance type under an additional assumption: the exclusion restriction.

It is important here that the receipt of treatment is unconfounded within subpopulations defined by compliance status. This follows from Assumption 23.1, that Z_i is randomly assigned, in combination with the fact that W_i^{obs} is a deterministic function of Z_i given compliance status.

Lemma 23.1 (Super-Population Unconfoundedness of Receipt of Treatment Given Compliance Status)
Suppose Assumption 23.1 holds. Then, for $g \in \{\mathrm{co}, \mathrm{nc}\}$,

$$\Pr\Big(W_i^{\mathrm{obs}} = 1 \,\Big|\, Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), G_i = g\Big) = \Pr\Big(W_i^{\mathrm{obs}} = 1 \,\Big|\, G_i = g\Big),$$

or

$$W_i^{\text{obs}} \perp\!\!\!\perp Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1) \,\big|\, G_i.$$

To see this, consider the two compliance types separately. First, for noncompliers ($G_i = \text{nc}$), we always have $W_i^{\text{obs}} = 0$, so unconfoundedness holds trivially. For compliers ($G_i = \text{co}$), $W_i^{\text{obs}} = Z_i$, and thus Lemma 23.1 holds by Assumption 23.1, random assignment of Z_i. The problem is that we cannot directly exploit the latent unconfoundedness result in Lemma 23.1 (latent, because it only holds given a partially unobserved covariate), because compliance type is only partially observed. We therefore rely on indirect methods for exploiting this latent unconfounedness property.

23.6 INSTRUMENTAL VARIABLES

In this section we discuss the key assumption underlying the method of instrumental variables, and present the main result of this chapter that, under that key assumption, we can estimate the average ITT effect for compliers, $\text{ITT}_{\text{Y,co}}$. We discuss the interpretation of this ITT effect and how it may be related to the causal effect of the receipt of treatment. We then discuss two approaches to inference for this average effect.

23.6.1 Exclusion Restriction for Noncompliers

First we discuss the key assumption that underlies, in some form or another, all instrumental variables analyses.

Assumption 23.2 (Exclusion Restriction for Noncompliers) *For all noncompliers, that is, all units with* $G_i = \text{nc}$*,*

$$Y_i(0,0) = Y_i(1,0).$$

This assumption, the exclusion restriction, rules out, for noncompliers, an effect of the assignment, the instrument Z_i, on the outcome of interest Y_i. It states that changing the assignment has no causal effect on the outcome, for those units for whom the level of the primary treatment W_i does not change with the change in assignment.

This exclusion restriction is the key assumption underlying the instrumental variables approach. Unlike the latent unconfoundedness assumption, however, it is not implied by the randomization of the assigned treatment. Instead, it is a substantive assumption that need not be appropriate in all randomized experiments with noncompliance, although it can be made plausible by design features such as double-blinding.

A slightly weaker version of the exclusion restriction for noncompliers requires the exclusion restriction to hold in distribution for the super-population:

Assumption 23.3 (Stochastic Exclusion Restriction for Noncompliers)

$$Z_i \perp\!\!\!\perp Y_i(Z_i, W_i(Z_i)),$$

for all noncompliers, that is, all units with $G_i = \text{nc}$.

This assumption implies that the super-population distribution of $Y_i(0,1)$ is the same as that of $Y_i(1,0)$ for noncompliers with $W_i(0) = W_i(1) = 0$. One advantage of this assumption is that there is a natural way to relax it in the presence of pre-treatment variables by requiring the independence to hold only conditional on the pre-treatment variables.

23.6.2 Exclusion Restriction for Compliers

Because of the central role of the exclusion restriction, some general comments about the applicability of this assumption are in order. Before doing so, let us also formulate a second exclusion restriction, this time for compliers.

Assumption 23.4 (Exclusion Restrictions for Compliers) *For all units with $G_i = \text{co}$, that is, all compliers,*

$$Y_i(0,w) = Y_i(1,w)$$

for both levels of the treatment w.

This is an assumption of a very different nature from the exclusion restriction for noncompliers. It restricts, for compliers, $Y_i(0,0)$ to be equal to $Y_i(1,0)$, and restricts $Y_i(0,1)$ to be equal to $Y_i(1,1)$. But for compliers, we observe either $Y_i(0,0)$ or $Y_i(1,1)$, and never observe $Y_i(0,1)$ or $Y_i(1,0)$, and so these restrictions have no empirical consquences, either in the current form or in a stochastic version, unlike the exclusion restriction for noncompliers. In a sense, this restriction is essentially an *attribution* of the ITT effect for compliers to the causal effect of the receipt of treatment, rather than to its assignment. It is primarily about the interpretation of this ITT effect, not about issues concerning estimating it from the data.

Note that the exclusion restriction for compliers is routinely made, often implicitly, in randomized experiments with full compliance (in that case all units are compliers). For instance, when analyzing and interpreting the results from double-blind randomized drug trials with full compliance, one often implicitly assumes that the estimated effect is due to the *receipt* of the drug, not to the *assignment* to receive the drug. Thus, the assumption is implicitly made that similar unit-level treatment effects will occur if the assignment mechanism is changed from randomized assignment to either voluntary assignment or full adoption. Specifically, suppose a drug company estimates the efficacy of a new drug in a randomized trial. Implicitly the assumption is that, had, at the start of the trial, all individuals been told that they would receive the new active drug and that no one would receive the control treatment, the typical outcome would have been approximately the same as the typical outcome observed in the subsample actually assigned to the treatment. Moreoever, after the drug is approved, physicians will presumably prescribe the new drug without using randomization. Again the presumption is that their patients will respond to the prescribed treatment in the same way that similar subjects in the randomized trial responded to assignment to the possibly unknown, blinded, treatment.

Yet the fact that this assumption is often implicitly made does not mean that this exclusion restriction is innocuous. There are many examples of studies where assignment did

make an important difference, separate from receipt of the active treatment. Concerns about potential complications from such direct effects of assignment motivate the use of placebos, and blinding or double blinding, in clinical trials with human subjects. If individuals do not know their values of assignment, it is difficult to see how the assignments could affect their outcomes, except through the biological effect of the treatment received. But, again, receipt of a known, approved drug is not necessarily the same as receipt of a blinded drug being evaluated in the experiment.

23.6.3 Discussion of the Exclusion Restrictions

In some settings where noncompliance is an issue, however, placebos and (double-) blinding are often infeasible. If the treatment is an invasive procedure or requires active participation on the part of the individual, the researcher typically cannot hide the nature of the treatment. Even in randomized eligibility designs, where access (eligibility) to the treatment is randomized, the exclusion restriction may be violated. Individuals assigned to the active treatment may refuse to accept it but, in response to the notification of eligibility, may take actions they would not have taken otherwise. For example, consider the evaluation of a smoking cessation program. Suppose the program is offered to a random sample of smokers. Some may be unwilling to go through the program if it takes a large amount of time or effort. Yet in response to the assignment such individuals may still change their lifestyles, including their smoking habits in ways that affect their subsequent health outcomes. In that case, health outcomes would differ by assignment for such individuals, even though they are noncompliers who do not participate in the program irrespective of their assignment. Examples such as these illustrate that the exclusion restriction requires careful consideration of the various paths through which assignment may affect outcomes.

One should note, however, that the exclusion restrictions, Assumptions 23.2 and 23.4, do *not* in any way restrict compliance behavior itself. For example, it allows for the possibility that individuals know their outcomes under both treatments and deliberately choose to comply when assigned to the active treatment only if it will benefit them. Specifically, suppose that all those with $Y_i(1,1) > Y_i(1,0)$ (those whose health status would improve with the receipt of the treatment) choose to comply, and all those with $Y_i(1,1) \leq Y_i(1,0)$ choose not to. Such behavior would imply that the receipt of treatment W_i^{obs} is confounded, and it is often exactly this type of systematic noncompliance behavior that motivates researchers to consider instrumental variable analyses. Such behavior is not, however, inconsistent with the exclusion restriction and thus will be compatible with the analyses developed here.

Let us consider the exclusion restriction for noncompliers for the Sommer-Zeger vitamin A supplement data. This restriction requires that, for those individuals who would not receive the supplements even if assigned to take them, the potential outcomes are unaffected by assignment. This assumption seems fairly plausible. If some mothers living in villages assigned to the treatment did not receive the supplements because of administrative mishaps, or through lack of interest, it is quite likely that the infants of such mothers would not have had different outcomes had their village been assigned to the control group, except if there are fewer contiguous infant diseases in the villages that were assigned the vitamin supplements. Nevertheless, this is a key assumption for

the validity of the IV approach, and even in this example it is not necessarily satisfied. Violations of this assumption could arise if the reason these women did not receive the supplements was related to other health improvement measures taken in some villages but not in others. For example, suppose that noncompliance was high in some villages because the administrators in those villages, if assigned to receive the supplements, diverted the program funding toward other health care improvements that would have been otherwise unaffordable. In that case, outcomes for noncomplying mothers would differ by assignment, even though none took the supplements, violating the exclusion restriction. Such a story may seem fairly implausible in this case, but such stories are important to consider. We will return to discuss such violations in other examples in subsequent chapters.

23.6.4 Local Average Treatment Effects

In this section we discuss the most important result in this chapter. Consider the average ITT effect in the population, decomposed by compliance status:

$$\text{ITT}_\text{Y} = \text{ITT}_{\text{Y,co}} \cdot \text{ITT}_\text{W} + \text{ITT}_{\text{Y,nc}} \cdot (1 - \text{ITT}_\text{W}), \tag{23.3}$$

using the fact that the one-sided nature of the noncompliance implies that $\text{ITT}_\text{W} = \pi_\text{co}$. The exclusion restriction for noncompliers implies that for noncompliers $Y_i(0,0) = Y_i(1,0)$, and thus,

$$\text{ITT}_{\text{Y,nc}} = 0.$$

Hence, the second term on the right-hand side of (23.3) is zero, reducing the global ITT on the outcome to the product of two ITT effects, the "local" ITT effect on the outcome for the compliers, and the global ITT effect on the receipt of treatment:

$$\text{ITT}_\text{Y} = \text{ITT}_{\text{Y,co}} \cdot \text{ITT}_\text{W}. \tag{23.4}$$

We now rearrange Equation (23.4) to give our formal result:

Theorem 23.1 (Local Average Treatment Effect)
Suppose that Assumption 23.2 holds. Then

$$\tau_\text{late} = \text{ITT}_{\text{Y,co}} = \frac{\text{ITT}_\text{Y}}{\text{ITT}_\text{W}}.$$

In other words, under the exclusion restriction for noncompliers, the ratio of the ITT effect on the outcome to the ITT effect on the treatment is equal to the ITT effect on the outcome for compliers, or what is called the Local Average Treatment Effect (LATE), or, synonymously, the Complier Average Causal Effect (CACE).

If we are also willing to assume the second exclusion restriction, the exlusion restriction for compliers given in Assumption 23.4, we can interpret this local average treatment effect as the average causal effect of the receipt of treatment for compliers. Thus, given both exclusion restrictions and the randomization assumption, we can learn about the effect of the primary treatment for the subpopulation of compliers, because we can unbiasedly estimate both the numerator and the denominator of τ_late.

To give a different interpretation for the result in Theorem 23.1, suppose for a moment that we could observe compliance status for all units. By Lemma 23.1, receipt of treatment is unconfounded given compliance status G_i, and so we could then analyze the data separately for noncompliers and compliers. Within these subpopulations, we can compare outcome by treatment status. For noncompliers, there would be no information in the data regarding the effect of the primary treatment on the outcome, because no noncomplier ever receives the active treatment. The data from noncompliers would therefore be discarded because of the absence of units who received the active treatment. For compliers, receipt of treatment is identical to assignment, and for this subpopulation we can therefore consistently estimate effects of the receipt of the treatment on the outcome, because, by the second exclusion restriction, it equals the intention-to-treat effect of the assignment on the final outcome. The only, but crucial, missing piece in this argument is that we do not observe the compliance status for all units. However, given the exclusion restriction, we can disentangle the potential outcome distributions for compliers and noncompliers from the mixture of noncompliers and compliers in the subpopulation assigned to the control treatment, through, for example, an imputation-based approach such as that outlined in Chapter 25 or the moment-based approach introduced here.

23.7 MOMENT-BASED INSTRUMENTAL VARIABLES ESTIMATORS

Summarizing the discussion so far, the overall ITT effect consists of two parts, the ITT effect for compliers and the ITT effect for noncompliers, weighted by their population proportions. The exclusion restriction for noncompliers implies that the ITT effect for noncompliers is zero. Hence, under the exclusion restriction for noncompliers, the ratio of the overall ITT effect, to the population proportion of compliers, is equal to the ITT effect for compliers.

In Section 23.4 we discussed how to estimate and conduct inference for ITT_W and ITT_Y. Given those two unbiased estimators, a simple moment-based instrumental variables (iv) estimator for τ_{late} is the ratio of estimated ITT effects,

$$\hat{\tau}^{\text{iv}} = \frac{\widehat{\text{ITT}_\text{Y}}}{\widehat{\text{ITT}_\text{W}}}.$$

This simple estimator has some drawbacks, and in Chapter 25 we discuss model-based methods that have more attractive statistical properties, especially in small samples. One of the reasons is that it does not necessarily satisfy all the restrictions implied by Assumptions 23.1 and 23.2. We will discuss these restrictions in more detail in Chapter 25, but as a simple example, suppose that $\text{ITT}_\text{W} = 0$. In that case there are no compliers, and by the exclusion restriction for noncompliers, it must be the case that $\text{ITT}_\text{Y} = 0$. More generally, the restrictions imply that the joint distribution of the data is consistent with the subpopulation of $(Z_i = 0, W_i^{\text{obs}} = 0)$ being a mixture of compliers and noncompliers, and the outcome distribution for noncompliers being the same as that for units with $(Z_i = 1, W_i^{\text{obs}} = 0)$.

The sampling variance calculations for the two ITT effects separately followed from the Neyman approach discussed in Chapter 6. Here we discuss the extension to the

sampling variance for $\hat{\tau}^{\text{iv}}$. Here we take explicitly a super-population perspective. That is, we view our sample as a random sample from a large population. In that large population, there is an average ITT effect for compliers, $\text{ITT}_{\text{Y,co}} = \mathbb{E}[Y_i(1, W_i(1)) - Y_i(0, W_i(0))|G_i = \text{co}$. We consider the sampling variance of $\hat{\tau}^{\text{iv}} - \text{ITT}_{\text{Y,co}}$. To calculate the sampling variance of the IV estimator $\hat{\tau}^{\text{iv}}$ requires estimation of the sampling covariance between $\widehat{\text{ITT}_{\text{W}}}$ and $\widehat{\text{ITT}_{\text{Y}}}$. With that covariance, we can use the delta method to estimate the large-sample sampling variance of the ratio of ITT effects. (See the Appendix to this chapter for more details on the delta method in general.) The result is that in large samples, $\hat{\tau}^{\text{iv}}$, as an estimator of the super-population ITT effect for compliers, will be approximately normally distributed with sampling variance

$$\begin{aligned}\mathbb{V}_{\text{sp}}(\hat{\tau}^{\text{iv}}) = &\frac{1}{\text{ITT}_{\text{W}}^2} \cdot \mathbb{V}(\widehat{\text{ITT}_{\text{Y}}}) + \frac{\text{ITT}_{\text{Y}}^2}{\text{ITT}_{\text{W}}^4} \cdot \mathbb{V}(\widehat{\text{ITT}_{\text{W}}}) \\ &- 2 \cdot \frac{\text{ITT}_{\text{Y}}}{\text{ITT}_{\text{W}}^3} \cdot \mathbb{C}(\widehat{\text{ITT}_{\text{Y}}}, \widehat{\text{ITT}_{\text{W}}}),\end{aligned} \tag{23.5}$$

where $\mathbb{C}(\cdot, \cdot)$ denotes the covariance of two random variables. A simple estimator for the sampling variance can be based on substituting estimates for the components of this sampling variance. Using this to construct confidence intervals raises some issues, such as if the denominator of (23.5), ITT_{W}, is close to zero, normality is likely to be a poor approximation to the sampling distribution of the estimator.

Returning to the vitamin A supplement data, using our earlier estimates for ITT_{Y}, $\mathbb{V}(\widehat{\text{ITT}_{\text{Y}}})$, ITT_{W}, and $\mathbb{V}(\widehat{\text{ITT}_{\text{W}}})$, in combination with the estimate for the covariance of $\widehat{\text{ITT}_{\text{Y}}}$ and $\widehat{\text{ITT}_{\text{W}}}$, $\widehat{\mathbb{C}}(\widehat{\text{ITT}_{\text{Y}}}, \widehat{\text{ITT}_{\text{W}}}) = -0.00000017$ (corresponding to a correlation between $\widehat{\text{ITT}_{\text{Y}}}$ and $\widehat{\text{ITT}_{\text{W}}}$ equal to -0.0502), we find that the method-of-moments IV estimate for the effect of taking vitamin A supplements on survival is

$$\hat{\tau}^{\text{iv}} = \frac{\widehat{\text{ITT}_{\text{Y}}}}{\widehat{\text{ITT}_{\text{W}}}} = 0.0032, \quad \text{and} \quad \mathbb{V}(\hat{\tau}^{\text{iv}}) = 0.0012^2,$$

leading to a 95% large-sample confidence interval for $\text{ITT}_{\text{Y,co}}$ (or τ_{late}) equal to

$$\text{CI}^{0.95}(\text{ITT}_{\text{Y,co}}) = \big(0.0010, 0.0055\big).$$

Because the ITT effect on the receipt of treatment is precisely estimated, and far from zero, the 95% confidence interval is likely to be valid (in the statistically conservative sense), with the qualification that we ignored the clustering of the experiment by village.

If in addition to the exclusion restriction for noncompliers, we are willing to assume the exclusion restriction for compliers, this estimated ITT effect for compliers can be interpreted as equal to the estimated average effect of the primary treatment on the primary outcome for compliers.

23.8 LINEAR MODELS AND INSTRUMENTAL VARIABLES

Even for readers familiar with traditional discussions of instrumental variables in econometric textbooks, the discussion thus far may look unfamiliar. In this section we discuss the link between the approach advocated in this book and conventional econometric

instrumental variables analyses. Readers not familiar with the textbook econometrics approach may wish to skip this section.

The traditional use of instrumental variables in the economics literature relies heavily on linear parametric specifications, even though some of these are not critical. It also takes a super-population perspective, where the sample at hand is assumed to be a random sample from an infinitely large population, and the estimands are population average causal effects. We maintain here both exclusion restrictions, for noncompliers and compliers. As a result we can drop the dependence of the potential outcome $Y_i(z, w)$ on z and write, without ambiguity, $Y_i(w)$, as a function of the receipt of treatment alone. In order to see the connection with our framework, it is useful to assume initially a constant treatment effect: $Y_i(1) - Y_i(0) = \tau$ for all i. We relax this assumption later. Define $\alpha = \mathbb{E}_{\text{sp}}[Y_i(0)]$ to be the super-population average outcome given the control treatment, so that we can write

$$\mathbb{E}_{\text{sp}}\,[Y_i(w)] = \alpha + \tau \cdot w,$$

for $w = \{0, 1\}$. We define the residual $\varepsilon_i = Y_i(0) - \alpha$ to be the unit-level deviation of the control outcome from its population mean, so that we can further write

$$Y_i(w) = \alpha + \tau \cdot w + \varepsilon_i. \tag{23.6}$$

Equation (23.6) is what is known in the econometric literature as a *structural* or *behavioral* equation: it relates treatments to outcomes in a causal way. For a given unit i (and thus, for a fixed value ε_i), $Y_i(w)$ is the outcome we would observe if we fixed (*set* in Pearl's (2000) terminology) $W_i = w$.

Equation (23.6) is *not*, however, a conventional regression function. Note that it is not written in terms of observed quantities. Substituting observed values for the treatment and outcome we can instead write

$$Y_i^{\text{obs}} = Y_i(W_i^{\text{obs}}) = Y_i(0) + W_i^{\text{obs}} \cdot (Y_i(1) - Y_i(0)) = \alpha + \tau \cdot W_i^{\text{obs}} + \varepsilon_i. \tag{23.7}$$

Yet, as written, Equation (23.7) remains a behavioral equation, not a conditional expectation: in general it is *not* true that $\mathbb{E}[Y_i^{\text{obs}}|W_i^{\text{obs}} = w] = \alpha + \tau \cdot W_i^{\text{obs}}$. The coefficient τ for the treatment indicator W_i^{obs} represents the *causal* effect of the treatment on the outcome; it is not equal to the ratio of the super-population covariance of Y_i^{obs} and W_i^{obs}, to the variance of W_i^{obs}.

The key factor distinguishing Equation (23.7) from a standard regression function is that the regressor, the receipt of treatment W_i^{obs}, is possibly correlated with $Y_i(0)$, and thus with the residual ε_i. To see this, let us first calculate the conditional mean of ε_i given W_i^{obs} in the super-population. Here let π_g be the share in the super-population of compliance type $G_i = g$. Remember that ε_i is defined as the difference between the observed and expected control outcome: $\varepsilon_i = Y_i(0) - \alpha = Y_i(0) - \mathbb{E}_{\text{sp}}[Y_i(0)]$. Given $W_i^{\text{obs}} = 1$ we have:

$$\begin{aligned}\mathbb{E}_{\text{sp}}[\varepsilon_i|W_i^{\text{obs}} = 1] &= \mathbb{E}_{\text{sp}}[\varepsilon_i|G_i = \text{co}] \\ &= \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] - \big(\mathbb{E}_{\text{sp}}[Y_i(0)|X_i = \text{co}] \cdot \pi_{\text{co}} + \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}] \cdot \pi_{\text{nc}}\big) \\
&= \pi_{\text{nc}} \cdot \big(\mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}]\big) \\
&= \pi_{\text{nc}} \cdot \Delta_{\text{co,nc}},
\end{aligned}
$$

where $\Delta_{\text{co,nc}}$ is defined as the difference in average control outcome for compliers and noncompliers, $\Delta_{\text{co,nc}} = \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}]$. To calculate $\mathbb{E}_{\text{sp}}[\varepsilon_i|W_i^{\text{obs}} = 0]$, first decompose $\mathbb{E}_{\text{sp}}[\varepsilon_i] = 0$:

$$
0 = \mathbb{E}_{\text{sp}}[\varepsilon_i] = \mathbb{E}_{\text{sp}}[\varepsilon_i|W_i^{\text{obs}} = 1] \cdot \Pr(W_i^{\text{obs}} = 1) + \mathbb{E}_{\text{sp}}[\varepsilon_i|W_i^{\text{obs}} = 0] \cdot \Pr(W_i^{\text{obs}} = 0).
$$

Given that the probability $\Pr(W_i^{\text{obs}} = 1)$ is equal to $p_Z \cdot \pi_{\text{co}}$, and thus $\Pr(W_i^{\text{obs}} = 0) = (1 - p_Z \cdot \pi_{\text{co}})$, it follows that

$$
\mathbb{E}_{\text{sp}}[\varepsilon_i|W_i^{\text{obs}} = 0] = -\frac{\pi_{\text{nc}} \cdot p_Z \cdot \pi_{\text{co}}}{1 - p_Z \cdot \pi_{\text{co}}} \cdot \Delta_{\text{co,nc}}.
$$

These expectations $\mathbb{E}_{\text{sp}}[\varepsilon_i|W_i^{\text{obs}} = w]$ are typically not zero. In econometric terminology, the explanatory variable W_i^{obs} is *endogenous*, and least squares methods do not lead to consistent estimation of τ.

Although the receipt of treatment, W_i^{obs}, is *not* independent of ε_i, the assignment to treatment, or the instrument Z_i *is* independent of ε_i. This follows from the random assignment assumption and the definition of ε_i in terms of the potential outcomes. This independence of Z_i and ε_i can be exploited through what is known in econometrics as Two-Stage-Least-Squares (TSLS) estimation. First, this independence implies that the conditional expectation of ε_i given Z_i is zero. This in turn implies that the conditional expectation of Y_i^{obs} given Z_i equals,

$$
\mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|Z_i] = \alpha + \tau \cdot \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i] + \mathbb{E}_{\text{sp}}[\varepsilon_i|Z_i] = \alpha + \tau \cdot \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i].
$$

This conditional expectation of Y_i^{obs} given Z_i is linear in $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]$, with coefficient equal to the treatment effect of interest τ. We can therefore write

$$
\begin{aligned}
Y_i^{\text{obs}} &= \alpha + \tau \cdot \Big(\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i] + \Big(W_i^{\text{obs}} - \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]\Big)\Big) + \varepsilon_i \\
&= \alpha + \tau \cdot \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i] + \eta_i, \qquad (23.8)
\end{aligned}
$$

where the composite residual is $\eta_i = \tau \cdot (W_i^{\text{obs}} - \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]) + \varepsilon_i$. By random assignment (Assumption 23.1), both ε_i and this unit-level difference $W_i^{\text{obs}} - \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]$ are uncorrelated with Z_i. Thus, the composite residual η_i is uncorrelated with Z_i. This in turn implies that least squares regression of Y_i^{obs} on the conditional expectation $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]$ will lead to an unbiased estimate of τ, the treatment effect of interest.

Unfortunately this linear regression is infeasible because we do not know the conditional expectation $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]$. However, we can estimate this conditional expectation. First let us write out the expected value of W_i^{obs} given Z_i as a function of Z_i – for those familiar with IV, the first-stage equation:

$$
\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i] = \pi_0 + \pi_1 \cdot Z_i,
$$

where $\pi_0 = \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i = 0]$ and $\pi_1 = \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i = 1] - \mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i = 0]$. Given one-sided noncompliance, $\pi_0 = 0$ ($Z_i = 0$ implies $W_i^{\text{obs}} = 0$), and π_1 equals $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i = 1]$, which is equal to the super-population proportion of compliers, π_{co}. Hence $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i] = \pi_1 \cdot Z_i = \pi_{\text{co}} \cdot Z_i$.

Using this expression we can rewrite Equation (23.8):

$$Y_i^{\text{obs}} = \alpha + \gamma \cdot Z_i + \eta_i, \quad \text{where } \gamma = \tau \cdot \pi_{\text{co}}. \tag{23.9}$$

Equation (23.9) is known as a *reduced form* in econometric terminology. Here the regression function does represent a conditional expectation, and as a result, its parameters can be consistently estimated by ordinary least squares. The least squares estimator, equal to the ratio of the covariance of Z_i and Y_i^{obs}, and the variance of Z_i will give an unbiased estimator of the composite coefficient $\gamma = \tau \cdot \pi_{\text{co}}$. With Z_i binary, this estimator will be equal to the difference in average outcomes by assignment, $\hat{\gamma} = \widehat{\text{ITT}_{\text{Y}}} = \overline{Y}_1^{\text{obs}} - \overline{Y}_0^{\text{obs}}$. Similarly, given the unconfoundedness of Z_i, regressing W_i^{obs} on Z_i will give an unbiased estimate of π_{co}. The estimator, with Z_i binary, equals $\hat{\pi}_{\text{co}} = \widehat{\text{ITT}_{\text{W}}} = \overline{W}_1^{\text{obs}} - \overline{W}_0^{\text{obs}}$.

Dividing the least squares estimator $\hat{\gamma} = \widehat{\text{ITT}_{\text{Y}}}$, by the estimator $\hat{\pi}_{\text{co}} = \widehat{\text{ITT}_{\text{W}}}$, gives the instrumental variables estimator $\hat{\tau}^{\text{iv}} = \widehat{\text{ITT}_Y}/\widehat{\text{ITT}_W}$ given earlier. For noncompliers, $W_i(z) = 0$ for $z = 0, 1$. Hence, given a binary assignment and treatment, using the linear parametric specification leads to an estimator identical to the moment-based estimator based on the potential outcomes approach. This estimator is also identical to that based on regressing Y_i on $\hat{\pi}_{\text{co}} \cdot Z_i$. The mechanical two-stage procedure of first regressing the receipt of treatment on the instrument to get an estimate of $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}}|Z_i]$, followed by regressing the outcome of interest on this predicted value of the receipt of treatment, is what led to the econometric terminology of TSLS, and the IV estimator is therefore also known as the TSLS estimator.

As just noted, we assumed in this derivation that the treatment effect is constant. Yet we did not make this same assumption in our potential outcomes discussion of the instrumental variables approach. As it turns out, this assumption is not necessary in either approach. Without it, we end up estimating the average treatment effect for compliers. More precisely, the numerical equivalence of the linear-equation IV estimand to the ratio of ITT effects does not rely on the assumption of a constant treatment effect. To see this, let τ_{late} be the average treatment effect for compliers, or the local average treatment effect, $\tau_{\text{late}} = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0)|G_i = c]$, and let ν_i be the unit-level difference between τ_i and τ, $\nu_i = Y_i(1) - Y_i(0) - \tau$. Again let $\alpha = \mathbb{E}_{\text{sp}}[Y_i(0)]$, and $\varepsilon_i = Y_i(0) - \alpha$. As before

$$Y_i^{\text{obs}} = Y_i(0) + W_i^{\text{obs}} \cdot (Y_i(1) - Y_i(0)),$$

which, given the definitions provided here, can be rewritten as

$$Y_i^{\text{obs}} = \alpha + W_i^{\text{obs}} \cdot \tau_{\text{late}} + \varepsilon_i + W_i^{\text{obs}} \cdot \nu_i. \tag{23.10}$$

We now have a new composite disturbance term, $\varepsilon_i + W_i^{\text{obs}} \cdot \nu_i$, which again is potentially correlated with W_i^{obs}. Thus an ordinary least squares regression of Y_i^{obs} on W_i^{obs} will not provide an unbiased estimate of τ.

However, just as ε_i is uncorrelated with Z_i, the second component of this new error term, $W_i^{\text{obs}} \cdot \nu_i$, is also uncorrelated with Z_i. To see this, consider this expectation

separately for $Z_i = 0$ and 1. Because $Z_i = 0$ implies $W_i^{\text{obs}} = 0$, it follows that $\mathbb{E}_{\text{sp}}[W_i^{\text{obs}} \cdot \nu_i | Z_i = 0] = 0$. To calculate the expectation given $Z_i = 1$, begin by expanding the expectation for both possible values of W_i^{obs}:

$$\begin{aligned}\mathbb{E}_{\text{sp}}[W_i \cdot \nu_i | Z_i = 1] &= \mathbb{E}_{\text{sp}}[0 \cdot \nu_i | Z_i = 1, W_i = 0] \cdot \Pr(W_i = 0 | Z_i = 1) \\ &\quad + \mathbb{E}_{\text{sp}}[1 \cdot \nu_i | Z_i = 1, W_i = 1] \cdot \Pr(W_i = 1 | Z_i = 1) \\ &= \mathbb{E}_{\text{sp}}[\nu_i | Z_i = 1, G_i = c] \cdot \pi_{\text{co}} = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0) - \tau | Z_i = 1, G_i = c] \\ &= \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0) - \tau | G_i = c] \cdot \pi_{\text{co}} = 0,\end{aligned}$$

by the definition of τ as the average treatment effect for compliers. Hence, looking at Equation (23.10), given that Z_i is uncorrelated with both elements of the error term, we can use the same argument as used earlier to motivate the moment estimator $\hat{\tau}^{\text{iv}}$.

23.9 NAIVE ANALYSES: "AS-TREATED," "PER PROTOCOL," AND UNCONFOUNDEDNESS

To put the simple instrumental variables analysis that is the main topic of this chapter in perspective, we conclude this chapter by discussing three other analyses, two of which are occasionally used in randomized experiments with noncompliance, and one of which serves to provide some perspective. (Note that we have already discussed one such alternative, the intention-to-treat analysis.) Like the IV approach, but unlike the ITT approach, these two additional analyses focus on the receipt of treatment, not merely on the causal effect of the assignment to treatment. Four analyses, IV, ITT, As-Treated, and Per Protocol, are identical when observed compliance is perfect, but they generally differ from one another when compliance is less than perfect. As will be seen here, however, in the presence of noncompliance, there is no compelling justification for these two other approaches. We present them merely to provide a better understanding of the competing intention-to-treat and instrumental variables methods.

23.9.1 As-Treated Analyses

The first of these two analyses is the "as-treated" approach. In this approach, the causal effect of the receipt of treatment is estimated as the difference in average outcomes by treatment received, W_i^{obs}:

$$\hat{\tau}_{\text{at}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}. \tag{23.11}$$

This approach would be justified, in the sense that it would give an unbiased estimate of the average treatment effect, if receipt of treatment W_i^{obs} were unconfounded. In general, however, it will not estimate a causal estimand. Here we explore the properties of this estimator. It will be convenient to take a super-population perspective, where we take the expectation over the randomization as well as over the distribution generated by random sampling from a large population.

The expectation of this estimator in the super-population is

$$\tau_{\text{at}} = \mathbb{E}_{\text{sp}}\left[Y_i^{\text{obs}} \middle| W_i^{\text{obs}} = 1\right] - \mathbb{E}_{\text{sp}}\left[Y_i^{\text{obs}} \middle| W_i^{\text{obs}} = 0\right].$$

Let us look at this difference in expectations under the two instrumental variables assumptions, random assignment and the exclusion restriction on noncompliers. Note that in our one-sided noncompliance case, units receiving the treatment must have $Z_i = 1$ and be compliers. Hence $\mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 1] = \mathbb{E}_{\text{sp}}[Y_i(1)|G_i = \text{co}]$. The second half of Equation (23.11) shows that units not receiving the treatment are a mixture of those assigned to the control and those assigned to the treatment who did not comply:

$$\begin{aligned}\mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 0] &= \mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 0, Z_i = 0] \cdot \Pr_{\text{sp}}(Z_i = 0|W_i^{\text{obs}} = 0) \\ &+ \mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 0, Z_i = 1] \cdot \Pr(Z_i = 1|W_i^{\text{obs}} = 0). \end{aligned} \tag{23.12}$$

With $p_Z = \Pr_{\text{sp}}(Z_i = 1)$, Bayes rule implies that, the probability that $Z_i = 1$ among those who do not take the treatment is equal to

$$\Pr_{\text{sp}}(Z_i = 1|W_i^{\text{obs}} = 0) = \frac{\pi_{\text{nc}} \cdot p_Z}{\pi_{\text{nc}} \cdot p_Z + 1 \cdot (1 - p_Z)}.$$

In the two expectations on the right-hand side of Equation (23.12), the second is simply the expected outcome for noncompliers under the control treatment. The first expectation in Equation (23.12) is a mixture of the expected value given the control treatment, for both compliers and noncompliers:

$$\mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|Z_i = 0, W_i^{\text{obs}} = 0] = \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] \cdot \pi_{\text{co}} + \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}] \cdot \pi_{\text{nc}}.$$

Combining all of the above, we can rewrite the expectation of the as-treated estimator as

$$\tau_{\text{at}} = \text{ITT}_{\text{Y,co}} + \Delta_{\text{co,nc}} \cdot \frac{\pi_{\text{nc}}}{p_Z \cdot \pi_{\text{nc}} + 1 - p_Z},$$

where, as before, $\Delta_{\text{co,nc}}$ is the expected difference in control outcomes for compliers and noncompliers:

$$\Delta_{\text{co,nc}} = \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}].$$

Unless compliance is perfect and there are no noncompliers ($\pi_{\text{nc}} = 0$), or the average control outcome is the same for compliers and noncompliers ($\Delta_{\text{co,nc}} = 0$, as implied by unconfoundedness of the treatment W_i), the expected value of $\hat{\tau}_{\text{at}}$ differs from the complier average causal effect.

This bias is easy to interpret: τ_{at} compares the average observed outcome given the active treatment to the average observed outcome given the control treatment. The first term is the average outcome given the active treatment for compliers, but the second term is an average of expected control outcome for compliers and noncompliers. If, as estimated in our example, noncompliers have lower average outcomes *without* the active treatment than compliers *without* the active treatment, this lowers the average outcome in the as-treated "control" group. Hence, the as-treated approach will overestimate the average treatment effect for compliers.

Let us illustrate this using the vitamin supplement data. In this sample the estimate of the average outcomes, with and without the supplements, are

$$\overline{Y}_{\text{c}}^{\text{obs}} = \frac{11,514 + 2,385}{11,514 + 2,385 + 74 + 34} = 0.9923,$$

and

$$\overline{Y}_{\text{t}}^{\text{obs}} = \frac{9,663}{9,663 + 12} = 0.9988.$$

Hence the as-treated estimate is

$$\hat{\tau}_{\text{at}} = 0.9988 - 0.9923 = 0.0065 \quad (\widehat{\text{s.e.}}\ 0.0008).$$

This estimator differs substantially from the IV estimate of 0.0033 calculated earlier. The reason can be seen by considering the estimates of the average outcomes of those assigned to the control for compliers and noncompliers separately. For noncompliers we estimated $\widehat{\mathbb{E}}_{\text{sp}}[Y_i(0)|G_i = \text{nc}] = 0.9859$, whereas for compliers we estimated $\widehat{\mathbb{E}}_{\text{sp}}[Y_i(0)|G_i = \text{co}] = 0.9955$, considerably higher. If the exclusion restriction holds, and hence our estimates of $\mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}]$ and $\mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}]$ are unbiased, the fact that the average outcome under the control treatment is higher for compliers than for noncompliers will lead the as-treated estimator to overestimate the complier average causal treatment effect.

23.9.2 Per Protocol Analyses

Now let us look at a second alternative to ITT and IV analyses, the per protocol analysis, in which only those units who are observed to comply with their assigned status are compared. In this analysis we therefore discard all observed noncompliers assigned to the treatment. Given the observable data, however, we cannot discard noncompliers assigned to the control. By one-sided noncompliance, these individuals automatically take the control; we would only be able to observe their compliance status if we instead saw them assigned to the treatment. If we could, in fact, discard *all* noncompliers, we would be left with only compliers, and then comparing their average outcomes by treatment status would estimate the average effect of receipt of treatment for compliers.

The per protocol analysis, however, discards only those noncompliers who do not comply with their *observed* treatment assignment and *not* those noncompliers who were assigned to the control group. The result is that the per protocol estimator, $\hat{\tau}_{\text{pp}}$, compares units receiving the treatment, that is, the compliers assigned to the treatment, to all units assigned to the control, with the latter a mixture of both compliers and noncompliers:

$$\hat{\tau}_{\text{pp}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{0}^{\text{obs}} = \frac{1}{N_t} \sum_{i=1}^{N} W_i^{\text{obs}} \cdot Y_i^{\text{obs}} - \frac{1}{N_0} \sum_{i=1}^{N} (1 - Z_i) \cdot Y_i^{\text{obs}},$$

which is biased for τ_{co}. Its expectation is:

$$\begin{aligned} \tau_{\text{pp}} &= \mathbb{E}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 1, Z_i = 1] - \mathbb{E}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 0, Z_i = 0] \\ &= \mathbb{E}_{\text{sp}}[Y_i(1)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)]. \end{aligned} \tag{23.13}$$

The last term in this expression is equal to $\mathbb{E}[Y_i(0)|G_i = \text{co}]\cdot\pi_{\text{co}} - \mathbb{E}[Y_i(0)|G_i = \text{nc}]\cdot\pi_{\text{nc}}$; hence we can rewrite τ_{pp} as

$$\begin{aligned}\tau_{\text{pp}} &= \mathbb{E}[Y_i(1) - Y_i(0)|G_i = \text{co}] \cdot \pi_{\text{co}} + (\mathbb{E}[Y_i(0)|G_i = \text{co}] - \mathbb{E}[Y_i(0)|G_i = \text{nc}]) \cdot \pi_{\text{nc}} \\ &= \text{ITT}_{\text{Y,co}} + \pi_{\text{nc}} \cdot \Delta_{\text{co,nc}}.\end{aligned}$$

Again, unless either π_{nc} or $\Delta_{\text{co,nc}}$ (or both) are equal to zero, $\hat{\tau}_{\text{pp}}$ will not give an unbiased estimate of the average effect of the treatment on compliers, even under the exclusion restriction for nomcompliers.

To illustrate this, we again use the Sommer-Zeger data to estimate τ_{pp}. Given these data, the first term of the estimand, $\mathbb{E}_{\text{sp}}[Y_i(1)|G_i = c] = \mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|W_i^{\text{obs}} = 1]$, is estimated as 0.9988 ($\widehat{\text{s. e.}}$ 0.0004), and the second, $\mathbb{E}_{\text{sp}}[Y_i(0)] = \mathbb{E}_{\text{sp}}[Y_i^{\text{obs}}|Z_i = 0]$, as $11,514/(11,514 + 74) = 0.9936$ ($\widehat{\text{s. e.}}$ 0.0007). Thus the per protocol estimate,

$$\hat{\tau}_{\text{pp}} = 0.9988 - 0.9936 = 0.0051 \qquad (\widehat{\text{s. e.}}\ 0.0008),$$

is again much larger than our estimate of the local average treatment effect, $\hat{\tau}_{\text{late}} = 0.0033$.

23.9.3 Analyses under Conditional Unconfoundedness Given the Instrument

A final analysis we wish to discuss briefly assumes unconfoundedness, like the "as-treated" analysis, but only conditional on the instrument. That is, it focuses on comparisons of units receiving and not receiving the treatment within subpopulations receiving the same level of assignment. Implicitly it treats the instrument as a covariate or pre-treatment variable that needs to be controlled for. In the current setting, with one-sided noncompliance among the subpopulation of units assigned to the control group, there are no units receiving the treatment, so we can do this only for the units assigned to the treatment.

The conditional unconfoundedness (cu) statistic focuses, for units assigned to the treatment, on the difference in average outcomes by receipt of treatment:

$$\hat{\tau}_{\text{cu}} = \overline{Y}_{\text{1t}} - \overline{Y}_{\text{1c}}.$$

This approach would be justified if, conditional on the assignment, receipt of treatment is random. Of course, the concern is that the very fact that these units, although assigned to the same level of the treatment, receive different levels of the treatment reflects systematic differences between these units. Let us look at the interpretation of this estimand under the instrumental variables assumptions. Given the definition of the compliance types, $\hat{\tau}_{\text{CU}}$ estimates

$$\tau_{\text{cu}} = \mathbb{E}_{\text{sp}}[Y_i(1)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}].$$

It is fundamentally comparing different subpopulations of units, under different treatment levels. More interesting, from a perspective of understanding the differences between the units, is to estimate the average outcomes for compliers and noncompliers

under the control treatment:

$$\Delta_{\text{co,nc}} = \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{co}] - \mathbb{E}_{\text{sp}}[Y_i(0)|G_i = \text{nc}],$$

because this compares the same potential outcomes for different subpopulations.

For the Sommer-Zeger data, we find

$$\hat{\tau}_{\text{cu}} = 0.9988 - 0.9859 = 0.0128 \qquad (\widehat{\text{s.e.}}\ 0.0024).$$

Survival rates for compliers assigned to the control treatment are substantially higher than for noncompliers assigned the active treatment, despite the fact that neither group took any active treatment.

23.10 CONCLUSION

The discussion in this chapter describes the instrumental variables approach to estimation of causal effects in randomized experiments with one-sided noncompliance, in settings where unconfoundedness of the receipt of treatment of interest is viewed as untenable. The approach exposited here relies on two key assumptions, which together replace the assumption of unconfoundedness of the receipt of treatment. The two assumptions are: unconfoundedness of the assignment to the active treatment (the instrument), rather than the receipt of treatment; and an exclusion restriction that rules out an effect of assignment on the outcome of interest for noncompliers. The first of these assumptions is implied by design in the randomized experiment setting. The second assumption relies more heavily on subject-matter knowledge, although it can be made more plausible by design measures such as double-blinding. Under those two assumptions, we can estimate the average effect of the treatment on a subset of the population, the so-called compliers, who comply with the treatment assignment irrespective of what that assignment is.

NOTES

Instrumental variables analyses have a long tradition in econometrics. The first cases of such analyses include S. Wright (1921, 1923), P. Wright (1928), Tinbergen (1930), and Haavelmo (1943). See Stock and Tregbi (2003) for a fascinating historical perspective. In these early analyses, as in most of the subsequent econometric discussions, models were typically specified in terms of linear equations. There was a clear sense, however, of what these equations meant: by assumption they describe behavioral or causal relationships between variables, not correlations, and thus they do not necessarily (although they may do so accidentally) describe conditional expectations.

Early on these models were characterized by constant treatment effects and tight parametric and distributional assumptions. More recently researchers have tried to relax these models by allowing for heterogeneity in the treatment effects and flexible functional forms. Heckman (1990) showed that conditions required for identification of the population average treatment effect in these models were very strong: essentially they