CHAPTER 20

Inference for General Causal Estimands

20.1 INTRODUCTION

Much of the discussion in the fourth part of the book focused on an average treatment effect as the causal estimand of primary interest. Although this is an important case, many of the analyses extend to other causal estimands in a conceptually straightforward manner. In this chapter we discuss some examples of other estimands, and show how some of the earlier analyses apply with other estimands.

In many cases concerning causal questions, average effects are the most obviously interesting objects. Sometimes the focus is on average effects after taking some transformation of the outcome, possibly involving pre-treatment variables, but this does not lead to any conceptual problems or operational difficulties when applying the analyses from the previous chapters. In other cases, however, the causal estimands are conceptually distinct from average treatment effects. This includes situations where the average effect is just one of the objects of interest, as well as settings where the primary object is not an average effect. For example, policy makers may be interested in the effect of a new program on specific parts of the distribution of outcomes. In a labor market training program, policy makers may be less interested in the effect of the program on relatively high-earning individuals, instead being more concerned about the effect on the left tail of the distribution. In that case, differences between quantiles of the two potential outcome distributions may be more interesting estimands. Alternatively, policy makers may be interested in the effect of a new program on inequality in outcomes, say, through the effect of the treatment on the variance or the inter-quartile range of the distribution of outcomes.

The approach to estimation and inference that is the focus here is model-based imputation, which has a number of conceptual advantages relative to other approaches. The most important one is that once the missing potential outcomes are imputed, any causal estimand of the type we consider can be directly calculated. As a result, under this approach, estimation of and inference for any causal estimand are conceptually straightforward. We can therefore consider a variety of estimands given the same model for the potential outcomes. In contrast, if one uses, say, regression estimates, one would implicitly be using different models for the potential outcomes when focusing on the average

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

effect versus the median effect of the treatment. The main alternative to using model-based imputation is weighting. Weighting approaches also can be used to estimate a variety of estimands, including some of the causal estimands considered in this chapter. As discussed in Chapter 12, a concern with weighting methods, specifically when weights must be estimated, is that the resulting estimators for causal effects can be particularly sensitive to the model for the propensity score. As a result, relatively minor changes in the specification for the propensity score can lead to substantial changes in the estimates of causal effects.

To implement the imputation of the missing potential outcomes, in our preferred approach we first estimate the propensity score. Next we block on the estimated propensity score. Within blocks defined by the estimated propensity score, we build parametric models for the outcome distributions conditional on the covariates, possibly with cross-block restrictions. We then use these models to impute the missing potential outcomes. Note that different models for imputation will generally be used for different outcome variables, an approach that fundamentally differs from the weighting or the pure propensity score approaches in important ways that give the model-based approaches substantial flexibility to obtain reliable causal effect estimates.

The rest of this chapter is organized as follows. In the next section we describe the data used in this chapter, originally collected and analyzed by Lalonde (1986), and previously used in Chapter 14, and we conduct some preliminary analyses on the data based on the previous chapters. In Section 20.3 we introduce some causal estimands that are of interest in the context of this application. In Section 20.4 we discuss the models for the potential outcomes used in this chapter. Next, in Section 20.5 we discuss the implementation of the methods. In Section 20.6 we return to the Lalonde data and report results for the application. Finally, Section 20.7 concludes.

20.2 THE LALONDE NSW OBSERVATIONAL JOB-TRAINING DATA

Here we return to the non-experimental part of the Lalonde data that we previously used in Chapter 14. The treated subsample consists of 185 men, and the control sample consists of 15,992 men. We first estimate the propensity score on the full sample of 16,177 men. As discussed in Chapter 14, there are substantial differences in the covariate distributions between the treated and control subsamples. We then use the trimming described in Chapter 16 to construct subsamples with more overlap. The estimated optimal threshold based on the methods from Chapter 16 is 0.0792. Dropping men with an estimated propensity score below 0.0792 or above $1-0.0792 = 0.9208$ leaves us with a subsample consisting of $N_{\text{c}} = 282$ men in the control sample and $N_{\text{t}} = 151$ men who received the job training. Table 20.1 gives summary statistics for this trimmed sample. In the trimmed sample, the overlap is still limited, with the normalized difference for some covariates as large as 0.54. Nevertheless, this is a substantial improvement over the original sample where some normalized differences were in excess of 2.0 (see Table 14.7 in Chapter 14).

Next we re-estimate the propensity score. This time the algorithm from Chapter 13 selects eight linear terms and six second-order terms. The parameter estimates for the propensity score models are reported in Table 20.2. Given this estimate of the propensity score, we construct blocks based on the methods from Chapter 17. The algorithm

Table 20.1. *Summary Statistics for Trimmed Lalonde Non-Experimental Data*

	Controls (N_c =282)		Trainees (N_c =151)		Nor Dif	Log Ratio of STD
	mean	(S.D.)	mean	(S.D.)		
black	0.92	(0.27)	0.95	(0.21)	0.15	−0.27
hispanic	0.06	(0.23)	0.03	(0.18)	−0.12	−0.26
age	25.13	(7.64)	25.70	(7.02)	0.08	−0.08
married	0.26	(0.44)	0.13	(0.34)	−0.32	−0.25
nodegree	0.64	(0.48)	0.74	(0.44)	0.22	−0.09
education	10.54	(3.05)	10.26	(2.05)	−0.11	−0.40
earn '74	2.75	(4.63)	1.67	(4.64)	−0.23	0.00
unempl '74	0.52	(0.50)	0.77	(0.42)	0.54	−0.17
earn '75	1.84	(2.66)	1.01	(1.97)	−0.36	−0.30
unempl '75	0.39	(0.49)	0.66	(0.48)	0.56	−0.03
pscore	0.26	(0.19)	0.51	(0.24)	1.15	0.22
linearized pscore	−1.26	(1.12)	0.07	(1.15)	1.18	0.03

Table 20.2. *Estimated Parameters of Propensity Score for the Trimmed Lalonde Non-Experimental Data*

Variable	Est	$\widehat{(\text{s.e.})}$	t-Stat
Intercept	−11.65	(0.13)	−92.6
Linear terms			
earn '74	0.15	(0.04)	3.4
unempl '74	−1.76	(1.17)	−1.5
earn '75	0.45	(0.38)	1.2
unempl '75	−0.95	(1.18)	−0.8
married	−3.15	(0.79)	−4.0
black	2.70	(0.55)	4.9
nodegree	1.33	(0.35)	3.8
age	0.55	(0.12)	4.7
Second-order terms			
age × age	−0.01	(0.00)	−5.1
married × nodegree	2.16	(0.86)	2.5
unempl '74 × age	0.12	(0.05)	2.4
earn '74× nodegree	−0.10	(0.05)	−2.0
earn '75 × black	−0.58	(0.38)	−1.5
unempl '74 × unempl '75	1.89	(1.20)	1.6

from that chapter leads to eight blocks. Summary statistics for the blocks are reported in Table 20.3.

Using the blocking estimator discussed in Chapter 17, including within-block regression adjustment, we obtain an estimate for the average effect for the treated equal to 2.33 (in thousands of dollars), with a standard error of 0.92. In this chapter, however, we are

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

Table 20.3. ***Optimal Subclassification for the Trimmed Lalonde Non-Experimental Data***

Subclass	Min P-Score	Max P-Score	# Controls	# Treated	t-Stat
1	0.00	0.17	96	4	0.8
2	0.17	0.18	11	5	−0.1
3	0.18	0.22	46	10	0.9
4	0.22	0.29	37	19	−0.2
5	0.29	0.40	38	19	0.4
6	0.40	0.47	15	19	0.4
7	0.47	0.73	29	38	0.5
8	0.73	1.00	10	38	1.6

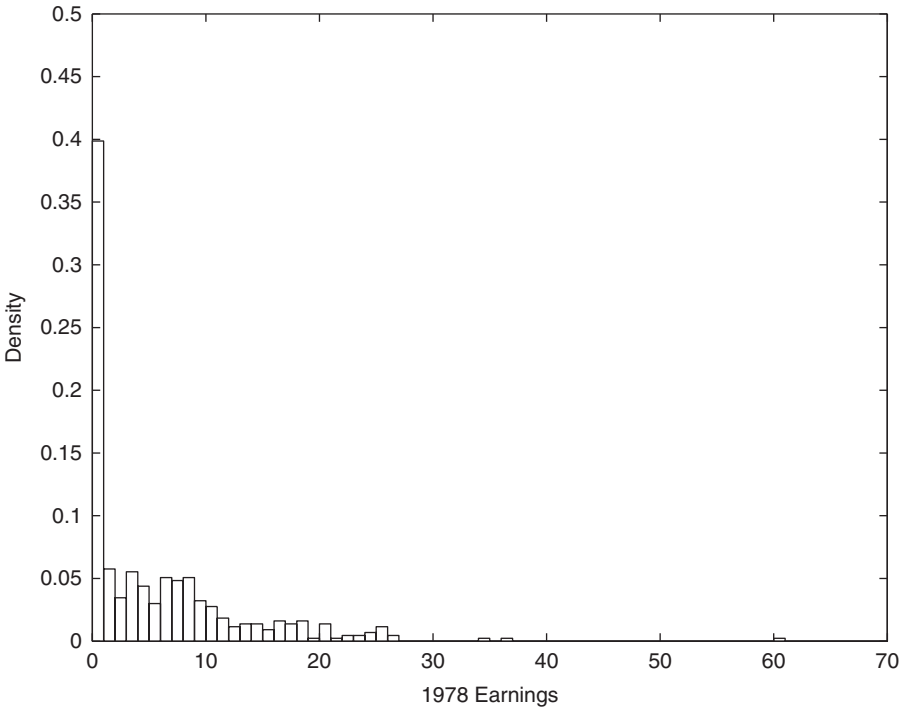

Figure 20.1. Histogram of 1978 earnings, trimmed Lalonde non-experimental data

interested in different causal estimands, and we will therefore build more flexible models for the conditional potential outcome distributions given the covariates and treatment levels. To inform the choice of such models, it is useful to inspect the marginal distributions of the observed outcomes, earnings in 1978 in thousands of dollars, either for the full trimmed sample, or separately by treatment group. Figure 20.1 presents a histogram of the outcome for the trimmed sample with 433 men. Two key features are the large proportion of individuals with zero earnings and the excess skewness and kurtosis of the distribution of earnings conditional on earnings being positive. In the trimmed sample, the proportion of men with zero earnings is 0.29, and the skewness among those positive earnings is 2.0, and the kurtosis is 10.7. It should also be noted that there is

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

an extreme value for the outcome. One individual in the trainee sample had subsequent 1978 earnings over $60,000. The next highest earning individual had yearly earnings less than $40,000. To put this in context, the average earnings for trainees in 1974 and 1975, respectively, are $1,670 and $1,010, with maximum values in the sample in those years equal to $31,000 and $11,500. Given that there are only 151 trainees in our sample, changing the 1978 earnings for this one man from over $60,000 to less than $40,000. would lower the point estimate of the average treatment effect substantially, from $2,327 to $2,170. We will attempt to take these features into account when developing models for the conditional distributions of the potential outcomes.

20.3 CAUSAL ESTIMANDS

At the very beginning of this book, in Chapter 1, we defined causal estimands to be a general function of the potential outcomes, the covariates, and the vector of treatment assignments,

$$\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}, \mathbf{W}). \tag{20.1}$$

Because of tradition and mathematical tractability we often focused on the finite-sample average effect

$$\tau_{\text{fs}} = \frac{1}{N} \sum_{i=1}^{N} \big(Y_i(1) - Y_i(0)\big),$$

or the super-population average treatment effect

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

In this chapter we consider two alternatives. For example, in a job-training program, policy makers may be interested in the effects of the program on the lower tail of the distribution. We can do this in a variety of ways. We may simply look at the average effect on a transformation of the original outcome. For example, we could define as the outcome whether an individual has positive earnings, or earnings exceeded some threshold level, such as some measure of the poverty level. Such transformations do not require any conceptual change in the methods discussed in previous chapters. Here we discuss some causal estimands that cannot be written as average effects on transformations of the original outcomes.

20.3.1 Quantile Treatment Effects

Distributional effects may conveniently be summarized by the difference in quantiles of the empirical distribution of the potential outcomes. For any outcome Y, with observations on N units $Y_1, \ldots, Y_N$, define q_Y^s to be the s^{th} quantile of the empirical

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

distribution of Y_i:

$$q_Y^s = \inf_q \left\{ q \in (-\infty, \infty) \;\middle|\; \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{Y_i \leq q} \geq s \right\}.$$

Then we can define the s^{th} quantile treatment effect as the difference of the s^{th} quantile of the $Y_i(1)$ and $Y_i(0)$ distributions:

$$\tau_{\text{quant}}^s = q_{Y(1)}^s - q_{Y(0)}^s.$$

We can estimate quantile treatment effects at different quantiles. Using the median gives a more robust estimate of a "typical" effect, although it should be kept in mind that the difference in medians by treatment status is generally *not* the median of the unit-level treatment effects. We can also look at differences in lower or higher quantiles to assess the effect of the treatment at the bottom or top of the distribution.

20.3.2 Causal Effects on Dispersion and Inequality

A conceptually very different estimand we consider in this chapter is a measure of inequality of the outcome distributions. A simple measure of this would be the difference in standard deviations in the two potential outcome distributions:

$$\tau_{\text{sd}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \big(Y_i(1) - \overline{Y}(1)\big)^2} - \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \big(Y_i(0) - \overline{Y}(0)\big)^2}.$$

Such a measure may be sensitive to the presence of outliers, in which case a more robust measure might be the interquartile range:

$$\tau_{\text{iqr}} = \big(q_{Y(1)}^{0.75} - q_{Y(1)}^{0.25}\big) - \big(q_{Y(0)}^{0.75} - q_{Y(0)}^{0.25}\big).$$

Alternatively, a common scale-free measure of inequality widely used in the social sciences is the so-called *Gini coefficient*. The Gini coefficient is often used to measure inequality of wealth. Given the ordered non-negative values $0 \leq Y_1 < Y_2 < \ldots < Y_N$, define the *Lorenz curve* as the piece-wise linear function $L^Y(y) : [0,1] \mapsto [0,1]$, going through the $N+1$ pairs of values $(F_0^Y, L_0^Y), \ldots, (F_N^Y, L_N^Y)$ where $(F_0^Y, L_0^Y) = (0,0)$, and, for $i = 1, \ldots, N$,

$$F_i^Y = \frac{i}{N}, \quad \text{and} \quad L_i^Y = \frac{\sum_{j=1}^{i} Y_i}{\sum_{j=1}^{N} Y_i}.$$

The Lorenz curve $L^Y(y)$ for, say, wealth, at a value $y \in [0,1]$, measures the share of the total wealth held by the bottom y proportion of the population. If wealth is shared equally, the Lorenz curve is equal to the forty-five-degree line, $L^Y(y) = y$. The Gini coefficient, denoted by G, is a scalar functional of the Lorenz curve, measuring the area

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

between the forty-five-degree line and the Lorenz curve as a share of the area underneath the forty-five-degree line:

$$G_Y = 1 - 2 \int_0^1 L^Y(y)\, dy. \tag{20.2}$$

If all values Z_i are identical, there is no inequality, the Lorenz curve equals to the forty-five-degree line, and the Gini coefficient is zero. The other extreme value is one, which occurs when all values Z_i are zero other than Z_N (and so all wealth is concentrated in the hands of one extremely wealthy individual). There are other measures of inequality available in the literature, but the Gini coefficient is widely used.

The causal estimand we focus on is the difference in Gini coefficients, the causal effect of the program on the Gini coefficient of the outcome distribution:

$$\tau^{\text{gini}} = G_{Y(1)} - G_{Y(0)}.$$

Policy makers may be interested to know whether the program increases inequality in earnings in the population.

20.3.3 Other Estimands

Here we focus primarily on two estimands, τ^s_{quant} for some specific values of s, and τ^{gini}. Many other estimands are possible. It is important, however, to note one common aspect of the two estimands we consider here. Both are functionals of the two marginal distributions of the potential outcomes, rather than functionals of the full joint distribution of the pair of potential outcomes. An example of a functional of the (full) joint distribution that cannot be written as a functional of the two marginal potential outcome distributions, and that is sometimes discussed in the literature, is the s^{th} quantile of the difference in potential outcomes, which is generally different from the difference in the s^{th} quantiles. In contrast to the distinction between the quantile of the difference and the difference in the quantiles, the average of the treatment effects is identical to the difference in the averages of the potential outcomes, because of the linearity of the expectations operator. In principle, the methods discussed in this chapter apply equally to estimands such as the median of the treatment effect, and we can directly apply the methods discussed in this chapter. In practice, though, it can be difficult to draw precise inferences about causal estimands that depend on the dependence structure of the potential outcomes. As discussed in the chapters on model-based inference in randomized experiments (Chapter 8), the data are not directly informative about the conditional dependence structure of the potential outcomes given covariates, and therefore prior information about the dependence structure may have important effects on posterior distributions, even in large samples. A question that sometimes arises is which object is of more interest, the median of the differences or the differences in the medians. In general, that question is difficult to answer without context. However, often policy makers contemplate exposing all units in a population (possibly homogeneous in characteristics) to the treatment versus no units. In that case, their decision should be based solely on the two marginal potential outcome distributions, not on the joint distribution of potential outcomes.

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

20.4 A MODEL FOR THE CONDITIONAL POTENTIAL OUTCOME DISTRIBUTIONS

The main model we consider is similar to the fourth model in Chapter 8 used for the model-based analysis of the experimental part of the Lalonde data. First we describe the general model and then extend it to the current setting that has substantial differences in the covariate distributions between the two treatment groups.

20.4.1 Single Block – Model I

We separately model the distribution of potential outcomes for each of the two treatment levels. For each treatment group we build a model for two parts of the conditional distribution given the pre-treatment variables X_i (all ten covariates other than the two indicators for ethnicity, `black` and `hispanic`, because there is little variation in ethnicity in the sample). Note that by unconfoundedness, these conditional distributions are free of dependence on W_i, that is, the same for units with $W_i = 0$ and $W_i = 1$. First, consider the probability of a positive value for $Y_i(0)$. A possible model for the event of a positive value for $Y_i(0)$ is a logistic model:

$$\Pr(Y_i(0) > 0|X_i, \theta) = \frac{\exp(\gamma_{c,0} + X_i\gamma_{c,1})}{1 + \exp(\gamma_{c,0} + X_i\gamma_{c,1})}, \tag{20.3}$$

and, analogously, we model the probability of a treated potential outcome for treated outcome as

$$\Pr(Y_i(1) > 0|X_i, \theta) = \frac{\exp(\gamma_{t,0} + X_i\gamma_{t,1})}{1 + \exp(\gamma_{t,0} + X_i\gamma_{t,1})}.$$

Second, we build models for the distributions of $Y_i(0)$ and $Y_i(1)$ conditional on a positive value for the potential outcome. Here, taking into account the excess skewness, we assume that the logarithm of the potential outcomes have normal distributions. Thus for the potential control outcome, we assume

$$\ln(Y_i(0))\,|Y_i(0) > 0, X_i, \theta \sim \mathcal{N}\left(\beta_{c,0} + X_i\beta_{c,1}, \sigma_c^2\right), \tag{20.4}$$

and for the potential treated outcome,

$$\ln(Y_i(1))\,|Y_i(1) > 0, X_i, \theta \sim \mathcal{N}\left((\beta_{t,0} + X_i\beta_{t,1}, \sigma_t^2\right),$$

where $Y_i(0)$ and $Y_i(1)$ are independent conditional on X_i and the parameters. Let $\theta = (\gamma_c, \gamma_t, \beta_c, \beta_t, \sigma_c^2, \sigma_t^2)$ denote the full parameter vector for these two distributions.

For convenience in conveying ideas, we specify a prior distribution for θ that is independent in its components and relatively dispersed, and for the regression parameters γ_c, γ_t, β_c, and β_t, we use normal prior distributions centered at zero with the variance equal to 10^2 times the identity matrix to capture relative ignorance about the components of these parameters. Similarly, for the variance parameters, σ_c^2 and σ_t^2, we use inverse Chi-squared distributions with parameters 1 and 0.01 respectively.

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

The implementation of this model using Markov-Chain Monte Carlo methods is similar to that in Chapter 8. In Table 20.4 we report summary statistics for the posterior distributions of the parameters. These are not of intrinsic interest but are useful to ensure that the posterior distribution is reasonable. Next we report in Table 20.5 the results for the causal estimands.

20.4.2 A Model with Multiple Blocks – Model II

The model in the previous subsection is a reasonable one in experimental settings where the covariate distributions are similar for treated and control units. However, in the current setting, even after the trimming, the covariate distributions are substantially different in the two treatment groups. We therefore consider a different model to allow for more flexibility. Specifically, we estimate separate models in each of the eight blocks of the propensity score. In this section we ignore the covariates. Using the methods from Chapter 17, we partition the range of the propensity score into $J = 8$ blocks, that is, intervals of the type $[b_{j-1}, b_j)$, where $b_0 = 0$ and $b_J = 1$, so that $\cup_{j=1}^{J}[b_{j-1}, b_j) = [0, 1)$, where $B_i(j) \in \{0, 1\}$ is an indicator. Let $B_i(j) \in \{0, 1\}$ be an indicator for unit i being in block j, for $j = 1, \ldots, J$:

$$B_i(j) = \begin{cases} 1 & \text{if } b_{j-1} \leq \hat{e}(X_i) < b_j, \\ 0 & \text{otherwise.} \end{cases}$$

Within each block and treatment level, we again specify a model for the event that the outcome is equal to zero, and a model for the outcome conditional on being positive. Specifically, for the control potential outcome $Y_i(0)$ in block j, we specify the model

$$\Pr(Y_i(0) > 0 | B_i(j) = 1, \theta) = \frac{\exp \gamma_c(j)}{1 + \exp \gamma_c(j)}, \tag{20.5}$$

and analogously, we model the probability of a positive outcome for treated outcome in this block as

$$\Pr(Y_i(1) > 0 | B_i(j) = 1, \theta) = \frac{\exp \gamma_t(j)}{1 + \exp \gamma_t(j)}.$$

Next, we build a model for the distribution of $Y_i(0)$ and $Y_i(1)$ in block j conditional on a positive value for the potential outcome. We assume

$$\ln\left(Y_i(0)\right) | Y_i(0) > 0, B_i(j) = 1, \theta \sim \mathcal{N}\left(\beta_c(j), \sigma_c^2\right), \tag{20.6}$$

and for the potential treated outcome,

$$\ln\left(Y_i(1)\right) | Y_i(1) > 0, B_i(j) = 1, \theta \sim \mathcal{N}\left(\beta_t(j), \sigma_t^2\right),$$

where $Y_i(0)$ and $Y_i(1)$ are assumed to be independent conditional on the block and the parameter. Here, for simplicity, we let the conditional variances differ by treatment status but not by block.

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

Table 20.4. *Single Block Model for Trimmed Lalonde Non-Experimental Data*

Sample	Controls						Treated					
	$q^{0.025}$	γ_c **med**	$q^{0.975}$	$q^{0.025}$	β_c **med**	$q^{0.975}$	$q^{0.025}$	γ_t **med**	$q^{0.975}$	$q^{0.025}$	β_t **med**	$q^{0.975}$
Intercept	0.25	**1.39**	3.20	1.02	**1.25**	1.48	0.89	**4.00**	9.05	1.43	**1.66**	1.89
`age`	−0.25	**−0.09**	0.05	−0.05	**−0.01**	0.03	−0.16	**0.06**	0.39	−0.04	**−0.01**	0.03
`married`	−2.82	**−0.00**	2.95	−0.38	**0.20**	0.78	−3.34	**4.15**	17.68	−0.52	**0.08**	0.69
`nodegree`	−2.55	**0.70**	4.02	−0.97	**−0.29**	0.38	−8.66	**−1.64**	4.19	−0.97	**−0.34**	0.30
`education`	−0.44	**0.04**	0.50	−0.03	**0.08**	0.18	−1.84	**−0.25**	0.95	−0.12	**0.01**	0.14
`earn '74`	−0.34	**0.06**	0.57	−0.07	**−0.01**	0.05	−0.71	**0.14**	1.45	−0.06	**0.01**	0.08
`unempl '74`	−3.02	**0.59**	4.85	−1.29	**−0.66**	−0.03	−3.98	**6.35**	23.43	−0.35	**0.47**	1.27
`earn '75`	−0.29	**0.44**	1.82	−0.01	**0.09**	0.20	−2.93	**−0.12**	2.22	−0.15	**0.03**	0.21
`unempl '75`	−4.87	**−0.94**	2.30	−0.42	**0.23**	0.88	−28.20	**−6.64**	4.07	−0.81	**−0.10**	0.61
σ				1.30	**1.43**	1.59				0.95	**1.08**	1.24

Table 20.5. ***Model-Based Analysis for Various Estimands for Trimmed Lalonde Non-Experimental Data***

	τ_{avg}			τ_{med}			$Y_i > 0$			$Y_i > 1$			Gini		
Model	$q^{.025}$	**med**	$q^{.975}$	$q^{.025}$	**med**	$q^{.975}$	$q^{.025}$	**med**	$q^{.975}$	$q^{.025}$	**med**	$q^{.975}$	$q^{.025}$	**med**	$q^{.975}$
I,fs	1.71	**3.11**	4.00	−0.52	**1.65**	3.35	−0.09	**0.11**	0.31	0.01	**0.18**	0.34	−0.23	**−0.16**	−0.07
I,sp	−0.10	**2.43**	4.82	−0.97	**2.17**	5.28	−0.16	**0.13**	0.40	−0.03	**0.22**	0.45	−0.25	**−0.13**	0.01
II,fs	1.12	**2.61**	4.00	−1.41	**0.85**	2.74	−0.07	**0.07**	0.24	−0.01	**0.12**	0.27	−0.20	**−0.13**	−0.06
II,sp	0.08	**2.25**	4.43	−1.81	**1.11**	4.02	−0.10	**0.12**	0.34	−0.01	**0.20**	0.40	−0.24	**−0.13**	−0.01
III,fs	2.09	**3.19**	4.00	−0.53	**1.48**	3.07	−0.02	**0.11**	0.24	0.07	**0.19**	0.30	−0.22	**−0.16**	−0.10
III,sp	0.38	**2.11**	3.85	−0.89	**1.77**	4.37	−0.08	**0.10**	0.26	0.02	**0.19**	0.34	−0.19	**−0.11**	−0.02

Note: Model I: single block, with covariates; Model II, eight blocks, no covariates; Model III: eight blocks, with covariates; fs, focus on finite sample causal estimand; sp, focus on super-population causal estimand.

Let $\theta = (\gamma_c(j), \gamma_t(j), \beta_c(j), \beta_t(j), j = 1, \ldots, J, \sigma_c^2, \sigma_t^2)$ denote the full parameter vector. Again, we use independent, fairly dispersed prior distributions for all elements of θ.

20.4.3 Multiple Blocks and Covariates – Model III

In our final model we incorporate both the block information and the covariates, and we combine the previous two specifications, using the analogous two-part model. We now specify for the control outcome the probability of a positive outcome as

$$\Pr\left(Y_i(0) > 0 | X_i, B_i(j) = 1, \theta\right) = \frac{\exp\left(\gamma_{c,0}(j) + X_i\gamma_{c,1}\right)}{1 + \exp\left(\gamma_{c,0}(j) + X_i\gamma_{c,1}\right)},$$

with, for positive income,

$$\ln\left(Y_i(0)\right) | Y_i(0) > 0, X_i, B_i(j) = 1, \theta \sim \mathcal{N}\left(\beta_{c,0}(j) + \beta_{c,1}X_i, \sigma_c^2\right),$$

and analogously for the treated outcome $Y_i(1)$. Thus we allow the intercepts in both models to be block-specific but restrict the slope coefficients to be identical. This restriction is partly motivated by the modest sample size. In some of the blocks there are only a few treated or only a few control units, so that it would be impossible to estimate precisely the slope coefficients separately within each block. An alternative would be a hierarchical structure where the parameters in each block are allowed to be different but are linked through a hierarchical structure through their prior distributions.

20.5 IMPLEMENTATION

We use Markov-Chain Monte Carlo methods to obtain draws from the posterior distribution of the parameters. Then we use two methods to obtain draws from the posterior distribution of the causal estimands. The first method follows closely that of Chapter 8. In this approach we draw values of the parameters from the posterior distribution given the observed data. We then use those parameter values in combination with the statistical model to impute the missing potential outcomes. Finally we calculate the estimand as a function of observed and imputed potential outcomes. Doing so repeatedly gives us the draws from the posterior distribution of the causal estimand.

However, this method does not always give credible results, and it is useful to sound a cautionary note. Specifically, in order for this first method to give accurate results, it relies heavily on the statistical model being a good approximation to the underlying distribution with regard to the particular estimand. For example, suppose we are interested in the average treatment effect for the treated, and we use the two-part model described in the previous section, with no covariates and a single block. We estimate the model for the control outcome using the control units. For this subsample, the proportion of zero outcomes is 0.31. Among the 69% control units with positive outcomes, the average and standard deviation of the logarithm of the outcome are 1.39 and 1.49 respectively. This implies, under the two-part model, that the expected value should be approximately $0.69 \cdot \exp\left(1.39 + 1.49^2/2\right) = 8.41$, whereas the actual average is 7.71. Because the model

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

is non-linear, at the fitted values the implied expectation is not necessarily equal to the sample average. In this simple example one could address this by estimating a linear model, but when we look at different estimands, unless the model fits the data well, it will not necessarily give good results for all estimands.

The second method addresses this as follows. We again draw parameter values from the posterior distribution of the parameters given the observed data. Now, however, for all units, we draw values for both potential outcomes. We then calculate the causal estimand as a function of these imputed potential outcomes, instead of combining observed and imputed outcomes. Implicitly this changes the focus from the sample causal estimand to the super-population causal estimand.

20.6 RESULTS FOR THE LALONDE DATA

For the Lalonde data we focus on estimands for the subsample of treated men. There is no interest in extending the labor market training program to the control individuals, only in assessing the benefits, if any, of the training program to those who took part in it. We focus on five estimands. The first is, for comparison purposes with earlier analyses, the average effect of the treatment on the treated:

$$\tau_{\text{fs},t} = \frac{1}{N_{\text{t}}} \sum_{i:W_i=1} \big(Y_i(1) - Y_i(0)\big).$$

The second estimand is, the difference in medians of $Y_i(1)$ and of $Y_i(0)$ for the treated units. First, extending the earlier definitions, we define the quantiles for the treated subsample as

$$q^s_{Y,t} = \inf_q \left\{ q : \frac{1}{N_t} \sum_{i:W_i=1} \mathbf{1}_{Y_i \le q} \ge s \right\}.$$

Then we define the s^{th} quantile treatment effect for the treated as the difference of the s^{th} quantile of the $Y_i(1)$ and $Y_i(0)$ distributions for the treated:

$$\tau^s_{\text{quant},t} = q^s_{Y(1),t} - q^s_{Y(0),t}.$$

Here we focus on the difference in medians,

$$\tau^{\text{med}}_t = q^{1/2}_{Y(1),t} - q^{1/2}_{Y(0),t}.$$

The next estimand is the causal effect of the treatment on the probability of having positive earnings,

$$\tau^{\text{pos}}_t = \frac{1}{N_t} \sum_{i:W_i=1} \big(\mathbf{1}_{Y_i(1)>0} - \mathbf{1}_{Y_i(0)>0}\big),$$

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

and the probability of having earnings exceeding 1 ($1,000),

$$\tau_t^{\geq 1} = \frac{1}{N_{\rm t}} \sum_{i:W_i=1} \left(\mathbf{1}_{\{Y_i(1)>1\}} - \mathbf{1}_{\{Y_i(0)>1\}} \right).$$

Finally, the fifth estimand is the difference in Gini coefficients in the $Y_i(1)$ and $Y_i(0)$ distributions for the treated units. Let $G_{Y(1),{\rm t}}$ denote the Gini coefficient for the $Y_i(1)$ distribution among the treated. Then the causal estimand we focus on is

$$\tau_t^{\rm gini} = G_{Y(1),t} - G_{Y(0),t}.$$

We estimate all three models, the one with covariates (Model I), with block indicators (Model II), and with both covariates and block indicators (Model III) on the Lalonde data.

In Table 20.4 we present posterior percentiles for the parameters of the first model. These parameter estimates are not of intrinsic interest and are presented here for completeness. In Table 20.5 we present posterior percentiles for the causal estimands. There are two rows for the first model, one for the finite-sample causal estimand, where only the control outcomes are imputed for all treated units, and one for the super-population causal estimand, where both control and treated outcomes are effectively imputed in the super-population. In general the estimates suggest that there is a positive effect of the treatment, as seen by the posterior medians for the average and median effects. It also suggests that the program may have led to a modest decrease in inequality as measured by the effect on the Gini coefficient.

20.7 CONCLUSION

In this chapter we discuss estimation of and inference for estimands other than average treatment effects. Under our preferred, model-based approach, there are no conceptual difficulties to studying general causal estimands. The approach of imputing the missing potential outcomes is valid in general. The main issue is that, in many cases, it becomes obvious that one has to be more careful in the choice of models. Depending on the choice of estimand, the results may be sensitive to particular modeling choices.

NOTES

Quantile treatment effects, defined as the difference in quantiles, as in the current chapter, have been considered previously by Lehman (1974). Causal effects of treatments on inequality measured through their effect on the Gini coefficient has been considered by Firpo (2003, 2007). In applied work, Bitler, Gelbach, and Hoynes (2006) study distributional effects beyond the average treatment effects and find, in the context of a randomized labor market program, that the effects are bigger at the lower tail of the distribution than in the upper tail.

https://doi.org/10.1017/CBO9781139025751.021 Published online by Cambridge University Press

For an early study of the sensitivity of estimates of causal effects to the choice of Bayesian model, see Rubin (1983); this example is discussed further in Gelman, Carlin, Stern, and Rubin (1995).

The discussion regarding the difference between, on the one hand, the difference between the medians of the potential outcomes by treatment status and, on the other hand, the median of the difference in potential outcomes is an old one. See for recent comments on this Manski (1996), Deaton (2010), and Imbens (2010).

The first general discussion of the imputation approach to inference for general causal estimands beyond average treatment effects is in Rubin (1978). Althauser and Rubin (1970) discuss computational issues.

Dehejia (2005b) and Manski (2013) discuss decision problems in a treatment effect context, where the intermediate focus is often on more complex estimands than simple average treatment effects.