Chapter 7

Donald B. Rubin

Yau Mathematical Sciences Center, Tsinghua University

June 30, 2021

Introduction

One of the more common ways of estimating causal effects with experimental, as well as observational, data in many disciplines is based on regression methods.

Typically an additive linear regression function is specified for the observed outcome as a function of a set of predictor variables and the treatment indicator.

Inferences, including point estimates, standard errors, tests, and confidence intervals, are based on standard least squares methods.

Introduction

Although popular, the use of these methods in this context is not without controversy, with some researchers arguing that experimental data should be analyzed based on randomization inference (see e.g. Freedman, 2006, Freedman, 2008a).

Both the Fisher approach and the Neyman methods as it core view the potential outcomes as fixed and the treatment assignments as the sole source of randomness. Here, as in Section 6.7, the starting point is an infinite super-population of units.

The chapter can be viewed as providing a bridge between the previous chapter and the next chapter, which is based on fully parametric models for imputation of the unobserved potential outcomes.

Introduction

There are four key features of the models considered.

1. We consider models for the observed outcomes, rather than models for the potential outcomes
2. We consider only models for the conditional mean, rather than for the full distribution
3. The estimand, here always an average treatment effect, is a parameter of the statistical model
4. In the current context of CRE, the validity of these models, is immaterial for the large—sample unbiasedness of the least squares estimator (OLS) of the average treatment effect

The LRC-CPPT Cholesterol Data

The Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT), designed to evaluate the effect of the drug cholestyramine on cholesterol levels is used to illustrate the concepts.

Here $N = 337$ individuals of which $N_t = 165$ were randomly assigned to receive cholestyramine and the remaining $N_c = 172$ received a placebo.
We observe two cholesterol measures recorded prior to the random assignment. The two measures differ in their timing.

1. `chol1`, was taken prior to a communication, sent to all 337 individuals in the study, about the benefits of a low-cholesterol diet
2. `chol2` was taken after this suggestion, but prior to the random assignment to cholestyramine or placebo.

The LRC-CPPT Cholesterol Data

We observe two outcomes.

1. `cholf` the an average of post-randomization cholesterol readings, averaged over two-month readings for a period of time averaging 7.3 years (primary outcome)
2. `comp`, the percentage of the nominally assigned dose (the same nominal dose of the drug or placebo, for the same time period)

The primary outcome in Efron and Feldman (1991) (who also analyzed the data) is the change in cholesterol level, relative to a weighted average of the two pre-treatment cholesterol levels, `cholp`$= 0.25 \cdot$ `chol1` $+ 0.75 \cdot$ `chol2`. This change in cholesterol levels is denoted `chold`=`cholf-cholp`.

Although individuals did not know whether they were assigned to cholestyramine or to the placebo, later we shall see that differences in side effects between the active drug and the placebo induced systematic differences in compliance behavior by treatment status.

The LRC-CPPT Cholesterol Data

In all RCTs compliance to the treatment is an issue when studying efficacy of the drug.

The availability of compliance data raises the issue of analyzing the differences between the effect of *being assigned* to the taking of cholestyramine and the effect of actually *taking* cholestyramine (this problem is discussed in Chapters 23-25).

Here we analyze the compliance measure solely as a secondary outcome. In general it is *not* appropriate to interpret either the difference in final cholesterol levels by assignment, *conditional* on observed compliance levels, or the difference in final cholesterol levels by actual dosage taken, as estimates of average causal effects.

The LRC-CPPT Cholesterol Data

Such causal interpretations would require strong additional assumptions beyond randomization. For example, to validate conditioning on observed compliance levels would require that observed compliance is a proper pretreatment variable unaffected by the assignment to treatment versus placebo.

Because observed compliance reflects behavior subsequent to the assignment, it may be affected by the treatment assigned. This is an assumption that can be assessed, and, in the current study we can reject the assumption that observed compliance is a proper covariate at conventional significance levels.

TThe LRC-CPPT Cholesterol Data

Table 7.1: Summary Statistics for PRC-CPPT Cholesterol Data

		Control (N_c =172)		Treatment (N_t =165)			
	Variable	Ave	Sample S.D	Ave	Sample S.D	min	max
pretreatment	chol1	297.1	23.1	297.0	20.4	247.0	442.0
	chol2	289.2	24.1	287.4	21.4	224.0	435.0
	cholp	291.2	23.2	289.9	20.4	233.0	436.8
posttreatment	cholf	282.7	24.9	256.5	26.2	167.0	427.0
	chold	-8.5	10.8	-33.4	21.3	-113.3	29.5
	comp	74.5	21.0	59.9	24.4	0	101.0

The Super-Population Average Treatment Effects

To be clear about this super-population perspective, define the two estimands

$$\tau_{\mathrm{FS}} = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i(1) - Y_i(0) \right),$$

the average effect of the treatment in the finite sample, and

$$\tau_{\mathrm{SP}} = \mathbb{E}_{\mathrm{SP}} \left[Y_i(1) - Y_i(0) \right],$$

the expected value of the unit-level treatment effect under the distribution induced by sampling from the super population, or, equivalently, the average treatment effect in the super population.

The Super-Population Average Treatment Effects

Define the super-population average and variance of the two potential outcomes conditional on the covariates or pretreatment variables, *e.g.*, $X_i = x$,

$$\mu_c(x) = \mathbb{E}_{\text{SP}}\left[Y_i(0)|X_i = x\right], \qquad \mu_t(x) = \mathbb{E}_{\text{SP}}\left[Y_i(1)|X_i = x\right],$$
$$\sigma_c^2(x) = \mathbb{V}_{\text{SP}}\left(Y_i(0)|X_i = x\right), \qquad \text{and} \quad \sigma_t^2 = \mathbb{V}_{\text{SP}}\left(Y_i(1)|X_i = x\right),$$

and let the mean and variance of the unit-level treatment effects at $X_i = x$ be denoted by

$$\tau(x) = \mathbb{E}_{\text{SP}}(Y_i(1) - Y_i(0)|X_i = x], \qquad \text{and} \quad \sigma_{ct}^2(x) = \mathbb{V}_{\text{SP}}\left(Y_i(1) - Y_i(0)|X_i = x\right),$$

respectively.

The Super-Population Average Treatment Effects

In addition, denote the marginal means and variances

$$\mu_c = \mathbb{E}_{\text{SP}}\left[Y_i(0)\right], \qquad \mu_t = \mathbb{E}_{\text{SP}}\left[Y_i(1)\right],$$

$$\sigma_c^2 = \mathbb{V}_{\text{SP}}\left(Y_i(0)\right), \qquad \text{and} \quad \sigma_t^2 = \mathbb{V}_{\text{SP}}\left(Y_i(1)\right).$$

Note that the two marginal means are equal to the expectation of the corresponding conditional means:

$$\mu_c = \mathbb{E}_{\text{SP}}\left[\mu_c(X_i)\right], \qquad \text{and} \quad \mu_t = \mathbb{E}_{\text{SP}}\left[\mu_t(X_i)\right],$$

but, by the law of iterated expectations, the marginal variance differs from the average of the conditional variance by the variance of the conditional mean:

The Super-Population Average Treatment Effects

$$\sigma_c^2 = \mathbb{E}_{\mathrm{SP}}\left[\sigma_c^2(X_i)\right] + \mathbb{V}_{\mathrm{SP}}\left(\mu_c(X_i)\right), \qquad \text{and } \sigma_t^2 = \mathbb{E}_{\mathrm{SP}}\left[\sigma_t^2(X_i)\right] + \mathbb{V}_{\mathrm{SP}}\left(\mu_t(X_i)\right). \quad (1)$$

Finally, let

$$\mu_X = \mathbb{E}_{\mathrm{SP}}\left[X_i\right], \qquad \text{and } \Omega_X = \mathbb{V}_{\mathrm{SP}}(X_i) = \mathbb{E}_{\mathrm{SP}}\left[(X_i - \mu_X)^T (X_i - \mu_X)\right],$$

denote the super-population mean and covariance matrix of the row-vector of covariates X_i, respectively.

Linear Regression With No Covariates

We maintain the assumption of a CRE and specify a linear regression function as

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + \varepsilon_i,$$

where the unobserved residual ε_i captures unobserved determinants of the outcome.

The OLS estimator for τ is based on minimizing the sum of squared residuals over α and τ,

$$(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}) = \arg\min_{\tau,\alpha} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i \right)^2,$$

with solutions

$$\hat{\tau}^{\text{ols}} = \frac{\sum_{i=1}^{N} \left(W_i - \overline{W} \right) \cdot (Y_i^{\text{obs}} - \overline{Y}^{\text{obs}})}{\sum_{i=1}^{N} \left(W_i - \overline{W} \right)^2}, \qquad \text{and } \hat{\alpha}^{\text{ols}} = \overline{Y}^{\text{obs}} - \hat{\tau}^{\text{ols}} \cdot \overline{W},$$

Linear Regression With No Covariates

where

$$\overline{Y}^{\text{obs}} = \frac{1}{N}\sum_{i=1}^{N} Y_i^{\text{obs}} \qquad \text{and} \quad \overline{W} = \frac{1}{N}\sum_{i=1}^{N} W_i = \frac{N_t}{N}.$$

Simple algebra shows that in this case

$$\hat{\tau}^{\text{ols}} = \overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}} = \hat{\tau}^{\text{dif}},$$

where, as before, $\overline{Y}_t^{\text{obs}} = \sum_{i:W_i=1} Y_i^{\text{obs}}/N_t$ and $\overline{Y}_c^{\text{obs}} = \sum_{i:W_i=0} Y_i^{\text{obs}}/N_c$.

Linear Regression With No Covariates

The assumptions traditionally used in the least squares approach are that the residuals ε_i are independent of, or at least uncorrelated with, the treatment indicator W_i.

This assumption is difficult to evaluate as the residuals in general is a vague notion of capturing unobserved factors affecting the outcome.

Statistical textbooks, therefore, often stress that in observational studies the regression estimate $\hat{\tau}^{\text{ols}}$ measures only the association between the two random variables W_i and Y_i^{obs}, and that a causal interpretation is generally not warranted.

In the current context, however, we already have a formal justification for the causal interpretation of $\hat{\tau}^{\text{ols}}$ because it is identical to $\overline{Y}_t^{\text{obs}} - \overline{Y}_c^{\text{obs}}$, which was shown to be unbiased for τ_{FS} and τ_{SP}.

Linear Regression With No Covariates

Nevertheless, it is useful to justify the causal interpretation of $\hat{\tau}^{\text{ols}}$ more directly in terms of the standard justification for regression methods, using the assumptions that random sampling created the sample and a CRE generated the observed data from that sample.

Let α be the population average outcome under the control, $\alpha = \mu_c = \mathbb{E}_{\text{SP}}[Y_i(0)]$, and recall that τ_{SP} is the super-population average treatment effect, $\tau_{\text{SP}} = \mu_t - \mu_c = \mathbb{E}_{\text{SP}}[Y_i(1) - Y_i(0)]$. Now,

$$\varepsilon_i = Y_i(0) - \alpha + W_i \cdot (Y_i(1) - Y_i(0) - \tau_{\text{SP}}) = \begin{cases} Y_i^{\text{obs}} - \alpha & \text{if } W_i = 0, \\ Y_i^{\text{obs}} - \alpha - \tau_{\text{SP}} & \text{if } W_i = 1. \end{cases}$$

Linear Regression With No Covariates

Then we can write

$$\varepsilon_i = Y_i^{\text{obs}} - (\alpha + \tau_{\text{SP}} \cdot W_i),$$

and thus we can write the observed outcome as

$$Y_i^{\text{obs}} = \alpha + \tau_{\text{SP}} \cdot W_i + \varepsilon_i.$$

Random sampling allows us to view the potential outcomes as random variables. In combination with random assignment this implies that assignment is independent of the potential outcomes,

$$\Pr\left(W_i = 1 | Y_i(0), Y_i(1)\right) = \Pr\left(W_i = 1\right),$$

or in Dawid's (1979) "$\perp\!\!\!\perp$" independence notation,

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

Linear Regression With No Covariates

The definition of the residual, in combination with random assignment and random sampling from a super-population, implies that the residual has mean zero conditional on the treatment indicator in the population:

$$\mathbb{E}_{\mathrm{SP}}[\varepsilon_i|W_i=0]=\mathbb{E}_{\mathrm{SP}}\left[Y_i(0)-\alpha|W_i=0\right]=\mathbb{E}_{\mathrm{SP}}\left[Y_i(0)\right]-\alpha]=0,$$

and

$$\mathbb{E}_{\mathrm{SP}}[\varepsilon_i|W_i=1]=\mathbb{E}_{\mathrm{SP}}\left[Y_i(1)-\alpha-\tau_{\mathrm{SP}}|W_i=1\right]=\mathbb{E}_{\mathrm{SP}}\left[Y_i(1)-\alpha-\tau_{\mathrm{SP}}|W_i=1\right]=0,$$

so that

$$\mathbb{E}_{\mathrm{SP}}[\varepsilon_i|W_i=w]=0, \qquad \text{for} \quad w=0,1.$$

The fact that the conditional mean of ε_i given W_i is zero in turn implies unbiasedness of the the least squares estimator, $\hat{\tau}^{\mathrm{ols}}$ for $\tau_{\mathrm{SP}}=\mathbb{E}_{\mathrm{SP}}\left[Y_i(1)-Y_i(0)\right]$, over the distribution induced by random sampling.

The [illegible] estimator [illegible]

Linear Regression With No Covariates

The fact that the conditional mean of ε_i given W_i is zero in turn implies unbiasedness of the the least squares estimator, $\hat{\tau}^{\text{ols}}$ for $\tau_{\text{SP}} = \mathbb{E}_{\text{SP}}[Y_i(1) - Y_i(0)]$, over the distribution induced by random sampling.

The above derivation shows how properties of residuals commonly asserted as assumptions in least squares analyses, actually follow from random sampling and random assignment, and thus have a scientific basis in the context of a CRE.

Linear Regression With No Covariates

Another way, which is closer to the way we will do this for the general case with covariates, is to consider the super population limits of the estimators.

The estimators are defined as

$$(\hat{\alpha}^{\text{ols}}, \hat{\tau}^{\text{ols}}) = \arg\min_{\alpha,\tau} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i \right)^2 .$$

Under some regularity conditions, these estimators converge, as the sample size goes to infinity, to the population limits (α^*, τ^*) which minimizes:

$$(\alpha^*, \tau^*) = \arg\min_{\alpha,\tau} \mathbb{E}_{\text{SP}} \left[\frac{1}{N} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i \right)^2 \right] = \arg\min_{\alpha,\tau} \mathbb{E}_{\text{SP}} \left[\left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i \right)^2 \right]$$

Linear Regression With No Covariates

This implies that the population limit is $\tau^* = \mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 1] - \mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 0]$.

Random assignment of W_i implies

$$\mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 1] - \mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 0] = \mathbb{E}_{\mathrm{SP}}[Y_i(1) - Y_i(0)] = \tau_{\mathrm{SP}},$$

so that the population limit of the OLS estimator is equal to $\tau^* = \tau_{\mathrm{SP}}$.

Linear Regression With No Covariates

Now let us analyze the least squares approach to inference.

Let us first assume homoskedasticity: $(\sigma^2_{Y|W} = \sigma^2_c = \sigma^2_t)$.

Using OLS, the variance of the residuals would be estimated as

$$\hat{\sigma}^2_{Y|W} = \frac{1}{N-2}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 = \frac{1}{N-2}\sum_{i=1}^{N}\left(Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}\right)^2,$$

where the estimated residual is $\hat{\varepsilon}_i = Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}$, and the predicted value $\hat{Y}_i^{\text{obs}}$ is

$$\hat{Y}_i^{\text{obs}} = \begin{cases} \hat{\alpha}^{\text{ols}} & \text{if } W_i = 0, \\ \hat{\alpha}^{\text{ols}} + \hat{\tau}^{\text{ols}} & \text{if } W_i = 1. \end{cases}$$

Linear Regression With No Covariates

The OLS variance estimate can be rewritten as

$$\hat{\sigma}^2_{Y|W} = \frac{1}{N-2}\left(\sum_{i:W_i=0}\left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right)^2 + \sum_{i:W_i=1}\left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}}\right)^2\right),$$

which is equivalent to our calculation of s^2 in Chapter 6.

The conventional estimator for the sampling variance of $\hat{\tau}_{\text{ols}}$ is then

$$\hat{\mathbb{V}}^{\text{homosk}} = \frac{\hat{\sigma}^2_{Y|W}}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)^2} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t}\right).$$

This expression is equal to $\hat{\mathbb{V}}^{\text{const}}$.

Linear Regression With No Covariates

For comparison with subsequent results, let $p = \text{plim}(N_t/N)$.

Then, as the sample size increases, the normalized sampling variance estimator converges in probability to

$$N \cdot \hat{\mathbb{V}}^{\text{homosk}} \xrightarrow{p} \frac{\sigma^2_{Y|W}}{p \cdot (1-p)}. \tag{2}$$

Note, that random assignment implies independence between assignments and potential outcomes, however it implies only zero correlation between the assignment and the residual.

Linear Regression With No Covariates

Yet we rely on independence when assuming the variance to be homoskedastic. In many cases, the homoskedasticity assumption will not be warranted, and one may wish to use an estimator for the sampling variance of $\hat{\tau}^{\text{ols}}$ that allows for heteroskedasticity.

The standard robust sampling variance estimator for OLS estimators is

$$\hat{\mathbb{V}}^{\text{hetero}} = \frac{\sum_{i=1}^{N} \hat{\varepsilon}_i^2 \cdot \left(W_i - \overline{W}\right)^2}{\left(\sum_{i=1}^{N} \left(W_i - \overline{W}\right)^2\right)^2}.$$

Defining, as the previous chapter,

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right)^2, \qquad \text{and} \quad s_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}}\right)^2,$$

Linear Regression With No Covariates

we can write the variance estimator under heteroskedasticity as

$$\hat{\mathbb{V}}^{\text{hetero}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} = (\hat{\mathbb{V}}^{\text{neyman}}.$$

So, in the case without additional predictors, the regression approach leads to sampling variance estimators that are familiar from the discussion in the previous chapter. It does, however, provide a different perspective on these results as it allows for a natural and simple extension to the case with additional covariates.

Linear Regression With Additional Covariates

Let

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + X_i\beta + \varepsilon_i, \tag{3}$$

where X_i is a row-vector of covariates (*i.e.*, pretreatment variables).

We estimate the regression coefficients again using OLS:

$$(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}, \hat{\beta}^{\text{ols}}) = \arg\min_{\tau,\alpha,\beta} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2.$$

The first question we address in this section concerns the causal interpretation of $\hat{\tau}^{\text{ols}}$. We are not interested *per sé* in the value of the "nuisance" parameters, β and α.

Moreover, we will *not* make the assumption that (3) is correctly specified nor that the conditional expectation of Y_i^{obs} is actually linear in X_i and W_i.

Linear Regression With Additional Covariates

However, in order to be precise about the causal interpretation of $\hat{\tau}^{\text{ols}}$, it is useful to define the limiting values to which the OLS estimators converge as the sample gets large.

We refer to these limiting values as the super–population values corresponding to the estimators, and denote them as before with a superscript $*$.

Using this notation, under some regularity conditions, $(\hat{\alpha}_{rols}, \hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}})$ converge to $(\alpha^*, \tau^*, \beta^*)$, defined as

$$(\alpha^*, \tau^*, \beta^*) = \arg\min_{\alpha,\beta,\tau} \mathbb{E}\left[\left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2\right].$$

Linear Regression With Additional Covariates

However, in order to be precise about the causal interpretation of $\hat{\tau}^{\text{ols}}$, it is useful to define the limiting values to which the OLS estimators converge as the sample gets large.

We refer to these limiting values as the super–population values corresponding to the estimators, and denote them as before with a superscript $*$.

Using this notation, under some regularity conditions, $(\hat{\alpha}_{\textit{rols}}, \hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}})$ converge to $(\alpha^*, \tau^*, \beta^*)$, defined as

$$(\alpha^*, \tau^*, \beta^*) = \arg\min_{\alpha,\beta,\tau} \mathbb{E}\left[\left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2\right].$$

Linear Regression With Additional Covariates

However, in order to be precise about the causal interpretation of $\hat{\tau}^{\text{ols}}$, it is useful to define the limiting values to which the OLS estimators converge as the sample gets large.

We refer to these limiting values as the super–population values corresponding to the estimators, and denote them as before with a superscript $*$.

Using this notation, under some regularity conditions, $(\hat{\alpha}_{rols}, \hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}})$ converge to $(\alpha^*, \tau^*, \beta^*)$, defined as

$$(\alpha^*, \tau^*, \beta^*) = \arg\min_{\alpha,\beta,\tau} \mathbb{E}\left[\left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2\right].$$

Linear Regression With Additional Covariates

These population values are generally well-defined (subject, essentially, only to finite moment conditions and positive definiteness of Ω_X), even if the conditional expectation of the observed outcome given covariates is not linear in the covariates.

In this case with additional predictors, it is no longer true that $\hat{\tau}^{\text{ols}}$ is unbiased for τ_{SP} in finite samples. However, irrespective of whether the regression function is truly linear in the covariates in the population, $\hat{\tau}^{\text{ols}}$ is unbiased in large samples for τ_{SP}.

In other words, τ^*, the probability limit of the estimator, is equal to the population average treatment effect τ_{SP}.

In addition,

Linear Regression With Additional Covariates

Theorem

Suppose we conduct a completely randomized experiment in a sample drawn at random from an infinite population. Then, (i)

$$\tau^* = \tau_{\mathrm{SP}},$$

and (ii),

$$\sqrt{N} \cdot \left(\hat{\tau}^{\mathrm{ols}} - \tau_{\mathrm{SP}}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}\left[(W_i - p)^2 \cdot \left(Y_i^{\mathrm{obs}} - \alpha^* - \tau_{\mathrm{SP}} \cdot W_i - X_i\beta^*\right)^2\right]}{p^2 \cdot (1-p)^2}\right).$$

The proof of (ii) and of subsequent results, are given in the Appendix to this chapter.

Linear Regression With Additional Covariates

Proof of Theorem 1(i): Consider the limiting objective function:

$$Q(\alpha, \tau, \beta) = \mathbb{E}[(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta)^2]$$

$$= \mathbb{E}\left[\left(Y_i^{\text{obs}} - \tilde{\alpha} - \tau \cdot W_i - (X_i - \mu_X)\beta\right)^2\right],$$

where $\tilde{\alpha} = \alpha + \mu_X\beta$, with $\mu_X = \mathbb{E}[X_i]$. Minimizing the right hand side over $\tilde{\alpha}$, τ and β leads to the same values for τ and β as minimizing the left hand side over α, τ, and β, with the least squares estimate of $\tilde{\alpha}$ equal $\hat{\alpha} + \hat{\beta}'\mu_X$. Next,

$$Q(\tilde{\alpha}, \tau, \beta) = \mathbb{E}_{\text{SP}}\left[\left(Y_i^{\text{obs}} - \tilde{\alpha} - \tau \cdot W_i - (X_i - \mu_X)\beta\right)^2\right]$$

$$= \mathbb{E}_{\text{SP}}\left[\left(Y_i^{\text{obs}} - \tilde{\alpha} - \tau \cdot W_i\right)^2\right] + \mathbb{E}_{\text{SP}}\left[((X_i - \mu_X)\beta)^2\right]$$

$$-2 \cdot \mathbb{E}_{\text{SP}}\left[\left(Y_i^{\text{obs}} - \tilde{\alpha} - \tau \cdot W_i\right) \cdot (X_i - \mu_X)\beta\right]$$

Linear Regression With Additional Covariates

$$= \mathbb{E}_{\mathrm{SP}}\left[\left(Y_i^{\mathrm{obs}} - \tilde{\alpha} - \tau \cdot W_i\right)^2\right] + \mathbb{E}_{\mathrm{SP}}\left[((X_i - \mu_X)\beta)^2\right] - 2 \cdot \mathbb{E}_{\mathrm{SP}}\left[Y_i^{\mathrm{obs}} \cdot (X_i - \mu_X)\beta\right], \tag{4}$$

because

$$\mathbb{E}_{\mathrm{SP}}\left[(X_i - \mu_X)\beta\right] = 0, \qquad \text{and} \quad \mathbb{E}_{\mathrm{SP}}\left[\tau \cdot W_i \cdot (X_i - \mu_X)\beta\right] = 0,$$

Because the last two terms in (4) do not depend on $\tilde{\alpha}$ or τ, minimizing (4) over τ and α is equivalent to minimizing the objective function without the additional covariates,

$$\mathbb{E}_{\mathrm{SP}}\left[\left(Y_i^{\mathrm{obs}} - \tilde{\alpha} - \tau \cdot W_i\right)^2\right].$$

Linear Regression With Additional Covariates

This leads to the solutions

$$\tilde{\alpha}^* = \mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 0] = \mathbb{E}_{\mathrm{SP}}\left[Y_i(0)|W_i = 0\right] = \mathbb{E}_{\mathrm{SP}}\left[Y_i(0)\right],$$

and

$$\tau^* = \mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 1] - \mathbb{E}_{\mathrm{SP}}[Y_i^{\mathrm{obs}}|W_i = 0] = \mathbb{E}_{\mathrm{SP}}[Y_i(1)|W_i = 1] - \mathbb{E}_{\mathrm{SP}}[Y_i(0)|W_i = 0] = \tau_{\mathrm{SP}}$$

Thus, the OLS estimator is consistent for τ_{SP}. $\square$

What is important in the first part of the result is that the consistency of the OLS estimator for τ_{SP} does *not* depend on the correctness of the specification of the regression function in a CRE.

No matter how nonlinear the conditional expectations of the potential outcomes given the covariates is in the super population, the OLS estimator is consistent for estimating the population average treatment effect.

Linear Regression With Additional Covariates

The key insight into this result is that, by randomizing treatment assignment, the super-population correlation between the treatment indicator and the covariates is zero.

Even though in finite samples the actual correlation may differ from zero, in large samples this correlation will vanish.

The fact that in finite samples the correlation may differ from zero is what leads to the possibility of finite sample bias.

Although the inclusion of the additional covariates does not matter for the limit of the corresponding estimator, it does matter for the sampling variance of the estimators.

Let us interpret the sampling variance in some special cases.

Linear Regression With Additional Covariates

Suppose that,

$$\mathbb{E}_{\mathrm{SP}}[Y_i(0)|X_i = x] = \alpha_c + x\beta, \qquad \text{and } \mathbb{E}_{\mathrm{SP}}[Y_i(1)|X_i = x] = \alpha_t + x\beta,$$

so that, in combination with random assignment, we have

$$\mathbb{E}_{\mathrm{SP}}\left[Y_i^{\mathrm{obs}} \middle| X_i = x, W_i = t \right] = \alpha_c + \tau_{\mathrm{SP}} \cdot t + \beta' x,$$

where $\tau_{\mathrm{SP}} = \alpha_t - \alpha_c$.

Instead of the unconditional variance of the potential outcomes, as in the expression for the sampling variance in the case without covariates in (2),we now have the conditional variance of the outcome given the covariates.

Linear Regression With Additional Covariates

If the covariates explain much of the variation in the potential outcomes, so that $\sigma^2_{Y|W,X}$ is substantially smaller than $\sigma^2_{Y|W}$, then including the covariates in the regression model will lead a considerable increase in precision.

The price paid for the increase in precision from including covariates is relatively minor. Instead of having (exact) unbiasedness of the estimator in finite samples, unbiasedness now only holds approximately, that is, in large samples.

The sampling variance for the average treatment effect can be estimated easily using standard least squares methods.

Linear Regression With Additional Covariates

Substituting averages for the expectations, and least squares estimates for the unknown parameters, we estimate the sampling variance as

$$\hat{\mathbb{V}}_{\mathrm{SP}}^{\mathrm{hetero}} = \frac{1}{N\left(N-1-\dim(X_i)\right)} \cdot \frac{\sum_{i=1}^{N}\left(W_i-\overline{W}\right)^2 \cdot \left(Y_i^{\mathrm{obs}} - \hat{\alpha}^{\mathrm{ols}} - \hat{\tau}^{\mathrm{ols}} - X_i\hat{\beta}^{\mathrm{ols}}X_i\right)^2}{\left(\overline{W}\cdot(1-\overline{W})\right)^2}.$$

A more precise estimator of the sampling variance under assumption of homoskedasticity is of the form:

$$\hat{\mathbb{V}}_{\mathrm{SP}}^{\mathrm{homo}} = \frac{1}{N\left(N-1-\dim(X_i)\right)} \cdot \frac{\sum_{i=1}^{N}\left(Y_i^{\mathrm{obs}} - \hat{\alpha}^{\mathrm{ols}} - \hat{\tau}^{\mathrm{ols}} - X_i\hat{\beta}^{\mathrm{ols}}X_i\right)^2}{\overline{W}\cdot(1-\overline{W})}.$$

Linear Regression With Covariates and Interactions

We specify the regression function as

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + X_i\beta + W_i \cdot (X_i - \overline{X})\gamma + \varepsilon_i.$$

We include the interaction of the treatment indicator with the covariates in deviations from their sample means to simplify the relationship between the population limits of the estimators for the parameters of the regression function and τ_{SP}.

Let $\hat{\alpha}^{\text{ols}}$, $\hat{\tau}^{\text{ols}}$, $\hat{\beta}^{\text{ols}}$, and $\hat{\gamma}^{\text{ols}}$ denote the least squares estimates,

$$(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}, \hat{\beta}^{\text{ols}}, \hat{\gamma}^{\text{ols}}) = \arg\min_{\tau,\alpha,\beta,\gamma} \sum_{i=1}^{N} \Big(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta - W_i \cdot (X_i - \overline{X})\gamma\Big)^2,$$

and let α^*, τ^*, β^*, and γ^* denote the corresponding population values:

$$(\alpha^*, \tau^*, \beta^*, \gamma^*) = \arg\min_{\alpha,\beta,\tau,\gamma} \mathbb{E}_{\text{SP}}\left[\Big(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta - W_i \cdot (X_i - \mu_X)\gamma\Big)^2\right].$$

Linear Regression With Covariates and Interactions

Results similar to Theorem 1 can be obtained for this case. The $\hat{\tau}^{\text{ols}}$ is consistent for τ_{SP}, and inference can be based on least squares methods.

Theorem

Suppose we conduct a CRE in a random sample from a super population. Then (i)

$$\tau^* = \tau_{\text{SP}},$$

and (ii)*,*

$$\sqrt{N} \cdot \left(\hat{\tau}^{\text{ols}} - \tau_{\text{SP}}\right) \xrightarrow{d} \mathcal{N}$$

$$\left(0, \frac{\mathbb{E}_{\text{SP}}\left[(W_i - p)^2 \cdot \left(Y_i^{\text{obs}} - \alpha^* - \tau_{\text{SP}} \cdot W_i - X_i\beta^* - W_i \cdot (X_i - \mu_X)\gamma^*\right)^2\right]}{p^2 \cdot (1-p)^2}\right).$$

Linear Regression With Covariates and Interactions

A slightly different interpretation of this result connects it to the imputation-based methods that are the topic of the next chapter.

Suppose we take the model at face value and assume that the regression function represents the conditional expectation:

$$\mathbb{E}_{\mathrm{SP}}\left[Y_i^{\mathrm{obs}} \middle| X_i = x, W_i = w \right] = \alpha + \tau \cdot t + \beta' x + w \cdot (x - \mu_X)\gamma. \tag{5}$$

In combination with the random assignment, this implies that
$\mathbb{E}_{\mathrm{SP}}\left[Y_i(0) | X_i = x \right] = \mathbb{E}_{\mathrm{SP}}\left[Y_i(0) | X_i = x, W_i = 0 \right] = \mathbb{E}_{\mathrm{SP}}\left[Y_i^{\mathrm{obs}} \middle| X_i = x, W_i = 0 \right] = \alpha + x\beta,$

and

$$\mathbb{E}_{\mathrm{SP}}\left[Y_i(1) | X_i = x \right] = \alpha + \tau + x\beta + (x - \mu_X)\gamma.$$

Linear Regression With Covariates and Interactions

Suppose that $W_i = 1$ for unit i, so $Y_i(1)$ is observed and $Y_i(0)$ is missing.

Under the model in (5), the predicted value for the missing potential outcome $Y_i(0)$ is

$$\hat{Y}_i(0) = \hat{\alpha}^{\text{ols}} + X_i\hat{\beta}^{\text{ols}},$$

so that for this treated unit the predicted value for the unit-level causal effect is

$$\hat{\tau}_i = Y_i(1) - \hat{Y}_i(0) = Y_i^{\text{obs}} - \left(\hat{\alpha}^{\text{ols}} + X_i\hat{\beta}^{\text{ols}}\right).$$

For a control unit i the predicted value for the missing potential outcome $Y_i(1)$ is

$$\hat{Y}_i(1) = \hat{\alpha}^{\text{ols}} + \hat{\tau}^{\text{ols}} + X_i\hat{\beta}^{\text{ols}} + (X_i - \overline{X})\hat{\gamma}^{\text{ols}},$$

Linear Regression With Covariates and Interactions

and the predicted value for the unit-level causal effect for this control unit i is

$$\hat{\tau}_i = \hat{Y}_i(1) - Y_i(0) = \hat{\alpha} + \hat{\tau} + X_i\hat{\beta} + (X_i - \overline{X})\hat{\gamma} - Y_i^{\text{obs}}.$$

Now we can estimate the overall average treatment effect τ_{FS} by averaging the estimates of the unit-level causal effects $\hat{\tau}_i$. Simple algebra shows that this leads to the OLS estimator:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i = \frac{1}{N}\sum_{i=1}^{N}\left\{W_i \cdot \left(Y_i(1) - \hat{Y}_i(0)\right) + (1 - W_i) \cdot \left(\hat{Y}_i(1) - Y_i(0)\right)\right\} = \hat{\tau}^{\text{ols}}.$$

Thus, the $\hat{\tau}^{\text{ols}}$ can be interpreted as averaging estimated unit-level causal effects in the sample, based on imputing the missing potential outcomes through a linear regression model.

Linear Regression With Covariates and Interactions

There is another important feature of the estimator based on linear regression with a full set of interactions.

As the above derivation shows, the estimator essentially imputes the missing potential outcomes. The regression model with a full set of interactions does so separately for the treated and control units.

When imputing the value of $Y_i(0)$ for the treated units, this procedure uses only the observed outcomes, Y_i^{obs}, for control units, without any dependence on observations on $Y_i(1)$ (and vice versa).

This gives the estimator attractive robustness properties, clearly separating imputation of control and treated outcomes. This will be important in the context of observational studies.

Transformations of the Outcome Variable

If one is interested in the average effect of the treatment on a transformation of the outcome, one can first transform the outcome, and then apply the methods discussed so far.

For example, in order to estimate the average effect on the logarithm of the outcome, we can first take logarithms and then estimate the regression function

$$\ln\left(Y_i^{\text{obs}}\right) = \alpha + \tau \cdot W_i + X_i\beta + \varepsilon_i.$$

Irrespective of the form of the association between outcomes and covariates, in a CRE, regression estimators of τ are consistent for $\mathbb{E}[\ln(Y_i(1)) - \ln(Y_i(0))]$.

Transformations of the Outcome Variable

There is an important issue, though, involving such transformations that relates to the correctness of the specification of the regression function.

Suppose one is interested in the average effect $\mathbb{E}[Y_i(1) - Y_i(0)]$, but suppose that one actually suspects that a model linear in logarithms provides a better fit to the distribution of Y_i^{obs} given X_i and W_i.

Estimating a model linear in logarithms and transforming the estimates back to an estimate of the average effect in levels requires assumptions beyond those on the conditional expectation of the logarithm of the potential outcomes: one needs to make distributional assumptions on the unobserved component (more on this in the next chapter).

Transformations of the Outcome Variable

As an extreme example, consider the case where the researcher is interested in the average effect of the treatment on a binary outcome.

Estimating a linear regression function by least squares will lead to a consistent estimator for the average treatment effect.
However, such a linear probability model is unlikely to provide an accurate approximation of the conditional expectation of the outcome given covariates and treatment indicator.

Logit models (where $\Pr(Y_i^{\text{obs}} = 1 | W_i = w, X_i = x)$ is modelled as $\exp(\alpha + \tau \cdot w + x\beta)/(1 + \exp(\alpha + \tau \cdot w + x\beta))$),
or probit models (where $\Pr(Y_i^{\text{obs}} = 1 | W_i = w, X_i = x) = \Phi(\alpha + \tau \cdot w + x\beta)$, with $\Phi(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} \exp(-z^2/2)$ the normal cumulative distribution function), are more likely to lead to an accurate approximation of the conditional expectation of the

Transformations of the Outcome Variable

As an extreme example, consider the case where the researcher is interested in the average effect of the treatment on a binary outcome.

Estimating a linear regression function by OLS will lead to a consistent estimator for the average treatment effect.

However, such a linear probability model is unlikely to provide an accurate approximation of the conditional expectation of the outcome.

Probit or logit models ($\Pr(Y_i^{\text{obs}} = 1 | W_i = w, X_i = x) = \Phi(\alpha + \tau \cdot w + x\beta)$ or $\Pr(Y_i^{\text{obs}} = 1 | W_i = w, X_i = x) = \exp(\alpha + \tau \cdot w + x\beta)/(1 + \exp(\alpha + \tau \cdot w + x\beta))$), are more likely to lead to an accurate approximation of the conditional expectation of the outcome given the covariates and the treatment indicator.

Transformations of the Outcome Variable

However, such a model will not generally lead to a consistent estimator for the average effect unless the model is correctly specified.

Moreover, the average treatment effect cannot be expressed directly in terms of the parameters of the logistic or probit regression model.

The issue is that in the regression approach, the specification of the statistical model is closely tied to the estimand of interest.

This separation is attractive for a number of reasons discussed in more detail in the next chapter, but it also carries a price, namely that consistency of the estimators will be tied more closely to the correct specification of the model. However, we do not view this as a major issue

Transformations of the Outcome Variable

In the setting of CRE. The bias is unlikely to be substantial with moderate sized samples, as flexible models are likely to have minimal bias.

Moreoever, this consistency property despite possible misspecification of the regression function only holds with completely randomized experiments. In observational studies, even regression models rely heavily on the correct specification for consistency of the estimator.

Furthermore, large sample results, such as consistency, are only guidelines for finite sample properties, and as such not always reliable.

The Limits on Increases in Precision due to Covariates

Suppose we do not include any predictor variables in the regression beyond the indicator variable for the treatment, W_i.

Normalized by the sample size, the sampling variance of the OLS estimator is equal to the difference in means estimator

$$N \cdot \mathbb{V}_{\text{nocov}} = \frac{\sigma_c^2}{1-p} + \frac{\sigma_t^2}{p},$$

Now suppose we have available a vector of covariates, X_i and that these covariates, their interactions with the treatment indicator, and possibly higher order moments of these covariates is included in the regression.

The Limits on Increases in Precision due to Covariates

The normalized sampling variance is the bounded from below by

$$N \cdot \mathbb{V}_{\text{bound}} = \frac{\mathbb{E}_{\text{SP}}[\sigma_c^2(X_i)]}{1-p} + \frac{\mathbb{E}_{\text{SP}}[\sigma_t^2(X_i)]}{p}.$$

The difference between the two expressions for the sampling variance

$$\begin{aligned}\mathbb{V}_{\text{nocov}} - \mathbb{V}_{\text{bound}} &= \left(\frac{\sigma_c^2}{1-p} + \frac{\sigma_t^2}{p}\right) - \left(\frac{\mathbb{E}_{\text{SP}}[\sigma_c^2(X_i)]}{1-p} + \frac{\mathbb{E}_{\text{SP}}\left[\sigma_t^2(X_i)\right]}{p}\right) \\ &= \frac{\mathbb{V}_{\text{SP}}(\mu_c(X_i))}{1-p} + \frac{\mathbb{V}_{\text{SP}}(\mu_t(X_i))}{p},\end{aligned}$$

where the last row is obtained from Equation (1), thus

The Limits on Increases in Precision due to Covariates

$$\sigma_c^2 = \mathbb{E}_{\mathrm{SP}}\left[\sigma_c^2(X_i)\right] + \mathbb{V}_{\mathrm{SP}}\left(\mu_c(X_i)\right), \qquad \text{and} \quad \sigma_t^2 = \mathbb{E}_{\mathrm{SP}}\left[\sigma_t^2(X_i)\right] + \mathbb{V}_{\mathrm{SP}}\left(\mu_t(X_i)\right).$$

The more the covariates help in explaining the potential outcomes, and thus the bigger the variation in $\mu_w(x)$, the bigger the gain from including them in the specification of the regression function.

When neither $\mu_c(x)$ nor $\mu_t(x)$ vary with the predictor variables, there is no gain from using the covariates. In small samples there will actually be a loss of precision due to the estimation of coefficients, that are, in fact, zero.

Testing for the Presence of Treatment Effects

In the current setting of CRE, tests for the presence of any treatment effects are not necessarily as attractive as the FEP calculations, but their extensions to observational studies are relevant.

In addition, we may be interested in testing hypotheses concerning the heterogeneity in the treatment effects that do not fit into the FEP framework because the associated null hypotheses are not sharp.

As in the discussion of estimation, we focus on procedures that are valid in large samples, irrespective of the correctness of the specification of the regression model.

Testing for the Presence of Treatment Effects

The most interesting setting is

$$Y_i^{\text{obs}} = \alpha + \tau_{\text{SP}} \cdot W_i + X_i\beta + W_i \cdot (X_i - \overline{X})\gamma + \varepsilon_i.$$

In that case we can test the null hypothesis of a zero average treatment effect by testing the null hypothesis that $\tau_{\text{SP}} = 0$.

However, we can construct a different test by focusing on the deviation of either $\hat{\tau}_{\text{SP}}$ or $\hat{\gamma}$ from zero.

If the regression model were correctly specified, that is

$$\mathbb{E}_{\text{SP}}\left[Y_i^{\text{obs}} \middle| X_i = x, W_i = w \right] = \alpha + \tau \cdot w + x\beta + w \cdot (x - \mu_X)\gamma',$$

Testing for the Presence of Treatment Effects

a test of $\tau = \gamma = 0$ would test the null hypothesis

$$H_0 : \mathbb{E}_{\mathrm{SP}}[Y_i(1) - Y_i(0)|X_i = x] = 0, \ \forall\ x,$$

against the alternative hypothesis

$$H_a : \mathbb{E}_{\mathrm{SP}}[Y_i(1) - Y_i(0)|X_i = x] \neq 0, \ \text{for some } x.$$

Without making the assumption that the regression model is correctly specified, it is still true that, if the null hypothesis that $\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] = 0$ for all x is correct, then the population values τ_{SP} and γ^* are both equal to zero.

Testing for the Presence of Treatment Effects

However, it is no longer true that for *all* deviations of this null hypothesis the limiting values of either τ_{SP} or γ^* differ from zero. It is possible that $\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$ differs from zero for some values of x even though τ_{SP} and γ^* are both equal to zero.

In order to implement these tests, one can again use standard least squares methods. The normalized covariance matrix of the vector $(\hat{\tau}^{\mathrm{ols}}, \hat{\gamma}^{\mathrm{ols}})$ is

$$\mathbb{V}_{\tau,\gamma} = \begin{pmatrix} \mathbb{V}_\tau & \mathbb{C}_{\tau,\gamma} \\ \mathbb{C}_{\tau,\gamma}^T & \mathbb{V}_\gamma \end{pmatrix}.$$

Testing for the Presence of Treatment Effects

The precise form of the components of the covariance matrix, as well as consistent estimators for these components, are given in the appendix.

In order to test the null hypothesis that the average effect of the treatment given the covariates is zero for all values of the covariates, we then use the quadratic form

$$Q_{\text{zero}} = \begin{pmatrix} \hat{\tau}^{\text{ols}} \\ \hat{\gamma}^{\text{ols}} \end{pmatrix}^T \hat{\mathbb{V}}_{\tau,\gamma}^{-1} \begin{pmatrix} \hat{\tau}^{\text{ols}} \\ \hat{\gamma}^{\text{ols}} \end{pmatrix}. \tag{6}$$

Note that this is not a test that fits into the FEP approach because it does not specify all missing potential outcomes under the null hypothesis.

Testing for the Presence of Treatment Effects

The second null hypothesis we consider is that the average treatment effect is constant as a function of the covariates:

$$H_0' : \mathbb{E}_{\mathrm{SP}}[Y_i(1) - Y_i(0)|X_i = x] = \tau_{\mathrm{SP}}, \qquad \text{for all } x,$$

against the alternative hypothesis

$$H_a' : \exists\, x_0, x_1, \text{ such that } \mathbb{E}_{\mathrm{SP}}[Y_i(1) - Y_i(0)|X_i = x_0] \neq \mathbb{E}_{\mathrm{SP}}[Y_i(1) - Y_i(0)|X_i = x_1].$$

This null hypothesis may be of some importance in practice. If there is evidence of heterogeneity in the effect of the treatment as a function of the covariates, one has to be more careful in extrapolating to different subpopulations.

Testing for the Presence of Treatment Effects

Lack of positive evidence for heterogeneity does not imply a constant treatment effect, but in cases with sufficient variation in the covariates, it does suggest that treatment effect heterogeneity may be a a second order problem why it may may be more credible to extrapolate estimates to different subpopulations.

In order to test this null hypothesis we can use the quadratic form

$$Q_{\text{const}} = (\hat{\gamma}^{\text{ols}})^T \hat{\mathbb{V}}_{\gamma}^{-1} \hat{\gamma}^{\text{ols}}. \tag{7}$$

Testing for the Presence of Treatment Effects

Theorem

Suppose we conduct a completely randomized experiment in a random sample from a large population. If $Y_i(1) - Y_i(0) = \tau$ *for some value* τ *and all units, then*
(i)*:* $\gamma^* = 0$,
and (ii)

$$Q_{\text{const}} \xrightarrow{d} \mathcal{X}(\dim(X_i)).$$

If $Y_i(1) - Y_i(0) = 0$ *for all units, then* (iii),

$$Q_{\text{zero}} \xrightarrow{d} \mathcal{X}(\dim(X_i) + 1).$$

Estimates for LRC-CPPT Cholesterol Data

We look at estimates for two average effects. The effects on

1. the cholesterol levels, the primary outcome of interest, denoted by `cholf`.
2. the level of compliance, `comp`

For each outcome, we present four regression estimates of the average effects.

1. only W.
2. include the composite prior cholesterol level, `cholp`, as a linear predictor.
3. include both prior cholesterol level measurements, `chol1` and `chol2`, as linear predictors.
4. add interactions of the `chol1` and `chol2` with W

Estimates for LRC-CPPT Cholesterol Data

Table 7.2: Regression Estimates for Average Treatment Effects for the PRC-CPPT Cholesterol Data from Table 7.1

Covariates	Effect of Assignment to Treatment On: Post Cholesterol Level est	(s.e.)	Compliance est	(s.e.)
No Cov	-26.22	(3.93)	-14.64	(3.51)
`cholp`	-25.01	(2.60)	-14.68	(3.51)
`chol1, chol2`	-25.02	(2.59)	-14.95	(3.50)
`chol1, chol2`, interacted with W	-25.04	(2.56)	-14.94	(3.49)

Estimates for LRC-CPPT Cholesterol Data

Compliance was far from perfect. On average, 75% and 60% in the control and treatment group took the nominal dose, respectively. Thus a difference of -15% (=60-70).

This means that the estimates of the effect on cholesterol levels, `cholf`, is estimates of *intention-to-treat* (ITT) effects, rather than estimates of the efficacy of the drug.

The consequence of the non-complience for the efficacy is studied in later chapters.

The left panel of Table 7.3 presents more detailed results for the regression.

The right panel of of Table 7.3 present the results using log transforming of `cholf`.

Estimates for LRC-CPPT Cholesterol Data

Table 7.3: Regression Estimates for Average Treatment Effects on Post Cholesterol Levels for the PRC-CPPT Cholesterol Data from Table 7.1

Covariates	Model for Levels est	Model for Levels (s.e.)	Model for Logs est	Model for Logs (s.e.)
Assignment	-25.04	(2.56)	-0.098	(0.010)
Intercept	-3.28	(12.05)	-0.133	(0.233)
`chol1`	0.98	(0.04)	-0.133	(0.233)
`chol2-chol1`	0.61	(0.08)	0.602	(0.073)
`chol1`× Assignment	-0.22	(0.09)	-0.154	(0.107)
(`chol2-chol1`)× Assignment	0.07	(0.14)	0.184	(0.159)
R-squared	0.63		0.57	

Estimates for LRC-CPPT Cholesterol Data

When using ln(`cholf`) this changes the estimand, and so the results are not directly comparable.

It is however useful to note, that in this case, the transformation does not improve the predictive power as shown by the R-squared.

In Table 7.4 we report p-values for some of the two tests: Q_{zero} (see equation (6) and Q_{const} (see equation 7) and the FEP on both outcomes.

Estimates for LRC-CPPT Cholesterol Data

Table 7.4: P-values for Tests for Constant and Zero Treatment Effects, Using `chol1` and `chol2-chol1` as Covariates for the PRC-CPPT Cholesterol Data from Table 7.1

		post chol level	compliance
zero treatment effect	$\mathcal{X}^2(3)$ approximation	0.000	0.000
	Fisher Exact P-value	0.000	0.001
constant treatment effect	$\mathcal{X}^2(2)$ approximation	0.029	0.270