

# Chapter 5 — Causal Inference for the Simple Linear Model

February 1, 2005

## 1 Introduction

Causal inference conventionally entails using data and a lot of auxiliary information to learn what might happen if there is an intervention in some social, biological, or physical process. For example, one might want to know what would happen to residential water consumption if the price of water were determined by its marginal cost rather than its average cost. Or one might want to know what would happen to a wild salmon population if a large number of hatchery salmon were released into the same river system. Or one might want to know how well first graders would learn to read if class size were reduced a certain amount.

## 2 Some Definitions: What’s a Causal Effect?

When regression analysis is used to make causal inferences about policy, there is often a model based on econometric traditions. Causal inferences are made from “causal parameters” that are part of that model. Heckman (1999:7-8) provides an excellent explanation of what this means.

Within the context of an economic model, the concept of a causal parameter is well defined. For example, in a model of production of output  $y$  based on inputs  $\mathbf{x}$  that can be independently varied, we write the function  $F : R^N \rightarrow R^1$  as

$$y = F(x_1, \dots, x_N), \tag{1}$$

where  $\mathbf{x} = (x_1, \dots, x_N)$  is a vector of inputs defined over the domain  $D(\mathbf{x} \in D)$ . They play the roles of causes, i.e. factors that produce  $y$ . These causes are the primitives of the relevant economic theory. Assuming that each input can be freely varied, so there are no functional restrictions connecting the components of  $\mathbf{x}$ , the change in  $y$  produced from the variation in  $x_j$  holding all other inputs constant is the causal effect of  $x_j$ . If  $F$  is differentiable in  $x_j$ , the marginal causal effect of  $x_j$  is

$$\frac{\partial y}{\partial x_j} = F_j(x_1, \dots, x_j, \dots, x_N) | \mathbf{x} = \tilde{\mathbf{x}}. \quad (2)$$

If  $F$  is not differentiable, finite changes replace the derivatives.

Several points are worth emphasizing now.

1. Each of the inputs can be manipulated.
2. These manipulations can be undertaken one at a time with the other inputs fixed at some particular values.
3. When an input is varied, the output varies in a manner described by a partial derivative (or a numerical approximation).
4. Causal effects are defined in the context of a model of how the world is supposed to work. The model should not just be a technical convenience.
5. Causal effects characterize what would happen *if* there were a particular “intervention.” They are hypothetical.
6. Causal effects are defined for a single observational unit.
7. Causal effects are comparative.

## 2.1 The Neyman-Rubin Model

Consider a *single* individual and a job training program. Let  $y$  be the response, the number of weeks after completing the program until a job is found. The observed intervention  $t$  equals 1 if that person participates in job training and 0 if not. Finally, and a source of potential confusion, let  $t^*$

equal 1 if the *hypothetical* intervention is job training and 0 if not. When one asks what *would* happen as a result of job training (compared with the lack thereof), one is asking about the hypothetical intervention that has not actually been introduced. It is the causal effects of this hypothetical intervention that we wish to estimate. We are seeking an answer to a “what if” question.

1. There are four possible pairings between the intervention that was received and the hypothetical intervention. The outcome, conditional on these four pairs, can be represented as follows:
  - (a)  $y|(t^* = 1, t = 1)$ : the outcome if hypothetically job training were received and it actually was received.
  - (b)  $y|(t^* = 1, t = 0)$ : the outcome if hypothetically job training were received but it was actually not received.
  - (c)  $y|(t^* = 0, t = 1)$ : the outcome if hypothetically job training were not received but it actually was received.
  - (d)  $y|(t^* = 0, t = 0)$ : the outcome if hypothetically job training were not received and it actually was not received.
2. Only the first and fourth conditional relationships are observable in principle. The second and third reflect “counterfactuals.” In practice, moreover, one will only be able to observe any given individual either for the first or last conditional relationship.
3. Within the Neyman-Rubin model, a causal effect can be defined as the difference between the outcomes under the treatment and control conditions:  $(y|t^* = 1) - (y|t^* = 0)$ .
4. Unfortunately, this definition cannot be implemented because the individual is either exposed to the treatment or not. But suppose we shift to groups rather than a single individual. And suppose we reformulate causal effects in distributional terms.
5. At the *group* level, a causal effect might be defined as a comparison between one or more features of  $(y|t^* = 1)$  and those same features of  $(y|t^* = 0)$ , where  $y$  is now a distribution of outcomes for a single group.

6. We concentrate on the special case of  $(\bar{y}|t^* = 1) - (\bar{y}|t^* = 0)$ . But the counterfactuals get us in trouble. For the “what if” all subject get the treatment we have:

$$\begin{aligned} (\bar{y}|t^* = 1) &= (\bar{y}|t^* = 1, t = 1) \times p(t = 1) + \\ &\quad (\bar{y}|t^* = 1, t = 0) \times p(t = 0). \end{aligned} \tag{3}$$

For the “what if” case when all are controls:

$$\begin{aligned} (\bar{y}|t^* = 0) &= (\bar{y}|t^* = 0, t = 1) \times p(t = 1) + \\ &\quad (\bar{y}|t^* = 0, t = 0) \times p(t = 0). \end{aligned} \tag{4}$$

7. The most straightforward way solve the counterfactual problem is to assign subjects randomly to treatments. It follows that a reasonable estimate of the average treatment effect is

$$(\bar{y}|t^* = 1, t = 1) - (\bar{y}|t^* = 0, t = 0). \tag{5}$$

8. The difference between the two means will be an unbiased estimate of the average treatment effect over individuals if the outcome for any one individual is unrelated to the treatment any other individual receives. In some circles, this is called the “stable unit treatment value assumption” (SUTVA), which can often be under the control of the researcher.
9. Note that we are no longer representing what is happening to particular subjects. For that you need a model.
10. Matching and covariance adjustments are then fallback positions. More on that soon.
11. The “what if” approach to causal effects can be easily generalized to comparisons between several interventions, including a “control” condition, and to interventions that are not categorical. That gets us back to a regression-like formulation developed by Freedman.

## 2.2 Thinking About Causal Effects as Response Schedules

1. Consider again a world in which nature constructs a population or generates the data directly using a linear model. So far, such models have been solely descriptions of how  $x$  and  $y$  are related. The mean of  $y$  is nothing more or less than a transformation of  $x$ . As such, the models are silent on what would happen to the mean of  $y$  if  $x$  were manipulated.
2. Imagine now a “response schedule” of the following form:

$$Y_{i,x} = a + bx + \delta_i, \tag{6}$$

where  $i$  refers to the observational units of interest,  $Y_{i,x}$  is the response,  $x$  is the “value” or “level” of some intervention, and  $\delta_i$  is a perturbation, commonly considered to be a random variable with an expected value of zero. We are now talking about random variables, which are theoretical entities with probability distributions. Hence, we work with expectations rather than means.

3. The expected value of  $Y_{i,x}$  is  $a + bx$ . Because the parameters are not subscripted, the same linear relationship applies to each observational unit.
4. A critical point is that the “disturbance” or “error,”  $\delta_i$ , does not vary with  $x$ . By some chance process that depends on context, a value of  $\delta_i$  is chosen and affixed permanently to observational unit  $i$ . Then as different values of  $x$  are introduced, systematic variation in the value of  $Y_{i,x}$  follows. The value of  $\delta_i$  is the same no matter which value for  $x$  is selected.
5. Finally, the response schedule requires that  $a$  and  $b$  are the same no matter what value of  $x$  or  $\delta_i$  applies to other relevant precincts.
6. Freedman provides a second kind of response schedule that does not require that errors  $\delta_i$  be invariant over interventions for a given observational unit. Let

$$Y_{i,x} = a + bx + \delta_{i,x}, \tag{7}$$

where the  $\delta_{i,x}$ , with an expected value of zero, is drawn independently and at random from some invariant distribution, given a value for  $x$ .

7. With  $x$  set to some value, the errors in principle generate a distribution for the values of  $Y_{i,x}$ . Thus, we have a response schedule in which, once the value of  $x$  is determined, the particular value  $Y_{i,x}$  is, in effect, sampled at random.
8. The linear form and  $a$  and  $b$  are still unaffected by  $x$ , the random perturbation, and by whatever may be going on for other units.
9. The main difference between the two models is that in the first, once a random error is attached to a case, it does not change. With invariant errors, the intervention is selected independently of  $\delta_i$ , and one gets to see  $(X_i, Y_i)$ , where  $Y_{i,x} = a + bX_i + \delta_i$ . With random errors,  $X_i$  is determined and then, conditional on that value,  $Y_i$  is chosen at random from a fixed distribution with a mean of  $a + bX_i$ .
10. The first model leads naturally to a causal interpretation for the role of  $x$ . Once  $\delta_i$  is fixed, variation in the response results only from variation in  $x$ .
11. In the second model, because both  $x$  and  $\delta_i$  can vary, both are potential culprits and isolating the role of  $x$  is more difficult. For the first model, a causal interpretation requires that  $x$  is selected independently of the single error attached to observational unit  $i$ . For the second model, a causal interpretation requires that the error is selected independently of  $x$ .
12. The second model is likely to be a bit more realistic precisely because the perturbations are free to vary. The second model also comports well with a regression formulation and is the conception we shall use. Then our causal inferences are to the conditional distribution of the response: *What would the conditional distribution of the response be if the value of the predictor were set to  $x$ ?*

### 2.3 What's an Intervention?

1. It is common to require that  $x$  is “manipulable.”
2. Matters get murky if one allows for interventions that depend on technology and/or political systems that do not exist. Once one opens the

door to interventions that do not appear to be fully realistic, almost anything would seem to qualify as an intervention.

3. Perhaps most commonly, race and sex are taken to be interventions, sometimes cloaked in “but for” legal language. But if Jane is believed to be making \$500 less a month than she would have made had she been male, the policy maker does not really have the direct option of making Jane male. The only real option is to compensate for the apparent inequality through policy instruments that *are* manipulable.
4. There are times when  $x$  can be manipulated but it cannot be manipulated without at the same time manipulating confounders. More on that soon.
5. Sometimes interventions are also not reversible.
6. Opening the door to natural phenomena as interventions brings into question the centrality of human action in any definition of a causal effect. the key is, can a plausible response schedule be constructed?
7. Consider an extended example of Amendments in 1990 to The Clean Air Act that mandated significant changes in vehicle inspection and maintenance programs for cities with air-quality problems. For light-duty vehicles of a certain age, there was to be scheduled testing of tail pipe and evaporative emissions.
8. There are other definitions of causal effects in the statistics literature. “Granger cause” is one example. But these definitions actually address rather different issues.
9. there will sometimes be ways to reformulate the intervention so that causal effects make sense. In the case of race, for instance, one can manipulate information about race if not race itself.

### 3 Studying Cause and Effect With Data

You **must** be able to construct a plausible response schedule as a necessarily condition to estimate causal effects from data. Otherwise, your “causal inferences” are nonsense. But even then, the going is not easy.

### **3.1 Using Nonstatistical Solutions for Making Causal Inferences**

Two kinds of generic approaches are common for what are nonstatistical efforts to make causal inferences. One can examine the same observational unit under different interventions or examine different units, each under a separate intervention.

### **3.2 Using Statistical Solutions for Making Causal Inferences**

As noted earlier, there are solutions available based on aggregating observational units. Comparability is established between *groups* of observational units.

1. The “gold standard” for comparability is achieved through random assignment to treatment groups. It is worth reiterating that one of the very desirable features of randomized designs is that there is no need to depend on a full response schedule framework; there need be no complete model for how a single observational units responds. If there is a desire to further model the response as a function of the treatment (e.g., impose an additive regression-like model), a response schedule framework can be applied.
2. Matching is another method commonly used to achieve comparability.
3. A third method commonly applied to achieve comparability is through “covariance adjustments” that are built into regression with more than one predictor.

### **3.3 Using the Simple Linear Model for Making Causal Inferences**

Suppose, for the moment, that the response schedule formulation is useful and reasonably plausible and that the invention is binary. Code the treatment condition as a “1” and the control coded as a “0.”

1. Consider Figure 1 in which, for each of 50 fictitious observations in a given city (some before and some after), the response of average carbon

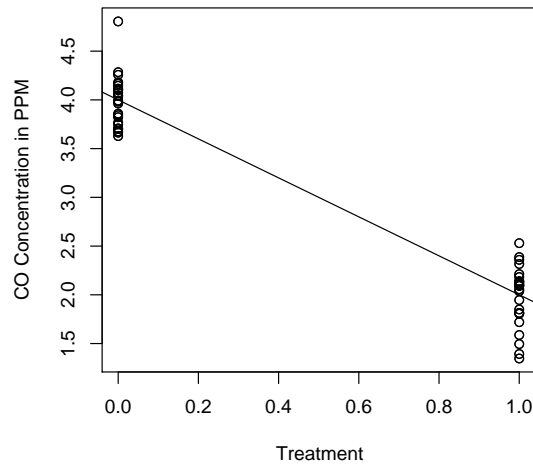


Figure 1: Regression of Carbon Dioxide Concentrations on Vehicle Inspection Intervention

monoxide (CO) concentrations in parts per million (PPM) is plotted against the binary variable for treatment and control conditions.

2. With only two values for the treatment, the regression line goes through both conditional means for CO concentrations. The value of  $\hat{\eta}_0$  is the mean of the response before the intervention (4 PPM). The value of  $\hat{\eta}_1$  is the difference between the mean of the before and after response ( $-2$  PPM). The sum of  $\hat{\eta}_0$  and  $\hat{\eta}_1$  is, therefore, the mean of the response after the intervention (2 PPM).
3. Any other two values could have been used instead of 0 and 1. Then, the fit would have been the same, and it would have been possible to reconstruct the two conditional means. But the 1-0 coding makes the calculations easy. And all of the features of the simple linear model carry over. More on categorical predictors later.