Harvard Business School

9-894-032
February 22, 1994

Causal Inference

In *Forecasting with Regression Analysis (894-007)* we saw how you could use regression analysis to forecast a future value of a dependent variable. In this note we take up a second use of regression: causal inference. We use regression to infer by how much a deliberate intervention that changes the value of some independent variable will **cause** the value of the dependent variable to change. For example, we may be interested in the amount by which our stock price will change if we increase our dividend by $1, or by how much Gross Domestic Product will change if the Fed lowers interest rates by 1%, or by how much an increase of $1 million in advertising expenditures will cause sales to change. Sometimes you can infer these causal effects by examining past data. When such data are "observational" (not obtained from carefully controlled experiments), you can use the regression model to estimate those effects.

Many textbooks in statistics assert that you cannot infer causation from observational data; rather, you can only infer statistical association. They state that you might observe in a representative sample of data that whenever x increases by one unit, y increases by three units, on average, but from that you cannot conclude that the expected value of y will necessarily increase by three units (or at all) if you deliberately cause x to increase by one unit.

Contrast this with what managers, scientists, and economists do. A manager of a fast-food chain might observe that those of the chain's restaurants that are located near a McDonald's restaurant do better than those that are not near a McDonald's, and as a result will seek new restaurant sites preferentially that are near McDonald's. A doctor may observe that patients who smoke incur heart and lung disease at a higher rate than nonsmokers, and therefore advise her patients not to smoke. An economist may observe that high inflation is associated with low unemployment and vice versa, and recommend that to control inflation the Government should take steps to increase unemployment. These are all violations of the textbook rule.

In this note, we shall show how you can sometimes infer from observational data how changing the value of one variable causes the value of some other variable to change. You will see that although you cannot "prove" causation, you may be able to make reasonable causal inferences. But you can easily make incorrect inferences unless you are very careful.

What Is Causation?

An assertion that reducing the price of a product from $18 per unit to $17 **causes** demand to increase from 22,000 units per month to 23,000 makes sense only in terms of an ideal experiment—one that can never be performed in practice. The experiment involves two scenarios. In the first, price is set at $18. The second scenario is identical to the first in all respects except that the price is set at

Professor Arthur Schleifer, Jr. prepared this note as a basis for class discussion.

$17. Under both scenarios, the monthly demand is observed. If demand was 22,000 units per month when price was $18, and 23,000 when price was $17, we can conclude that the price change **caused** the change in demand.

"Identical in all respects…" means just that. The two scenarios cannot take place in different periods of time, or in different geographical regions. Although it is common practice to assert that an increase in demand that followed a reduction in price was "caused" by the price reduction, such an assertion involves measurement of demand in two different periods, between which other factors that affect demand may have changed.

Given that the "ideal" experiment can never be performed in practice, we can never obtain an exact measure of the amount by which a change in the value of one variable causes the value of another variable to change. Our challenge is to find methods that come as close as possible to mimicking the ideal experiment. But before discussing such methods, let's explore what we would learn if we could actually carry out this two-scenario experiment.

The price reduction—the variable whose value we deliberately change—is called a "treatment" or an "intervention." What is the effect of that treatment? Although we have focused on just one effect that was of particular interest to us—the change in demand—it should be clear that the price change causes not just this one effect, but many effects. An increase in demand of 1,000 units next month means that some customers who would not have purchased had the price not been changed decided to purchase as a result of the intervention. This, in turn, may mean that some of them will not purchase some other product, that some will reduce their savings or increase their debt. Perhaps a competitor will respond to our price reduction by reducing his price as well (something he would not have done had we not lowered our price), and this too might have an impact on the demand for our product.

Some of those effects may have little to do with our "bottom line," but others may. If lowering the price of one item in our product line diverts demand away from other items, then the total effect of the intervention is the increase in demand for the item whose price was reduced, less the decrease in demand for substitute products, more appropriately measured in dollars than in units. A single intervention usually can change the values of many variables, but one of them—the net change in dollar sales across all items in our product line—is the one we select to be the dependent variable, the one that most appropriately measures the total effect of the intervention.

What Is an Experiment?

The nearest we can come in the real world to measuring the true effect of a treatment is to conduct an experiment by finding "matched pairs"—pairs of individuals (or experimental units) that are as alike as possible in all respects—and to apply the treatment to one member of the pair and not to the other. If the pairs are truly matched in all respects, we will have essentially done with matched pairs what the "ideal experiment" does with a single individual or experimental unit: the difference in their responses will be a true measure of the effect. Unfortunately, from a practical point of view, it is not possible to find individuals who are exactly alike in all respects. (Even identical twins have almost certainly been exposed to different environmental influences.) Thus "matched pairs" may be alike in many respects, but they may differ with respect to other, unmeasured variables, that have an effect on the dependent variable. If these unmeasured variables happen to be correlated with the treatment, the observed treatment effect will include a **proxy effect** for these other variables.

Even the choice of which experimental unit gets the treatment and which does not may make it difficult to sort out effects. For example, the average effect on longevity of giving up smoking (the "treatment") may be different for people who voluntarily give it up than for people who are forced to give it up. In the extreme, we could imagine a situation in which those who voluntarily give up smoking are the only ones whose longevity is increased. If this were true, we would observe that

those who gave up smoking lived longer than those who did not, but we would also discover that applying coercion or providing incentives to nonvolunteers to give up smoking would provide no benefits to them.

Unwanted proxy effects of unmeasured variables can be eliminated by using a random device to choose which experimental unit in a matched pair receives the treatment. Random assignment of treatments assures that, on average, whatever unmeasured variables affect the dependent variable will not be correlated with the treatment. Thus, the treatment will not capture unwanted proxy effects.

There are situations where this random assignment can be carried out relatively easily. Returning to our price-reduction problem, suppose the context is that of a direct-mail company. The company could prepare two sets of catalogs, both identical except for the item whose price reduction was under consideration. One set of catalogs would show the standard price; the other set, the reduced price. Matched pairs of customers could be selected based on the recency, frequency, and monetary value of their previous purchases, and for each pair the catalog with the reduced price could be assigned at random. In drug testing, it is routine to assign the drug to one member of a matched pair, and a placebo to the other, with the determination of who gets what decided by a randomizing device (e.g., the flip of a coin).

There are other situations where random assignment is virtually impossible, either because it is too hard to implement, or is socially unacceptable. In the smoking experiment, it would be impossible to justify and enforce a policy in which some people, chosen at random, were instructed to keep smoking, while others were told to stop smoking. In dealing with the economy, we cannot segment the population into two groups, on only one of which we take measures to increase unemployment, and observe the difference in the rate of inflation in the two segments. Even in the pricing example, it might be unacceptable to the mail-order company's management to have two catalogs in circulation with different prices. We are often reduced to relying on observational data.

Observational Data

When we seek to estimate from observational data the effect of a "treatment"—an independent variable whose value we will be able to manipulate in the future—on a dependent variable, the estimation problem is made more difficult by the presence of other independent variables that (1) may also affect the dependent variable, and (2) may be correlated with the treatment variable. An independent variable may be correlated with the treatment for one of four reasons: (a) there may be no causal relationship between the two variables, but they might be correlated by chance alone; (b) the independent variable may **affect** the treatment; (c) it may be **affected by** the treatment; or (d) it and the treatment may both be affected by some other variable: they may be correlated due to a "common cause."

An Example

We may, for example, be interested in learning by how much changes in the posted speed limit on highways (the treatment) affect motor-vehicle death rates—deaths per thousand drivers per year (the dependent variable). The reason for our interest is that if we discover that reducing the speed limit reduces death rates, we might want to propose legislation to lower the speed limit.

Suppose we have a cross section of the fifty states in the United States, with data on each state's maximum speed limit and motor-vehicle death rate. Even if lowering the speed limit really reduced the death rate, a scatter diagram of speed limit vs. death rate might show that, in the data, death rate declines as speed limit increases. How could this be? It might be that states with very high death rates are states where bad weather makes driving conditions hazardous, where drivers drive long distances, where driving under the influence of alcohol is prevalent, etc. These states may have

lowered the speed limit to reduce the carnage, but still have higher death rates than states in which driving is safer, but the speed limit remains high. If this story is correct, then low speed limits may reduce the death rate but also proxy for variables that increase the death rate.

If those other variables—weather conditions, miles driven per capita per year, alcohol consumption, etc.—are included, along with speed limit, as independent variables in a regression model, then the regression coefficient on speed limit will show how death rate varies with speed limit when the other variables in the model are held constant. Speed limit will no longer proxy for these other variables. If lower speed limits reduce the death rate, then the regression coefficient on speed limit will be negative. Weather conditions, miles driven, and alcohol consumption are examples of other independent variables that **affect** death rate and that are **correlated** with speed limit. The correlation occurs because these variables have **caused** the speed limit to be lowered in states where they are major contributors to highway deaths. These variables **should** be included in the model to eliminate their unwanted proxy effects on the treatment variable.

Suppose we discover that states whose citizens have pronounced concerns for public safety tend to have both low speed limits and rigorous automobile-inspection standards. This is a case where the treatment and another variable (automobile inspection) that may affect death rates are correlated because they are both affected by another variable that is a **common cause**—concern for public safety. Clearly, a measure of the rigor of inspection standards should be included as an independent variable; otherwise, speed limit will capture the unwanted proxy effect of inspection on death rate.[1]

Now suppose that reduction in the posted speed limit causes drivers to drive slower, on average, and it is the actual reduction in speed driven, not the posted speed limit, that causes the death rate to decrease. Should we include average speed driven as an independent variable in our model? Clearly not: if we did, the regression coefficient on posted speed limit would show its relationship to death rate when average speed driven remained constant, and since actual speed driven, not posted speed limit, causes fatal accidents, the coefficient would indicate that the effect of posted speed limit on death rate was zero, and thus make it appear that changing posted speed has no effect on death rate. To assure that the regression coefficient correctly captures the causal relationship, we want posted speed limit to **include** the "good" proxy effect of driving speed, and thus we want to **exclude** driving speed from the regression model. Driving speed is an example of a variable that affects the dependent variable but is **affected by** the treatment variable. Such a variable should be excluded from the model, so that the treatment variable will capture its proxy effects.

Finally, consider the case where some other variable, say the average age of cars in the various states, affects death rates, but has no causal relationship to posted speed limits. Nonetheless, average age may be correlated with speed limits in the sample data: even variables that have nothing to do with one another are seldom perfectly uncorrelated in observational data. In this case, failure to include average age of cars as an independent variable in the model will cause speed limit to carry an unwanted proxy effect for age of cars. We should, therefore, **include** age of cars as an independent variable.

Which Independent Variables Should Be Included?

Based on this example, we can state the following rules. When you want to estimate the effect that a treatment or intervention will have on a dependent variable, you should

---

[1] If inspection and speed limit are perfectly correlated, it will be impossible to sort out to what degree each contributes to lowering the death rate. This is an example of **collinearity,** a problem that occurs frequently in regression. Collinearity is not a defect in the methodology; it is a defect in the data. The only way you can sort out the effects in this extreme case is to change the relationship between speed limit and inspection in the data.

- include as an independent variable any variable that you believe might affect the dependent variable and that is correlated with the treatment variable because (a) it affects the treatment variable, or (b) it and the treatment variable are both affected by a common cause, or (c) the correlation occurred purely by chance;

- exclude from the model any variable that you believe might affect the dependent variable and that is correlated with the treatment variable because it is affected by the treatment variable.

If a variable affects the dependent variable but is uncorrelated with the treatment variable, whether you include it or not makes no difference as regards the regression coefficient for the treatment variable: there are no proxy effects from an uncorrelated variable. As a matter of practice, you should include it if it affects the treatment variable; at the very least, it will improve the fit of the model, and in a sample of observational data only rarely are two variables completely uncorrelated.

The consequences of these rules may seem counterintuitive. A variable that affects the dependent variable and is correlated with the treatment variable must be included in the model; the higher the correlation, the more important it is to include it. Including such a variable will not greatly improve the fit (increase R^2, decrease RSD). Nevertheless, omitting a variable like this distorts the apparent effect of the treatment variable by causing it to pick up the unwanted proxy effects of the omitted variable.

On the other hand, omitting an independent variable that affects the dependent variable and is affected by the treatment variable assures that the treatment variable captures its "good" proxy effects. Nevertheless, omitting such a variable invariably results in a poorer fit (lower R^2, higher RSD).

In both cases, what is good for proper causal inference is bad for forecasting. This seemingly counterintuitive result is resolved by recognizing that causal inference involves correctly estimating a particular regression coefficient, while forecasting involves providing a good fit to past data. What is good practice for dealing with one of these problems is not necessarily good practice for dealing with the other.

How to Identify the Relevant Independent Variables

Given that you should include any independent variable that might affect the dependent variable and that is correlated with the treatment variable, but is not affected by it, how do you decide what variables to include? The answer depends on your understanding of what causes what, and on your ingenuity in finding ways of measuring crucial variables. Sometimes you have to settle for a variable that is correlated with such a crucial variable. In our speed-limit example, you may have trouble in obtaining data on drunk-driving convictions in a state, but probably can easily get statistics on alcohol consumption. For many reasons this may be an imperfect measure of driving under the influence, but it may be good enough for our purposes. Think about how you would measure weather conditions that are dangerous for drivers, or amount of driving per person.

A useful tool for depicting causal relationships is the **influence diagram**. *Figure A* shows such a diagram schematically. A treatment variable and a dependent variable are shown. Other variables are classified by type: Type A variables affect the dependent variable directly as well as indirectly through their effect on the treatment variable. Type B variables affect the dependent variable and are correlated with the treatment variable by virtue of a common cause. Type C variables affect the dependent variable and are correlated with the treatment variable by chance. Type D variables affect the dependent variable but are uncorrelated with the treatment variable. All of these variables should be included as independent variables in the regression model.

Type E variables, on the other hand, affect the dependent variable directly, but are affected in turn by the treatment variable. They should not be included in the model.

It is not always clear which way the causation goes. Advertising expenditures by your competitor may be correlated with your advertising. The correlation may be due to a common cause (seasonality, business conditions), in which case your competitor's advertising should be included as an independent variable. On the other hand, your competitor may simply be reacting to your advertising, raising expenditures when you raise yours, and vice versa, in which case it should be excluded. About all you can do under such circumstances is perform the regression with and without the competitive-advertising variable, and weight the resulting regression coefficients on the treatment variable by the probability you assign to the two competing causal models.

Figure 1 Influence Diagram