WHAT IS ENDOGENEITY BIAS AND HOW CAN WE ADDRESS IT?

Professor Victor Menaldo
University of Washington, Political Science
January 2011

SELECTION BIAS AND ENDOGENEITY

Suppose you want to explain what determines the observed variation in an outcome of interest. For example, what explains differences in economic development between nations? Say your hypothesis for explaining variation in economic development across countries is the quality of a country's institutions. Countries with broad and secure property rights, and the rule of law, generate consistent economic growth. Those with weak property rights enforcement and rife with corruption and cronyism, do not. If you were to try to establish whether there is a causal relationship running from a country's institutions to its level of economic development, how would you go about it? Would you attempt to establish whether there is a correlation between these two variables?

There are good reasons to eschew this approach. Correlation does not equal causation. Higher levels of economic development may drive higher levels of institutional quality. Alternatively, an unobserved variable may jointly determine both high levels of institutional quality and high levels of economic development. Or both might be true. But the fundamental reason why a correlation between institutional quality and the level of Per Capita Income may not allow you to conclude this correlation is causal, and running from institutional quality to economic development, is that observational data is usually not randomly assigned. This is certainly the case when it comes to institutions.

If institutions were randomly assigned, establishing causality would be as easy as ascertaining whether institutions and economic development are correlated. Does variation in the independent variable map onto variation in the dependent variable? If they were in fact correlated, one could safely infer that institutions are indeed a cause of persistent economic growth. The reason is that random assignment ensures that the units in the treatment group, those exposed to high quality institutions, will be matched with an equal number of similar units in the control group, save for the fact that the latter do not receive exposure to the treatment.

How are the units in the treatment group similar to those in the control group? They are similar in that their expected value of the dependent variable is statistically identical BEFORE exposure to the treatment variable. In other words, if one were to calculate the average value of Per Capita Income for countries slated to be exposed to high quality institutions and the average value of Per Capita Income for countries that will not be exposed to high quality institutions, one would arrive at the same answer. One way to think about this is that for every unit in the treatment variable there is likely to be a similar, if not identical, unit in the control group—a twin, if you will.

What is the ultimate consequence of the similarity between treatment and control groups for causal inference? Any baseline differences in the dependent variable of interest between the treatment and control groups BEFORE exposure to the treatment will have been eliminated. The upshot is that any differences in the mean level of the dependent variable of interest between treatment and control groups AFTER exposure to treatment can be accurately attributed to the treatment variable. This eliminates the possibility that an unobserved factor correlated with the units in the treatment group, and also correlated ex ante with a higher level of the dependent variable, is driving any difference in the average outcome between both groups observed AFTER exposure to treatment.[1] In other words, it eliminates the problem that unobserved confounders might be driving any correlation between independent and dependent variables.

To bring this back to institutional quality and economic development, there are both observed and unobserved processes that lead to the adoption and perpetuation of institutions across countries, and these factors are correlated with economic development, either directly or indirectly. These factors therefore have to be neutralized to avoid inducing a biased calculation of the treatment effect of institutions on growth. Otherwise, they will engender a difference in the baseline measures of the outcome of interest between the control and treatment group before exposure to treatment. Therefore, any difference in the control and treatment groups after exposure to treatment has to be adjusted to account for these preexisting differences.

Another example may help. Suppose that we wanted to discern the treatment effect of aspirin on headache pain. Non-random selection into the treatment group would be akin to having a mild headache and receiving an aspirin versus having a strong headache and receiving no aspirin and then using the level of headache pain afterward to vet the effectiveness of aspirin. This will bias upward the (purported) positive effect of aspirin and make it appear like aspirin is really great for headaches. However, this conclusion will be an artifact of selection bias.[2]

To summarize, the key idea here is that correlation between the independent variable and other variables that are correlated with the outcome of interest render selection into the "treatment group" non-random; instead, assignment to the treatment group will have been a function of some other factor and, more importantly, that other factor will be correlated with a higher (or lower) level of the outcome

---

[1] The Central Limit Theorem underpins this notion: if the sample is composed of data that is normally distributed and as the sample size approaches infinity, you will be more likely to match the units assigned to the treatment group with units assigned to the control group across any dimension of variation that might be correlated with their observed levels regarding the outcome of interest. And even in the event that you cannot eliminate baseline differences between the treatment and control groups before exposure to treatment via random assignment, you can nonetheless calculate an unbiased average treatment effect if you can calculate the difference-in-the-differences: the difference in the average outcome between the treatment and control groups after calculating the difference in each group's outcome over time: the difference between the post and pre experiment phase for each group.

[2] Because the double blind studies that have established the efficacy of aspirin vis-à-vis headache pain employed random assignment to the treatment group, I have no doubt that aspirin is actually quite good for headaches.

of interest BEFORE the treatment variable is even assigned. This makes it more (or less) likely to erroneously attribute a causal effect to the treatment variable when comparing the difference between treatment and control groups AFTER assignment.

GAINING TRACTION ON THE PROBLEM

One way of addressing the potential for endogeneity bias is to use instrumental variables. This is an alternative to attempting to identify and control for all possible factors that might be correlated with both the "treatment" variable of interest and the outcome of interest when random assignment was not used to allocate units to the treatment group. Instead, the logic of an instrumental variable is that it is not correlated with these alternative factors whatsoever. In fact, it must only be correlated with the independent variable of interest to qualify as an instrumental variable. Because it is only correlated with the independent variable of interest and not any other variable, such a variable will only be correlated with the dependent variable of interest indirectly: the instrumental variable works exclusively through the independent variable to affect the dependent variable. This is called the "exclusion restriction".

In political economy, instrumental variables often exploit "quasi-natural experiments". These are "situations where the forces of nature or government policy have conspired to produce an environment somewhat akin to a randomized experiment" Angrist and Krueger 2001, p. 73). The goal is to identify sources of variation that could not have possibly been determined by the outcome of interest, nor are correlated with other factors that might affect the outcome of interest other than the independent variable of interest.

Acemoglu, Johnson and Robinson (2001)

Again, suppose that you hypothesize that broad and secure property rights and the rule of law cause economic development. A friendly skeptic may say to you: but doesn't the causal arrow run from economic development to good institutions, since wealthy countries can afford good institutions and poor countries cannot? One way around this is to use an instrumental variable that captures the exogenous variation in institutions. If there is a high likelihood that modern day institutions were induced by a source of exogenous variation, and that these exogenous factors are exclusively working through these modern day institutions to affect growth today, then you've got a good instrument.

Acemoglu, Johnson and Robinson (2001) employ a clever strategy to show that the causal arrow runs from a country's political institutions to its economic development, rather than vice-versa. They argue that settler mortality rates during colonialism, a function of disease environments that are exogenous—not caused by human institutions or behavior—influenced colonizers' colonization strategies and, in turn, colonization strategies are linked to a path of institutional development that culminated in the modern day institutions that are observed today, even after colonialism ended. Finally, contemporary institutions are the only path by which colonies' erstwhile settler mortality rates affect economic development today. Specifically, in the American and African colonies in which settler mortality rates were high, the migration of Europeans was kept in check—either because Europeans moved over and

died or were deterred from migrating because of the fear of death. This encouraged the few Europeans who were rich and brave enough to make the trip over to impose predatory institution in which natives were forced into servitude or slaves were brought over from Africa to work in large plantations and mines. Conversely, in the colonies in which settler mortality rates were low, Europeans migrated in droves, seeking a better life. And the critical mass of migration from the European mainland made it easier for colonizers to establish institutions to establish institutions that codified political institutions that protected property rights broadly. While in both the colonies established on the principle of predation against natives and imported slaves and those established on the principle of equal rights for European colonizers early forms of institutions persisted until today, it was only in the latter case that these institutions provided the incentives conducive to long-run economic growth. This therefore explains why places with high settler mortality rates and low quality contemporary institutions are associated with low levels of economic development and places with low settler mortality rates and high quality contemporary institutions are associated with high levels of economic development.

SOME PITFALLS

1. Not all policy interventions are exogenous. Political decisions and past realizations of the outcome of interest can affect existing policies or the decision to introduce a new policy. For example, in the AJR story outlined above, what if levels of wealth yesterday drove settlers' colonial strategies and those levels of wealth are correlated with the level of wealth today? This is one of the reasons why AJR later take pains to show that there has been a "reversal of fortune" (AJR 2002): they show that the New World's wealthiest regions, both before and during colonialism—for example Mexico and Peru—were eventually overtaken by some of the new world's poorest regions (for example, the United States).

2. The effect that is registered and reported in the study may be conditional on some unobserved factor uniquely associated with the quasi-natural experiment and therefore any conclusions gleaned about this study cannot be extrapolated beyond that context. This issue of heterogeneous treatment effects differs fundamentally from the issue of baseline differences between a treatment and control group. As outlined above, to eliminate bias the mean level of the dependent variable should be identical across groups before the experiment is run—before units are assigned to the treatment variable versus the control group. However, even if there is no baseline difference between groups, the group that receives exposure to the treatment might respond systematically differently than a treatment group in another context. Again, this is the case *even if random assignment determines selection into the treatment group and therefore eliminates baseline differences or, similarly, an instrumental variable isolates the exogenous variation in the independent variable of interest*. This is because the direction and magnitude of the *response* to the treatment variable might be conditional on attributes possessed by the units in the treatment group. Or it may be something about the setting in which the first experiment was conducted. The methods to address this other source of bias are beyond the scope of this note.

3. You may have weak instruments only weakly correlated with the explanatory variable that you fear is contaminated by endogeneity bias. This issue and its solutions are beyond the scope of this note.