

Jesper B Sørensen

August 2012

Endogeneity is a fancy word for a simple problem. So fancy, in fact, that the Microsoft Word spell-checker does not recognize it.

Technically, in a statistical model you have an endogeneity problem when there is a correlation between your X variable and the error term in your model.

What does this mean? Well remember that the error term in your model is due to all of the stuff in your dependent variable that is not due to the variables you have in your model.

So in the broadest sense an endogeneity problem arises when there is something that is related to your Y variable that is also related to your X variable, and you do not have that something in your model. Call that something Z, although notice that I have not claimed that Z is a variable – it is just a “something.”

What this broad meaning of endogeneity suggests is that there are a wide variety of sources of endogeneity problems. In fact one of the problems with the use of the term is that people use it to cover many different things, yet the solutions to those different things can be quite different.

For example, endogeneity in this broad sense can be caused by omitted variables, or unobserved heterogeneity. In this case, the endogeneity complaint is a complaint that you left a variable (or two) out of your model.

This is obviously very familiar to everyone in this room. One reason why it is familiar is that we all know how to deal with it: measure the variable and put it in. And we all know how to fight with the reviewer about such things.

Unfortunately, there are other sources of endogeneity that are not so easily dealt with. And in fact, I think that in most cases where the charge of endogeneity is filed, people are not so much worried about omitted variables. Rather, what they are worried about is things like simultaneity – i.e., X causes Y but Y also causes X, -- and self-selection. The problem with such endogeneity problems is that no amount of control variables will address them.

For an example of simultaneity, consider a very nice paper by Simcoe and Waguespack on status signals. Sociologists are somewhat obsessed with the idea that rewards accrue to actors because of their status, and claim that status affects the performance of those actors – i.e., quality. This is the Matthew Effect – because people defer to high-status actors and wish to affiliate them, these actors reap higher rewards and get more recognition. Yet the problem is that quality also affects status – people get recognized because they do good work.

But if all there is to the status effect is the effect of accumulated quality, then there really is not a whole lot for sociologists to talk about, at least with respect to status. So trying to see whether the status signal has an independent effect is very important – but controlling for variables (even measures of quality) is not good enough.

Simcoe and Waguespack take advantage of a situation in which the identity of authors is sometimes obscured. The reason why the identity is obscured in some cases and not others is unrelated to the quality of the author's work. So if you can detect a difference in attention when the identity is obscured and when it is not, you can plausibly attribute that to an effect of status.

Self-selection is another source of endogeneity. Firm size and entrepreneurship...

I think we are starting to see in organizations and management research an increased concern with these kinds of endogeneity issues. If you think about the different kinds of endogeneity concerns, what you see is an increasing concern with more complex forms of endogeneity – everyone takes it for granted that we should be worried about omitted variables, a fair number of people think seriously about self-selection problems, and we are seeing the beginnings of an emphasis in management research on worrying about simultaneity problems.

I think this is a good thing. But we have to be careful not to get too obsessed.

One of the reasons it is a good thing is that I think a lot of quantitative researchers – both in the organizations and management fields, and in my own field of sociology – take too much comfort in the power of control variables.

The important thing to take away from the examples of simultaneity and self-selection is that it is impossible to include enough control variables in a regression model, but that does not mean that you should try to collect more.

In general I think a lot of researchers, in thinking about their research designs, misallocate effort. They dedicate too much time to collecting a wide range of control variables, even though it is often the case that many of those variables either don't affect the X or affect the Y (or both). Simcoe and Waguespack did not add value by expanding the list of control variables.

Instead, the efforts of researchers are often better focused on thinking about actual research design -- how they might anticipate and address concerns about endogeneity in the form of things like simultaneity and self-selection.

This involves thinking deeply about the X and the Y – the two constructs you are most interested in. Of course we already do that when we develop our theoretical claims. So really what I am saying is that we need to think deeply about the X – and in particular the processes that bring about the X.

What you will notice about simultaneity and self-selection is that they are both processes that bring about the level of X. And the extent to which we are concerned about them is the extent to which those processes are related to the Y variable.

Yet all too often, I think, researchers don't think very deeply about the social processes that generate their independent variables of interest. I think this is due to a false analogy between regression models on observational or field data and experimental data. (An observation made long ago by Stan Lieberson.)

In an experiment – of course the beauty of an experiment – different people are assigned to different levels of X by chance. This is the beauty of an experiment – the reason why it allows us to feel confident in our causal inferences.

In other words, when we look at our X variable, we all too often think that different people are assigned to different levels of the X variable by chance. Or we think that once we have controlled for enough variables, the assignment to different levels of X is by chance.

Of course if you asked people that explicitly, they would probably be appropriately skeptical and say that the assignment process is non-random. But the problem is that people don't think about this explicitly.

And I think a real benefit of the growing emphasis on endogeneity concerns is to force people to be explicit about this assumption. Right now, people often come to face this issue in the review process. But we can hope that over time people will internalize this concern in the same way that they have internalized the concerns with omitted variables.

What I really want to emphasize is this last point: that you want to internalize the concern with selectivity and self-selection *early* in the process: at the research design stage. A lot of the language and techniques around endogeneity is statistical – instrumental variables, sample selection models, etc. But these techniques are often times solutions to problems caused by the limitations of the research design.

I've emphasized some of the positives of the growing emphasis on endogeneity concerns. I do not have any doubt that this is a positive force in management research. And the creativity and effort that has been put into a lot of this work is impressive.

But there can always be too much of a good thing. It is important to think about the tradeoffs.

In thinking about the possible tradeoffs, I think it is most helpful to think about the difference between theory generation and theory testing. A lot of people think this is a distinction between abstract, conceptual work – thinking – on the one hand, and empirical analysis on the other. But I think this is wrong, since the abstract conceptual work does not happen without engagement with the empirical world, and since the empirical testing that we do does not happen without a conceptual overlay.

From a research design perspective, there is a tradeoff between theory generation and theory testing. In my experience, the more compelling the test that an author performs, the less creative is the theorizing. This is not to say that the theorizing is dull or uninteresting. But rather it is to say that the paper with the awesomely convincing test of the causal mechanism typically does not do much to generate a new insight – typically you are testing an idea that already exists. In short, as many others have noted, there is a tradeoff between fertility and rigor.

There are several reasons for this. One is that you only have so much creative energy, and things need to get published and get out into the debate. And so in any given project you have to allocate your creative energy, I think, either to generating new ideas and theories, or to designing the compelling test of the causal mechanism. Second, creativity demands imagination and speculation; a willingness to say “what if” and work out the details later. Yet a rigorous causal test is all about making the causal inference as immune as possible to the imaginative speculations of your critics. The better they can do in terms of dreaming up alternative interpretations, the worse you have done in providing a robust test. There are likely other explanations as well.

If you believe that this tradeoff is real – and you should – then it has a number of important implications. The first is that if there are real tradeoffs, the worst papers are those that pretend that there is no tradeoff – i.e., that pretend that it does not exist. For example, they are theoretically creative papers that vastly over-claim with respect to the strength of their evidence. Or they are papers with elegantly designed critical tests of existing theory that simultaneously try to convince you that they contain path-breaking ideas. Second, as a consequence, you should try to decide for yourself which side of the tradeoff you prefer. This does not have to be a lifelong decision, but for any given paper you should make a decision. Don't do a half-assed job on theory and a half-assed job on the empirics. Decide where your strengths lie, and where your project's strengths lie, and have the courage of your convictions. Finally, as a reviewer and critic of other people's work, keep the tradeoff in mind. If you review a paper that is clearly positioned as providing a more rigorous test of an important idea, it is not fair game to complain about the fact that there is no new theory. Similarly, if a paper goes a bit beyond what the

empirical evidence might support, but is circumspect about its limitations, don't pile on because the paper lacks true random assignment. Any one paper cannot do everything; our business is all about triangulating.