11.5.3 Defending the Causal Interpretation of SEM (or, SEM Survival Kit)

Question to Author:

J. Wilson from Surrey, England, asked about ways of defending his Ph.D. thesis before examiners who do not approve of the causal interpretation of structural equation models (SEM). He complained about “the complete lack of emphasis in PhD programmes on

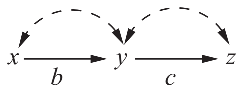

Figure 11.14 The graph underlying equations (11.21)–(11.22).

how to defend causal interpretations and policy implications in a viva when SEM is used … if only causality had been fully explained at the beginning of the programme, then each of the 70,000 words used in my thesis would have been carefully measured to defend first the causal assumptions, then the data, and finally the interpretations … (I wonder how widespread this problem is?) Back to the present and urgent task of trying to satisfy the examiners, especially those two very awkward Stat Professors – they seem to be trying to outdo each other in nastiness."

***Author's Reply*:**

The phenomenon that you complain about is precisely what triggered my writing of Chapter 5 – the causal interpretation of SEM is still a mystery to most SEMs researchers, leaders, educators, and practitioners. I have spent hours on SEMNET Discussion List trying to rectify the current neglect, but it is only students like yourself who can turn things around and help reinstate the causal interpretation to its central role in SEM research.

As to your concrete question – how to defend the causal interpretation of SEM against nasty examiners who oppose such interpretation – permit me to assist by sketching a hypothetical scenario in which you defend the causal interpretation of your thesis in front of a hostile examiner, Dr. EX. (Any resemblance to Dr. EX is purely coincidental.)

A Dialogue with a Hostile Examiner
or
SEM Survival Kit

For simplicity, let us assume that the model in your thesis consists of just two-equations,

$$y = bx + e_1 \tag{11.21}$$

$$z = cy + e_2, \tag{11.22}$$

with e_2 uncorrelated with x. The associated diagram is given in Figure 11.14. Let us further assume that the target of your thesis was to estimate parameter c, that you have estimated c satisfactorily to be $c = 0.78$ using the best SEM methods, and that you have given a causal interpretation to your finding.

Now your nasty examiner, Dr. EX, arrives and questions your interpretation.

Dr. EX: What do you mean by "c has a causal interpretation"?

You: I mean that a unit change in y will bring about a c units change in $E(Z)$.

Dr. EX: The words "change" and "bring about" make me uncomfortable; let's be scientific. Do you mean $E(Z \mid y) = cy + a$??? I can understand this last expression, because the conditional expectation of Z given y, $E(Z \mid y)$, is well defined mathematically, and I know how to estimate it from data. But "change" and "bring about" is jargon to me.

You: I actually mean "change," not "an increase in conditional expectation," and by "change" I mean the following: If we had the physical means of fixing y at some constant y_1, and of changing that constant from y_1 to y_2, then the observed change in $E(Z)$ would be $c(y_2 - y_1)$.

Dr. EX: Well, well, aren't we getting a bit metaphysical here? I never heard about "fixing" in my statistics classes.

You: Oh, sorry, I did not realize you have statistics background. In that case, let me rephrase my interpretation a bit, to read as follows: If we had the means of conducting a controlled randomized experiment, with *y* randomized, then if we set the control group to y_1 and the experimental group to y_2, the observed difference in $E(Z)$ would be $E(Z_2) - E(Z_1) = c(y_2 - y_1)$ regardless of what values y_1 and y_2 we choose. (Z_1 and Z_2 are the measurements of *z* under the control and experimental groups, respectively.)[14]

Dr. EX: That sounds much closer to what I can understand. But I am bothered by a giant leap that you seem to be making. Your data was nonexperimental, and in your entire study you have not conducted a single experiment. Are you telling us that your SEM exercise can take data from an observational study, do some LISREL analysis on it, and come up with a prediction of what the outcome of a controlled randomized experiment will be? You've got to be kidding!! Do you know how much money can be saved nationwide if we could replace experimental studies with SEM magic?

You: This is not magic, Dr. EX, it is plain logic. The input to my LISREL analysis was more than just nonexperimental data. The input consisted of two components: (1) data, (2) causal assumptions; my conclusion logically follows from the two. The second component is absent in standard experimental studies, and that is what makes them so expensive.

Dr. EX: What kind of assumptions? "Causal"? I never heard of such strangers. Can you express them mathematically the way we normally express assumptions – say, in the form of conditions on the joint density, or properties of the covariance matrix?

You: Causal assumptions are of a different kind; they cannot be written in the vocabulary of density functions or covariance matrices. Instead, they are expressed in my causal model.

Dr. EX: Looking at your model, equations (11.21)–(11.22), I do not see any new vocabulary; all I see is equations.

You: These are not ordinary algebraic equations, Dr. EX. These are "structural equations," and if we read them correctly, they convey a set of assumptions with which you are familiar, namely, assumptions about the outcomes of hypothetical randomized experiments conducted on the population – we call them "causal" or "modeling" assumptions, for want of better words, but they can be understood as assumptions about the behavior of the population under various randomized experiments.

Dr. EX: Wait a minute! Now that I begin to understand what your causal assumptions are, I am even more puzzled than before. If you allow yourself to make assumptions about the behavior of the population under randomized experiments, why go through the trouble of conducting a study? Why not make the assumption directly that in a randomized experiment, with *y* randomized, the observed difference in $E(Z)$ should be $c'(y_2 - y_1)$, with c' just any convenient number, and save yourself agonizing months of data collection and analysis. He who believes your other untested assumptions should also believe your $E(Z_2) - E(Z_1) = c'(y_2 - y_1)$ assumption.

---

[14] Just in case Dr. EX asks: "Is that the only claim?" you should add: Moreover, I claim that the distribution of the random variable $Z_1 - cy_1$ will be the same as that of the variable $Z_2 - cy_2$.

You: Not so, Dr. EX. The modeling assumptions with which my program begins are much milder than the assertion $E(Z_2) - E(Z_1) = 0.78(y_2 - y_1)$ with which my study concludes. First, my modeling assumptions are qualitative, while my conclusion is quantitative, making a commitment to a specific value of $c = 0.78$. Second, many researchers (including you, Dr. EX) would be prepared to accept my assumptions, not my conclusion, because the former conforms to commonsense understanding and general theoretical knowledge of how the world operates. Third, the majority of my assumptions can be tested by experiments that do not involve randomization of y. This means that if randomizing y is expensive, or infeasible, we still can test the assumptions by controlling other, less formidable variables. Finally, though this is not the case in my study, modeling assumptions often have some statistical implications that can be tested in nonexperimental studies, and, if the test turns out to be successful (we call it "fit"), it gives us further confirmation of the validity of those assumptions.

Dr. EX: This is getting interesting. Let me see some of those "causal" or modeling assumptions, so I can judge how mild they are.

You: That's easy, have a look at our model, Figure 11.14, where

z – student's score on the final exam,
y – number of hours the student spent on homework,
x – weight of homework (as announced by the teacher) in the final grade.

When I put this model down on paper, I had in mind two randomized experiments, one where x is randomized (i.e., teachers assigning weight at random), the second where the actual time spent on homework (y) is randomized. The assumptions I made while thinking of those experiments were:

1. Linearity and exclusion for y: $E(Y_2) - E(Y_1) = b(x_2 - x_1)$, with b unknown (Y_2 and Y_1 are the time that would be spent on homework under announced weights x_2 and x_1, respectively.) Also, by excluding z from the equation, I assumed that the score z would not affect y, because z is not known at the time y is decided.
2. Linearity and exclusion for z: $E(Z_2) - E(Z_1) = c(y_2 - y_1)$ for all x, with c unknown. In words, x has no effect on z, except through y.

In addition, I made qualitative assumptions about unmeasured factors that govern x under nonexperimental conditions; I assumed that there are no common causes for x and z.

Do you, Dr. EX, see any objection to any of these assumptions?

Dr. EX: Well, I agree that these assumptions are milder than a blunt, unsupported declaration of your thesis conclusion, $E(Z_2) - E(Z_1) = 0.78(y_2 - y_1)$, and I am somewhat amazed that such mild assumptions can support a daring prediction about the actual effect of homework on score (under experimental setup). But I am still unhappy with your common cause assumption. It seems to me that a teacher who emphasizes the importance of homework would also be an inspiring, effective teacher, so e_2 (which includes factors such as quality of teaching) should be correlated with x, contrary to your assumption.

You: Dr. EX, now you begin to talk like an SEM researcher. Instead of attacking the method and its philosophy, we are beginning to discuss substantive issues – e.g., whether it is reasonable to assume that a teacher's effectiveness is uncorrelated with the weight

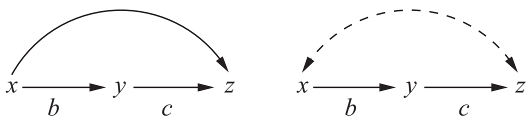

Figure 11.15 Statistically equivalent models to Figure 11.14.

that the teacher assigns to homework. I personally have had great teachers that could not care less about homework, and conversely so.

But this is not what my thesis is about. I am not claiming that teachers' effectiveness is uncorrelated with how they weigh homework; I leave that to other researchers to test in future studies (or might it have been tested already?). All I am claiming is: Those researchers who are willing to accept the assumption that teachers' effectiveness is uncorrelated with how they weigh homework will find it interesting to note that this assumption, coupled with the data, logically implies the conclusion that an increase of one homework-hour per day causes an (average) increase of 0.78 grade points in student's score. And this claim can be verified empirically if we are allowed a controlled experiment with randomized amounts of homework (y).

Dr. EX: I am glad you do not insist that your modeling assumptions are true; you merely state their plausibility and explicate their ramifications. I cannot object to that. But I have another question. You said that your model does not have any statistical implications, so it cannot be tested for fitness to data. How do you know that? And doesn't this bother you?

You: I know it by just looking at the graph and examining the missing links. A criterion named *d*-separation (see Section 11.1.2, "*d*-separation without tears") permits students of SEM to glance at a graph and determine whether the corresponding model implies any constraint in the form of a vanishing partial correlation between variables. Most statistical implications (though not all) are of this nature. The model in our example does not imply any constraint on the covariance matrix, so it can fit perfectly any data whatsoever. We call this model "saturated," a feature that some SEM researchers, unable to shake off statistical-testing traditions, regard as a fault of the model. It isn't. Having a saturated model at hand simply means that the investigator is not willing to make implausible causal assumptions, and that the mild assumptions he/she is willing to make are too weak to produce statistical implications. Such a conservative attitude should be commended, not condemned. Admittedly, I would be happy if my model were not saturated – say, if e_1 and e_2 were uncorrelated. But this is not the case at hand; common sense tells us that e_1 and e_2 are correlated, and it also shows in the data. I tried assuming $cov(e_1, e_2) = 0$, and I got terrible fit. Am I going to make unwarranted assumptions just to get my model "knighted" as "nonsaturated"? No! I would rather make reasonable assumptions, get useful conclusions, and report my results side by side with my assumptions.

Dr. EX: But suppose there is another saturated model, based on equally plausible assumptions, yet leading to a different value of c. Shouldn't you be concerned with the possibility that some of your initial assumptions are wrong, hence that your conclusion $c = 0.78$ is wrong? There is nothing in the data that can help you prefer one model over the other.

You: I am concerned indeed, and, in fact, I can immediately enumerate the structures of all such competing models; the two models in Figure 11.15 are examples, and many more. (This too can be done using the *d*-separation criterion; see pp. 145–8.) But note that the existence of competing models does not in any way weaken my earlier stated claim: "Researchers who accept the qualitative assumptions of model *M* are compelled to accept the conclusion $c = 0.78$." This claim remains logically invincible. Moreover, the claim can be further refined by reporting the conclusions of each contending model, together with the assumptions underlying that model. The format of the conclusion will then read:

If you accept assumption set A_1, then $c = c_1$ is implied,
If you accept assumption set A_2, then $c = c_2$ is implied,

and so on.

Dr. EX: I see, but still, in case we wish to go beyond these conditional statements and do something about deciding among the various assumption sets, are there no SEM methods to assist one in this endeavor? We, in statistics, are not used to facing problems with two competing hypotheses that cannot be submitted to some test, however feeble.

You: This is a fundamental difference between statistical data analysis and SEM. Statistical hypotheses, by definition, are testable by statistical methods. SEM models, in contrast, rest on *causal* assumptions, which, also by definition (see p. 39), cannot be given statistical tests. If our two competing models are saturated, we know in advance that there is nothing more we can do but report our conclusions in a conditional format, as listed above. If, however, the competition is among equally plausible yet statistically distinct models, then we are facing the century-old problem of model selection, where various selection criteria such as AIC have been suggested for analysis. However, the problem of model selection is now given a new, causal twist – our mission is not to maximize fitness, or to maximize predictive power, but rather to produce the most reliable estimate of causal parameters such as *c*. This is a new arena altogether (see Pearl 2004).

Dr. EX: Interesting. Now I understand why my statistician colleagues got so totally confused, mistrustful, even antagonistic, upon encountering SEM methodology (e.g., Freedman 1987; Holland 1988; Wermuth 1992). One last question. You started talking about randomized experiments only after realizing that I am a statistician. How would you explain your SEM strategy to a nonstatistician?

You: I would use plain English and say: "If we have the physical means of fixing *y* at some constant y_1, and of changing that constant from y_1 to y_2, then the observed change in $E(Z)$ would be $c(y_2 - y_1)$." Most people understand what "fixing" means, because this is on the mind of policy makers. For example, a teacher interested in the effect of homework on performance does not think in terms of randomizing homework. Randomization is merely an indirect means for predicting the effect of fixing.

Actually, if the person I am talking to is really enlightened (and many statisticians are), I might even resort to counterfactual vocabulary and say, for example, that a student who scored *z* on the exam after spending *y* hours on homework would have scored $z + c$ had he/she spent $y + 1$ hours on homework. To be honest, this is what I truly had in mind when writing the equation $z = cy + e_2$, where e_2 stood for all other characteristics of the student that were not given variable names in our model and that are not affected by *y*. I did not even think about $E(Z)$, only about *z* of a typical student. Counterfactuals are the most precise linguistic tool we have for expressing the meaning of scientific relations. But I refrain from mentioning counterfactuals when I talk to statisticians because, and this is regrettable, statisticians tend to suspect deterministic concepts, or concepts that are not immediately testable, and counterfactuals are such concepts (Dawid 2000; Pearl 2000).

Dr. EX: Thanks for educating me on these aspects of SEM. No further questions.

You: The pleasure is mine.