Non-recursive SEM and Instrumental Variables

Say we are interested in the total causal effect of X on Y. But, we know there is a W out there such that the following is true. As described earlier in Path Analysis, in order to unbiasedly estimate the causal effect of X on Y, W should be included in a regression equation. That is, \( Y = \beta_0 + \beta_1 X + \beta_2 W + \text{error} \) will give an unbiased estimate of the total causal effect, but \( Y = \beta_0 + \beta_1 X + \text{error} \) without W will NOT.

![Figure 1](image1)

But what if we do not observe W nor have any sort of measure of it? Another way to describe the situation is that X is correlated with the error in the regression of Y on X, i.e.

![Figure 2](image2)

Recall that one of the basic assumptions of OLS regression producing unbiased estimates (one that I said is not usually emphasized and often ignored) is that the predictors are uncorrelated with the errors, i.e. X is uncorrelated with error. Well if we do not measure W, then in the regression of \( Y = \beta_0 + \beta_1 X + \text{error}_1 \), X will be correlated with the error1.

Without further information, it is not possible to unbiasedly estimate the causal effect of X on Y. In particular, it is not possible to separate the effect that X has on Y from the effect that X has on Y through its correlation with the error. But, what if we have a variable Z at our disposal that satisfies the following diagram:

![Figure 3](image3)

Figure 3. Z is called an **instrumental variable** for X. That is, it is a direct cause of X AND it is not a direct cause of Y. Specifically it is not correlated with the error2 nor the error1.

If we can find and measure such a variable Z, then it is possible to estimate the causal effect of X on Y because now all the paths in the model are identifiable. That is, the correlated errors can be separated from the direct effect X has on Y.
Below is a demonstration where I have generated data such as that in Figure 1, i.e. where there is a confounding variable $W$. The true causal effect of $X$ on $Y$ equals 1 in the simulated data and I’ve allowed $W$ to have a positive influence on both $X$ and $Y$. I have also generated an instrumental variable $Z$ like in Figure 3 that will allow us to estimate the causal effect of $X$ on $Y$ even though we don’t have $W$ at our disposal.

The comparison is between two estimation methods, OLS regression and a full normally distributed structural equation model. Note, there is another estimation method called Two Stage Least Squares (2SLS) and in the case of normal data it is equivalent to full SEM (though the standard errors are less efficient).

Result of 100 simulated datasets (in 2 scenarios for sample size = 200 and also 1000) and associated estimator of the direct causal effect of $X$ on $Y$ using simple OLS regression of $Y = \beta_0 + \beta_1 X + \text{error}$ verses using a structural equation model that simultaneously estimates the two equations $Y = \beta_0 + \beta_1 X + \text{error}_1$, and $X = \alpha_0 + \alpha_1 Z + \text{error}_2$ allowing error1 and error2 to be correlated.

The true causal effect of $X$ on $Y$ is 1.0.

Notice that the 100 different estimators under the two sample size scenarios show OLS produces a biased effect (bigger on average than it should be), while the SEM approach on average is giving the right answer of 1.

Side note: The method of 2SLS proceeds as follows, OLS is applied to $Y = \beta_0 + \beta_1 X_{\text{hat}} + \text{error}$ (2nd stage) where $X_{\text{hat}}$ is obtained from OLS regression of $X = \alpha_0 + \alpha_1 Z + \text{error}$ (1st stage).

NOTE: If $Z$ does not satisfy the requirements of being an “instrument” then the results will still be biased. It is often hard to come up with a variable $Z$ that can serve as an instrument since usually things that are causes of $X$ will often also be direct causes of $Y$ or correlated with other causes of $Y$.

Recall that a non-recursive SEM is one that cannot be fit by a series of OLS regressions. A non-recursive SEM involves models where there are reciprocal feedback loops or correlated error terms (as in Figure 3).
One example where this set up is reasonable is in the “As Treated” analysis of a clinical trial. In clinical trials we often talk about the “Intent to treat” analysis which is usually the main planned analysis which looks at an outcome by the randomized treatment groups. The lack of any possible confounding in an “Intent to treat” analysis is why it is and will remain the gold standard (e.g. random treatment assignment is independent of any potential confounder variable such as disease severity, hence no variables W need to be included for the “intent to treat” analysis).

Since often in the real world not everyone complies with treatment, the causal effect being estimated by an “intent to treat” analysis is aptly named, that is, it is the causal effect of telling someone to take the drug versus not to take the drug, or in other words, intending to treat someone whether they actually follow through or not. The estimated effect represents a realistic view of what would happen if the drug was used in the real world where people don’t always do what they are intended to do. So the estimated causal effect is averaging over non-compliance.

But, as researchers wanting to know whether a treatment potentially could work or not, we may be interested in estimating the “truly treated” or “As treated” effect. That is, estimating the causal effect of actually taking the drug, not just being randomly assigned to take the drug. If we let Z be the random assignment of a person to an Active Drug versus a Placebo and we let X be the actual truth of whether the person took the Active Drug or not and Y be the Main outcome of interest we are trying to influence. The problem is with an “As treated” analysis, we might expect there is some confounding variable W that influences both the likelihood that someone may comply as well as causing the outcome itself regardless of treatment. An example might be disease severity. Perhaps someone who is more sick will be more likely to comply with the treatment but they also will be more likely to have poor outcomes. Hence if this confounder was ignored in an “as treated” analysis simply looking at how X is related to Y, it might end up looking like the treatment is causing worse outcomes.

So rather than just regress Y on X, we can use Z as an instrument and fit the SEM in Figure 3. In most cases it is reasonable to think that Z (randomized group) is only able to directly influence X but not anything else that is influencing Y. In other words, to estimate the causal effect of X on Y, we do not need to have measured the confounder(s) W since we have the instrument Z to allow us to separate the causal effect from the confounded correlation between X and Y.

-Bound Jaeger Baker (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, JASA, 90, 443-450.