Should We Worry About Unmeasured Variables in Organizational Research?

Adam W. Meade & Tara S. Behrend

Download slides at:
www4.ncsu.edu/~awmeade/

The Legend

- In path analysis and structural equation modeling, failure to include all relevant causes may invalidate study results
This issue is also referred to as:

- Unmeasured variables problem
  - Duncan, 1975; James, 1980
- Omitted variables problem
  - James, 1980; Kenny, 1979; Sackett, Laczo, & Lippe, 2003
- Left out variables error (L.O.V.E.)
  - Mauro, 1990
- Lack of perfect isolation (i.e., pseudo-isolation)
  - Bollen, 1989
- Lack of self-containment
  - James, Mulaik, & Brett, 1982
- Is a particular type of
  - specification error (Kenny, 1979)
  - model misspecification (Hanushek & Jackson, 1977)

Presentation Goals

1. Define the issue (there will be some math)
2. Discuss conditions under which omitted variables may be problematic or not problematic
3. Provide recommendations for reducing the risk of L.O.V.E.
Basics of the problem

Regression coefficients estimate the unique effect of a predictor on a criterion.

\[ Y = \beta_{yx}X + d \]

\[ \beta_{yx} = r_{yx} \]

\[ Y = \beta_{yx1}X_1 + \beta_{yx2}X_2 + d \]

\[ \beta_{yx1} = \frac{r_{yx1} - r_{yx2}r_{x1x2}}{1 - r^2_{x1x2}} \quad \beta_{yx2} = \frac{r_{yx2} - r_{yx1}r_{x1x2}}{1 - r^2_{x1x2}} \]

So, what happens when a variable is omitted?

First, let’s define some basic terms in a path model.

- \( X_1 \) and \( X_2 \) are exogenous causes of \( Y \)
- \( d \) is a disturbance term composed of:
  - Random shocks (unstable influences)
  - Measurement error
  - Omitted relevant causes of \( Y \)
  - Omitted non-relevant causes of \( Y \)

From James, Mulaik, & Brett, 1982
What happens when a variable is omitted?

- Model B represents a situation where one cause is omitted.

In Model A:

- Regression coefficients are estimated as:

\[
\beta_{yx1} = \frac{r_{yx1} - r_{yx2}r_{x1x2}}{1 - r_{x1x2}^2} \quad \beta_{yx2} = \frac{r_{yx2} - r_{yx1}r_{x1x2}}{1 - r_{x1x2}^2}
\]
What happens when a variable is omitted?

- In Model B

```
(B) X \beta_{xy} \quad d
```

- Only $\beta_{yx}$ is estimated
- $\beta_{yx}$ should be estimated as:

$$
\beta_{yx} = \frac{r_{yx} - r_{xo}r_{yo}}{1 - r_{xo}^2}
$$

- However, $\beta_{yx}$ is estimated as $r_{yx}$

What happens when a variable is omitted?

- In Model B

```
(B) X \beta_{xy} \quad d
```

- $\beta_{yx}$ is estimated as $r_{yx}$
- Where $r_{xy} = \beta_{yx} + r_{xo}\beta_{yo}$ in the true model
- So, $\beta_{yx}$ is biased by a factor of $r_{xo}\beta_{yo}$. 
Sources of bias

- $\beta_{yx}$ is biased by a factor of $r_{xo}^2 \beta_{yo}$.
- So, if $r_{xo} = 0$, there is no bias and O is an non-relevant cause of Y.

The role of $\beta_{yo}$ is more complicated

\[
\beta_{yo} = \frac{r_{yo} - r_{yx} r_{xo}}{1 - r_{xo}^2}
\]

- $r_{yo}$ also has an impact on the amount of bias present.
- However, the effect of $r_{yo}$ is tempered by $r_{xo}$.

- Note that if $r_{yo}$ is small, $\beta_{yo}$ can take on negative values. In this case, $r_{yx}$ may actually underestimate $\beta_{yx}$.
- This is a suppressor effect.
What happens when a variable is omitted?

- Disturbance term in model includes effects of relevant cause O.
- O correlates with X.
- By assumption, causes cannot be correlated with the disturbance term of the endogenous variable.
- In this case, X will correlate with d by a magnitude of $r_{xo} \beta_{yo}$.
So in sum…

- If O is uncorrelated with the predictor:
  - \( r_{xy} \) is an unbiased estimator of \( \beta_{yx} \)
  - There is no effect of the unmeasured variable
  - O is a non-relevant cause of Y
- If the variance in Y accounted for by O is highly redundant with the predictors in the model:
  - Its unique effect (\( \beta_{yo} \)) will be near zero
  - It will have little biasing effect
- If O is uncorrelated with the criterion, but strongly correlated with the predictor:
  - \( r_{xy} \) may underestimate \( \beta_{yx} \)
  - i.e., a suppressor effect

More summary

- In order for an omitted variable to cause positive bias in estimated path coefficients, the variable must:
  A. Correlate at a non-zero level with other determinants of Y,
  B. Not be completely redundant with other variables included in the path model, and
  C. Correlate with the endogenous variable
- If (a) and (b) are true, but (c) is not, the omitted variable may serve to artificially deflate the estimate of the regression coefficient of the variables included in the model.
Larger models produce more complexity

Where $X_1$ is a measured exogenous variable, $O$ is an omitted exogenous variable, $X_2$ and $Y$ are measured endogenous variables.

The true effect is:

$$
\beta_{yx1} = \frac{r_{yx1}(1-r_{x2o}^2) + r_{yx2}(r_{x1o}r_{x2o} - r_{x1x2}) + r_{yo}(r_{x1x2}r_{x2o} - r_{x1o})}{1 + 2r_{x1x2}r_{x2o}r_{x1o} - r_{x1o}^2 - r_{o2o}^2 - r_{x1x2}^2}
$$

But the estimate will be:

$$
\hat{\beta}_{yx1} = \frac{r_{yx1} - r_{yx2}r_{x1x2}}{1-r_{x1x2}^2}
$$
Path Coefficient Bias Versus Significance Testing

- So, omitted variables can bias path coefficients in SEM and path analysis.
- However, perhaps the question of greater interest is how often omitted variables change our conclusions about the effects of observed variables in the model.
- There can be bias in the path coefficient, perhaps even severe bias, yet the significance test associated with this coefficient may not change.

In other words, the answer to the question “does the exogenous variable have an effect on the endogenous variable?” would seem more important than the question “what is the magnitude of the unique effect of the exogenous variable on the endogenous variable?” (Mauro, 1990)

Significant effects will likely lead to the inclusion of the predictor in future (perhaps more comprehensive) models.
Is your research at risk?

- Are there a large number of determinants of outcome variables?
- Do you typically deal with a small subset of these potential determinants?
- Is it likely that the omitted variables have moderate or large correlations with your predictors?
- It is likely that the omitted variables would account for unique variance in the outcome variables?
- If you answered yes to all of the above, you might be at risk for L.O.V.E.

Tips for dealing with omitted variables

1. When possible, randomly assign
   - With large samples, random assignment will lead to a correlation of zero between observed and omitted variables.
   - A zero correlation means no bias in regression coefficients is possible.
   - Thus, random assignments makes other causes non-relevant.
   - Examples: Training courses, reward systems, equipment and other environmental factors, or organizational interventions
Tips for dealing with omitted variables

2. Include as many known causes of the endogenous variable as possible in the path model
   - Bias comes from the unique effect of an omitted variable
   - When more variables are included in the model, less unique variance will be remain that could be explained by the omitted variable
   - However, note that when you add variables to a model, the same assumptions regarding self-containment apply to those variables too!

3. Use what is known from previous research to make the case for the exclusion of entire classes of potential predictors
   - For example, ability, personality, and situational predictors of job performance are not typically highly correlated
Tips for dealing with omitted variables

4. Lastly, consider the purpose of the research.
   - If the goal is to provide a precise estimate of path coefficients, or to compare the relative variance accounted for by different predictors, omitted variables are considerably more problematic than if the goal is to test the statistical significance of the effect of a predictor on a criterion.
   - Most hypotheses in organizational research are proposed as the latter.
   - The latter also has a larger contribution to theory building.

Summary

- In organizational research, omitted variables are a fact of life and they can be problematic.
- However, the notion that omitted variables are always problematic is a myth, as the threat to the inferences that we tend to draw may not be as serious as some believe.