ENDOGENEITY: AN OVERLOOKED THREAT TO VALIDITY OF CROSS-SECTIONAL RESEARCH

John Antonakis

Professor of Organizational Behavior
Faculty of Business and Economics
University of Lausanne

Research seminar

28 February 2012

What is the point of research?

The ultimate aim of research is to develop theory (Kerlinger, 1986) and that theory is:

"a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying [causal] relations among the variables, with the purpose of explaining and predicting the phenomena" (p. 9)

….which has important practical and policy implications.

What is Causality?

To measure causal effects, we need an effect (y) and a presumed cause (x). Classically, x is said to have an effect on y if the following three conditions are met (Holland, 1986; Kenny, 1979):

   (a) y follows x temporally.

   (b) y changes as x changes (and this relationship is statistically significant).

   (c) no other causes should eliminate the relation between x and y.

Causal claims, whether explicit or implicit, are ubiquitous and important in non-experimental social science research (and policy):

- *Explicit*: "x causes, predicts, affects, influences, explains, or is an antecedent of y" or that "y depends on x."

- *Implicit*: "y is associated or related to x"

If x is not exogenous then its relation to y is not interpretable, not even as an association or relation (i.e., it is *inconsistent* because of *endogeneity*).

What does exogenous mean? An exogenous variable:

- Varies randomly in nature or is manipulated

- Is not determined by variables in, or omitted from, the model: *x* does

  not correlate with the error term:

$$\psi 1 = 0 \qquad\qquad\qquad\qquad \psi 2 \neq 0$$

Figure 1A

Figure 1B
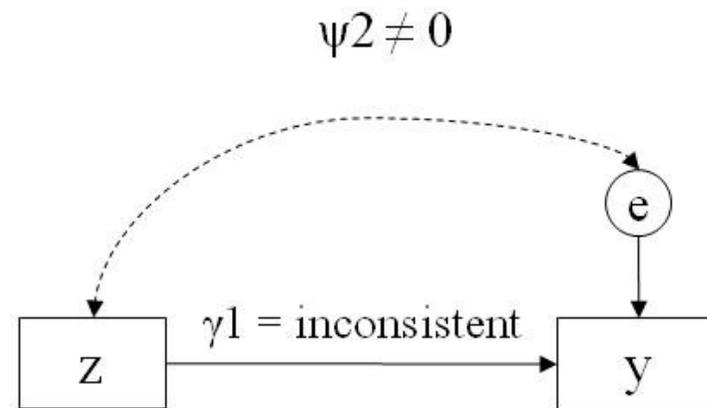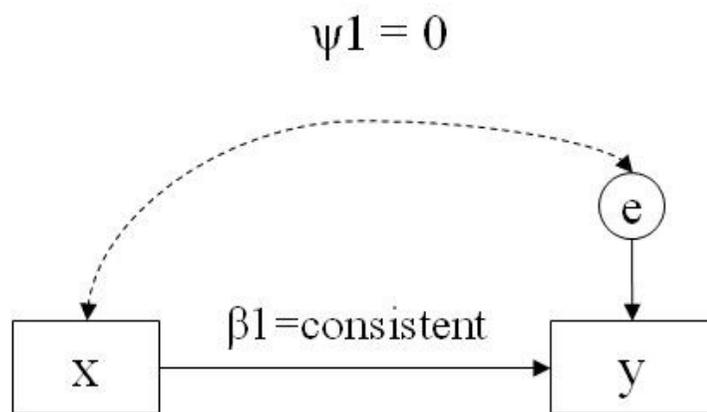
Video presentation
(available on Youtube as "Endogeneity: An inconvenient truth")
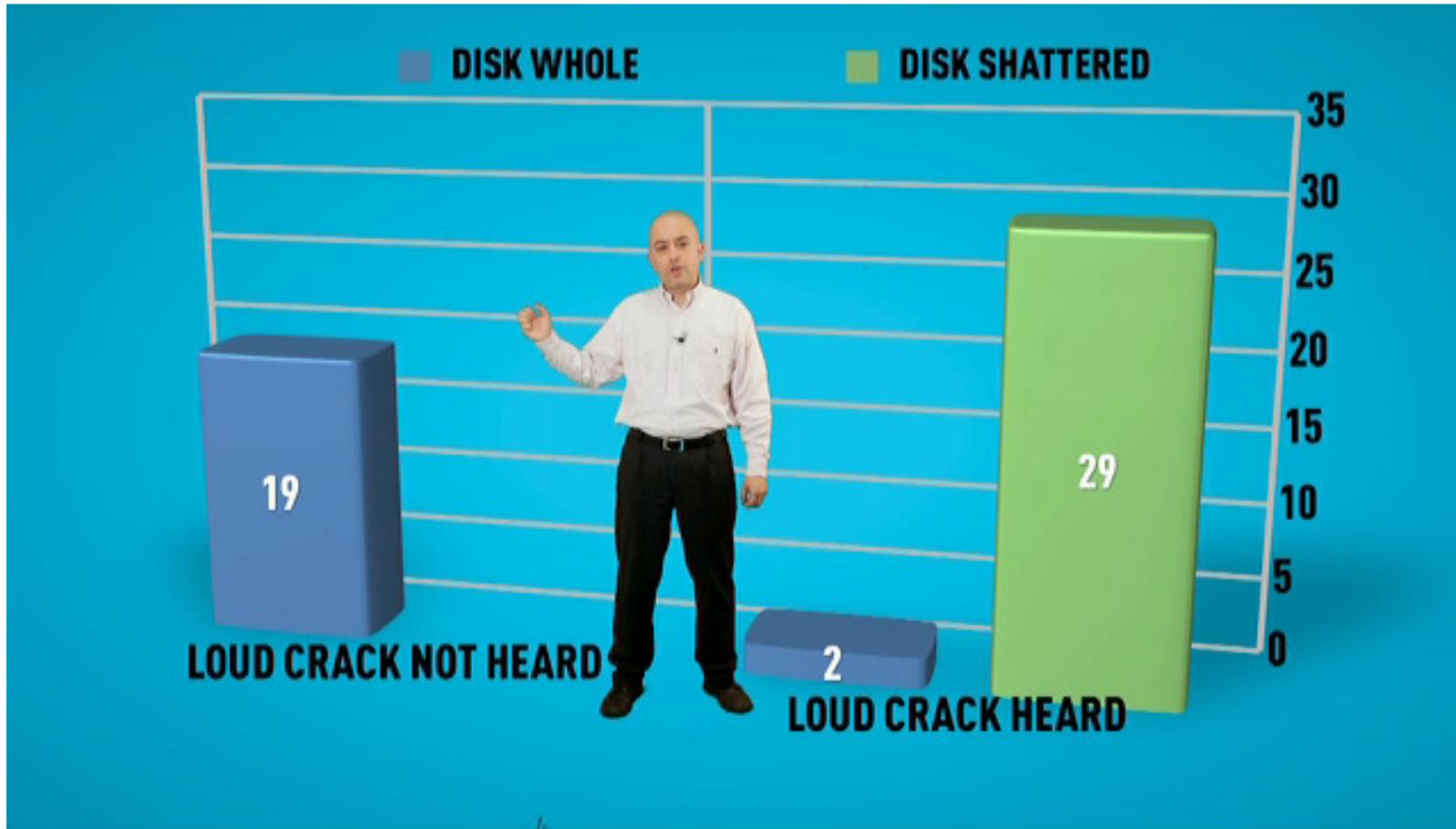
01:50 – 08:50

The simple case of Endogeneity: Omitted variables

Suppose the true model is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i \qquad \text{Eq. 1}$$

Instead of Eq. 1, we estimate:

$$y_i = \varphi_0 + \varphi_1 x_i + v_i \qquad \text{Eq. 2}$$

In the presence of endogeneity is: $\varphi_1 = \beta_1$?

If z and x correlated (irrespective of the direction), then we can note that:

$$z_i = \gamma_1 x_i + u_i \qquad \text{Eq. 3}$$

The endogeneity is evident when substituting Eq. 3 into Eq. 1:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (\gamma_1 x_i + u_i) + e_i, \qquad \text{Eq. 4}$$

Multiplying out gives (notice, the error term v_i, which is the error term of Eq. 2):

$$y_i = \beta_0 + \beta_1 x_i + \underbrace{(\beta_2 \gamma_1 x_i + \beta_2 u_i + e_i)}_{v_i} \qquad \text{Eq. 5}$$

Or, rearranging as a function of x gives

$$y_i = \beta_0 + (\beta_1 + \beta_2 \gamma_1) x_i + (\beta_2 u_i + e_i) \qquad \text{Eq. 6}$$

In the presence of endogeneity, $\beta_1(Eq.\,1) \neq \varphi_1(Eq.\,2)$ unless:

(a) $\beta_2 = 0$ or

(b) $\gamma_1 = 0$.

Whether φ_1 is increases or decreases when excluding z will depend on the signs and magnitudes of β_2 and γ_1.

The common-methods variance problem

Researchers sometimes ask respondents to provide answers on x and y. The problem with doing so is that individuals may "align" the ratings in ways that:

- make intuitive sense

- are socially desirable

- may be driven by an omitted cause like a "halo" effect (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff, MacKenzie, & Podsakoff, 2012; Podsakoff & Organ, 1986).

Assume the following true model, where q is the common-source effect:

$$y_{ij}^* = \beta_0 + \beta_1 x_{ij}^* + \sum_{k=2}^{50} \beta_k D_{kj} + e_{ij}$$

Eq. 7

We do not directly observe y^* or x^*, but y and x, which such that:

$$y_{ij} = y_{ij}^* + \gamma_y q_{ij}$$

Eq. 8

$$x_{ij} = x_{ij}^* + \gamma_x q_{ij}$$

Eq. 9

The two later equations can be rearranged as follows:

$$y_{ij}^* = y_{ij} - \gamma_y q_{ij}$$

Eq. 10

$$x_{ij}^* = x_{ij} - \gamma_x q_{ij}$$

Eq. 11

We can substitute y^* and x^* in Eq. 7, which gives:

$$(y_{ij} - \gamma_y q_{ij}) = \beta_0 + \beta_1(x_{ij} - \gamma_x q_{ij}) + \sum_{k=2}^{50} \beta_k D_{jk} + e_{ij} \qquad \text{Eq. 12}$$

This equation can be rearranged to obtain:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sum_{k=2}^{50} \beta_k D_{jk} + (e_{ij} - \beta_1 \gamma_x q_{ij} + \gamma_y q_{ij}) \qquad \text{Eq. 13}$$

Thus, omitted causes (expanded and rearranged error term) correlates with x; in this case the bias may cause *inflated* or *attenuated* estimates (see lab example).

Solving the problem with 2SLS:

I have used the procedure to solve it in a couple of recent papers (Antonakis, Fenley, & Liechti, 2011; Lee & Antonakis, in press; Lee, Stettler, & Antonakis, 2011).

The procedure is simple: Find "instruments" (i.e., exogenous sources of variance) to purge the error term of endogeneity bias. Instruments can be found in interesting places…..one just has to look!

Simulation:

I generated this data for a sample size of n = 10,000, with q being the common method variance. Assume the true model that generated the data is (note that e and u are normally distributed and independent of each other, m and n are exogenous instruments):

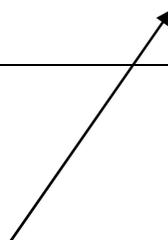$$x = \alpha_0 + q + .7m + .7n + e \qquad \text{Eq. S1}$$

$$y = \beta_0 + q - .3x + u \qquad \text{Eq. S2}$$

Table S1: Correlation matrix for 2SLS demonstration

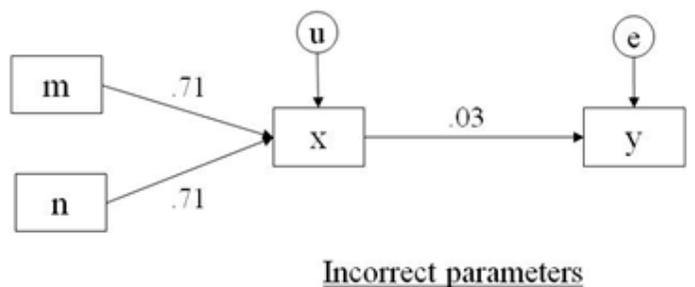| Variable	Mean	SD	q	m	n	x	y
q	0.00	1.00	1.00				
m	0.00	1.01	0.00	1.00			
n	0.01	1.00	0.01	-0.01	1.00		
x	0.01	1.73	0.58	0.41	0.41	1.00	
y	-0.01	1.30	0.54	-0.16	-0.16	0.04	1.00

N=10,000.

Notice the observed correlation!!

Data available at: http://www.hec.unil.ch/jantonakis/2slsdata.xls
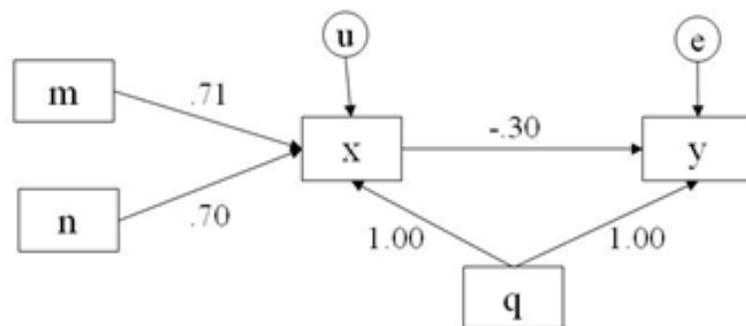
Graphically (unstandardized estimates):

True model parameters

Incorrect parameters
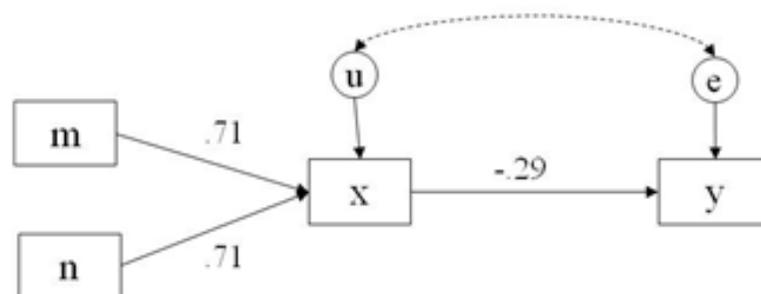

Incorrect parameters
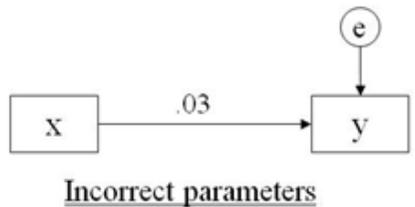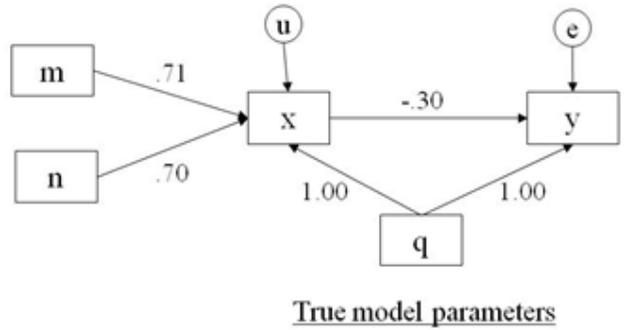
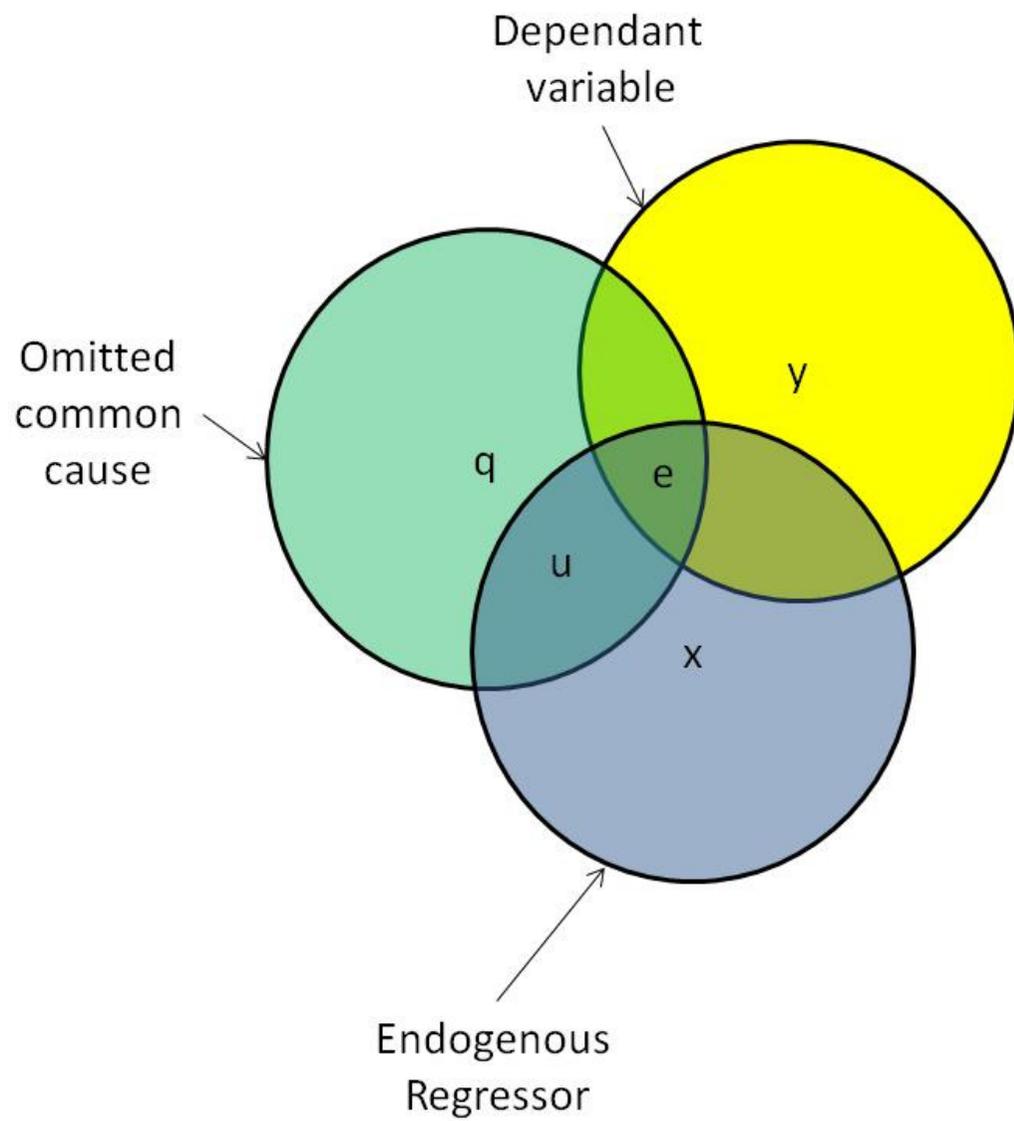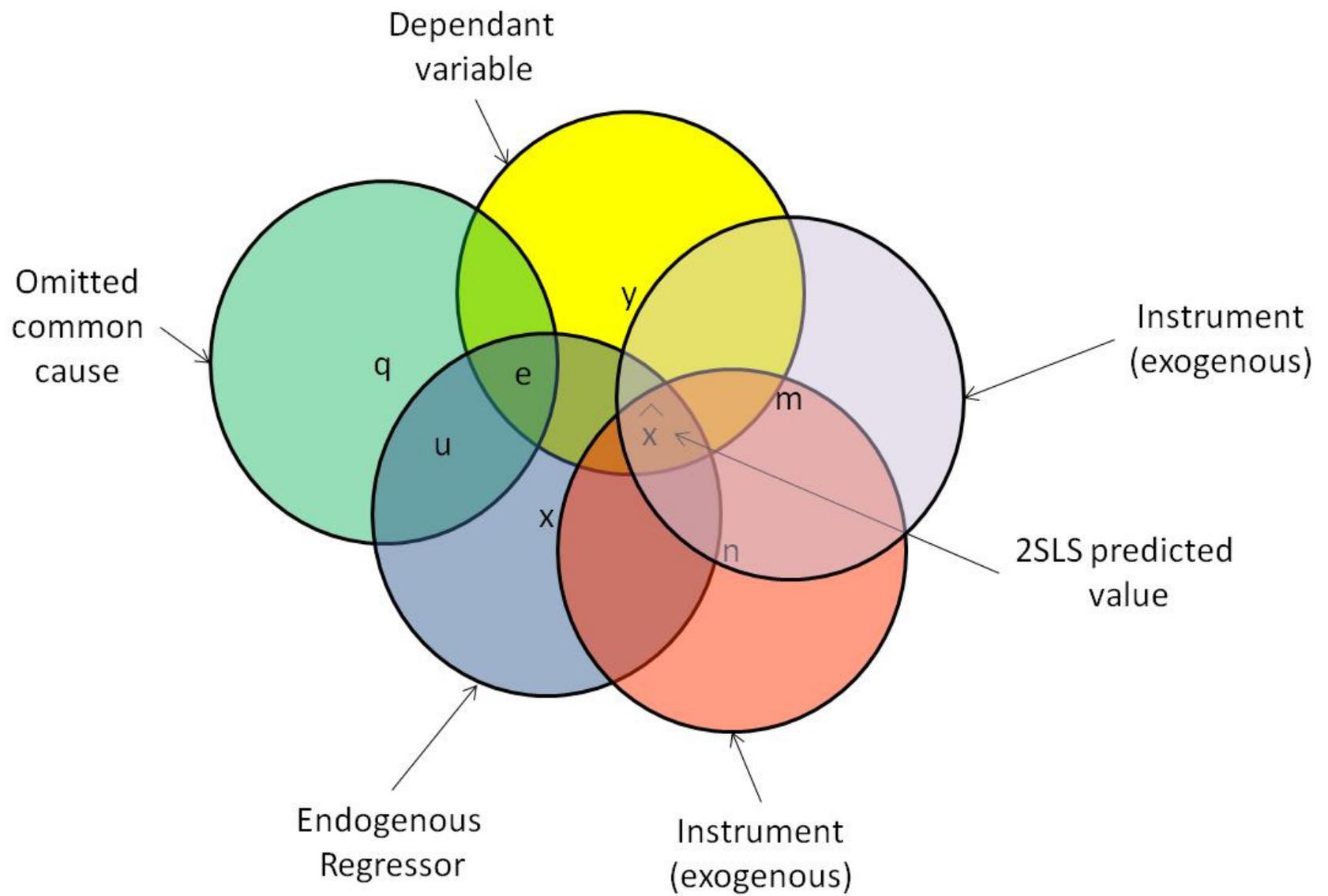(à la Baron & Kenny, 1986)

Correct parameters with 2SLS estimation:

True model parameters

Correct model parameters using 2SLS

Let's see this problem and solution graphically:

Dependant variable

Omitted common cause

Instrument (exogenous)

2SLS predicted value

Endogenous Regressor

Instrument (exogenous)

q

e

y

m

u

x̂

x

n

Research is rife with endogeneity

- psychology (Foster & McLanahan, 1996)

- management (Shaver, 1998)

- accounting (Larcker & Rusticus, 2010)

- in strategy about 90% of papers published in the premier strategy

  journal *Strategic Management Journal* (SMJ), did not account for

  endogeneity (Hamilton & Nickerson, 2003).

Now……this is on the radar of journal editors (just like in economics)!

Our LQ paper (Antonakis, et al., 2010)

- Described different cases of endogeneity and how to deal with it, drawing particularly from econometrics (and other fields).

- Reviewed (and "graded") a sample of papers ($n = 110$, between 1999-2008).

  o The population of journals are all top-tier journals according to objective criteria (i.e., 5-year ISI impact factor) publishing leadership research in management or applied psychology.

Criteria coded (14 criteria):

1. Omitted variables (4 criteria)

2. Omitted selection (3 criteria)

3. Simultaneity (1 criterion)

4. Measurement error (1 criterion)

5. Common-method variance (1 criterion)

6. Consistency of inference (2 criteria)

7. Misspecification (2 criteria)

Results
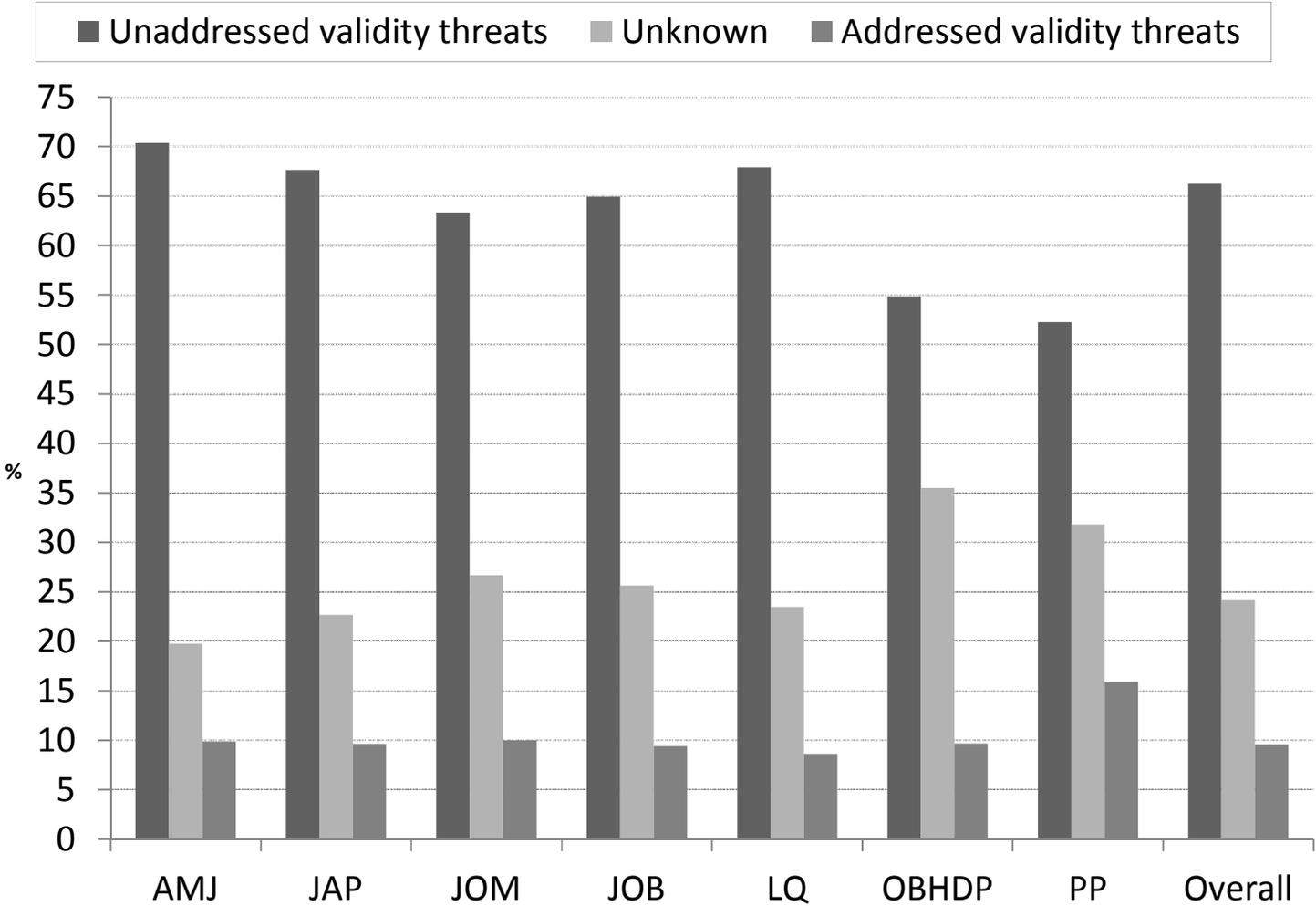
Summary statistics on criteria

	Relevancy percentage (100%-% of 0)	% of 1 (relevant not corrected)	% of 2 (relevant, not known if corrected)	% of 3 (relevant, corrected)
Omitted regressors	80.9	100.0	0.0	0.0
Omitted fixed-effects	80.0	80.7	5.7	13.6
Random-effects (without justification)	13.6	93.3	6.7	0.0
x's not exogenous	79.1	96.6	0.0	3.4
Selection: groups different at start	2.7	66.7	0.0	33.3
Selection is endogenous	4.5	100.0	0.0	0.0
Self-selection	58.2	48.4	18.8	32.8
Simulaneity	74.5	98.8	0.0	1.2
Measurement error	94.5	70.2	13.5	16.3
Common-methods variance	83.6	77.2	2.2	20.7
Robust SE's	94.5	1.0	92.3	6.7
Cluster-robust SE's	78.2	5.8	91.9	2.3
Correlating disturbances/Hausman test	31.8	100.0	0.0	0.0
Using MLE (not checking if consistent)	10.0	100.0	0.0	0.0

Discussion

- Methodological practices regarding causal modeling in the domain of leadership are unsatisfactory.

- Except for *The Leadership Quarterly*, the articles we coded were published in general management and organizational behavior (applied psychology) journals.

   o Practices of others disciplines publishing in those journals should be very similar to the practices we identified

Why is current practice not where it should be?

- Doctoral training; in psychology at least, it appears that adequate training in field research and quantitative methods in general is not provided, even at elite universities (Aiken, West, & Millsap, 2008). Ditto in management.

- Users of statistical programs (like SPSS/AMOS) have been very slow to adopt software that can do the job correctly when causal analysis in non-experimental settings is concerned.

To download readings:

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (submitted). Causality

and endogeneity: Problems and solutions. In D.V. Day (Ed.), The Oxford

Handbook of Leadership and Organizations.

http://www.hec.unil.ch/jantonakis/Causality_and_endogeneity_final.pdf

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal

claims: A review and recommendations. The Leadership Quarterly, 21(6).

1086-1120.

http://www.hec.unil.ch/jantonakis/Causal_Claims.pdf

References:

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology - Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist, 63*(1), 32-50.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086-1120.

Antonakis, J., Fenley, M., & Liechti, S. (2011). Can charisma can be taught? Tests of Two Interventions. *Academy of Management Learning & Education, 10*(3), 374-396.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Foster, E. M., & McLanahan, S. (1996). An Illustration of the Use of Instrumental Variables: Do neighborhood conditions affect a young person's change of finishing high school? *Psychological Methods, 1*(3), 249-260.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization, 1*(1), 51-78.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-960.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley-Interscience.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3 ed.). New York: Holt, Rinehart and Winston.

Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics, 49*(3), 186-205.

Lee, Y. T., & Antonakis, J. (in press). When Preference Is Not Satisfied but the Individual Is: How Power Distance Affects Person-Job Fit *Journal of Management*.

Lee, Y. T., Stettler, A., & Antonakis, J. (2011). Incremental Validity and Indirect effect of Ethical Development on Work Performance. *Personality and Individual Differences, 50*(7), 1110-1115.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology, 89*(5), 879-903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology, 63*(1), 539-569.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science, 44*(4), 571-585.