

Endogeneity: A Review and Agenda for the Methodology-Practice Divide Affecting Micro and Macro Research

Aaron D. Hill 

University of Florida

Scott G. Johnson

Iowa State University

Lindsey M. Greco

Oklahoma State University

Ernest H. O’Boyle 

Sheryl L. Walter

University of Indiana

An expanding number of methodological resources, reviews, and commentaries both highlight endogeneity as a threat to causal claims in management research and note that practices for addressing endogeneity in empirical work frequently diverge from the recommendations of the methodological literature. We aim to bridge this divergence, helping both macro and micro researchers understand fundamental endogeneity concepts by: (1) defining a typology of four distinct causes of endogeneity, (2) summarizing endogeneity causes and methods used in management research, (3) organizing the expansive methodological literature by matching the various methods to address endogeneity to the appropriate resources, and (4) setting an agenda for future scholarship by recommending practices for researchers and gatekeepers about identifying, discussing, and reporting evidence related to endogeneity. The resulting review builds literacy about endogeneity and ways to address it so that scholars and reviewers can better produce and evaluate research. It also facilitates communication about the topic so that both

Acknowledgments: The authors would like to thank Associate Editor Karen Schnatterly and two anonymous reviewers for their constructive feedback and guidance during the revision process.

Supplemental material for this article is available with the manuscript on the JOM website.

Corresponding author: Aaron D. Hill, University of Florida, Warrington College of Business, Stuzin Hall 219C, Gainesville, FL 32611, USA.

E-mail: aaron.hill@warrington.ufl.edu

micro- and macro-oriented researchers can understand, evaluate, and implement methods across disciplines.

Keywords: *endogeneity; research methods; research design; Heckman modeling (2-stage); microeconomic analysis of panel data; sample selection (Heckman-type) models; causality; instrumental variables; omitted variables bias*

If models are misspecified, causal variables or paths omitted, or there is systematic error in measures of focal constructs, model estimates will be biased; together, these effects have been grouped under the broad heading of endogeneity (e.g., Bascle, 2008; Bergh et al., 2016; Bliese, Schepker, Essman, & Ployart, 2020; Hamilton & Nickerson, 2003; Semadeni, Withers, & Certo, 2014). While more prominently used in macro research, this term has recently gained traction in the micro literature (e.g., Antonakis, Bendahan, Jacquart, & Lalive, 2010; MacKinnon & Pirlott, 2015) to label similar concerns discussed with different terminology (e.g., control variables, common method variance [CMV]). Formally defined, endogeneity occurs when a predictor (independent variable, explanatory variable, regressor) correlates with the unexplained residual (disturbance, error term) of the outcome (dependent variable) in a predictive model (for clarity, we utilize the terms predictor, residual, and outcome in the paper).¹ What makes endogeneity particularly pernicious is that the bias cannot be predicted with methods alone and the coefficients are just as likely to be overestimated as underestimated. As such, endogeneity is often noted as one of the greatest threats to management researchers' ability to correctly specify models and make causal claims (e.g., Antonakis, Bendahan, Jacquart, & Lalive, 2014; Certo, Busenbark, Woo, & Semadeni, 2016; Clougherty, Duso, & Muck, 2016; Shaver, 1998; Wolfolds & Siegel, 2019).

Such a dire threat to the veracity of research claims warrants serious attention, and papers increasingly discuss and attempt to address endogeneity concerns with a combination of research design, theoretical logic, and statistical analysis. Further, some journals now explicitly instruct authors to address endogeneity, and concerns over the issue are a "frequent reason for manuscript rejection" (Semadeni et al., 2014: 1070). If one reads the resulting empirical papers, it may seem the increased awareness means that the endogeneity problem is in hand: It is widely seen as problematic, institutional safeguards have been or are being implemented, and researchers are attempting to address the problem with design, theory, and analysis. As such, researchers often claim endogeneity does not affect their results and/or that any problems have been mitigated.

At the same time, a series of reviews, commentaries, and best practices papers question how well endogeneity concerns are addressed in published work. For example, Wolfolds and Siegel (2019) found only 33% of articles they reviewed in top management journals correctly use Heckman's method to address endogeneity; Antonakis et al. (2010; 2014), Certo et al. (2016), and Clougherty et al. (2016) document problems in applying and explaining other methods used to address endogeneity; and Semadeni et al. (2014: 1071) point to "alarming inconsistencies" in approaches and remedies to endogeneity. Our question is: Why is there such a disconnect between practices for addressing and claims about the effects of

endogeneity in empirical work compared to the methodological and technical literature on endogeneity?

An analogy may help explain the situation. Endogeneity is a disease (problem) that infects an unknown portion of empirical studies in management. Various medicines (methods) can treat the disease, so with increasing concern about the disease, more people are using the medicines either to treat a known problem or reduce concerns about the disease. Medical experts (i.e., methodologists) have studied how various medicines are being used and have found they are often administered incorrectly (e.g., Antonakis et al., 2010; Certo et al., 2016; Semadeni et al., 2014; Wolfolds & Siegel, 2019). As a result, many of the papers thought to be “cured” are in fact not cured due to inappropriate use of methods to treat endogeneity. Future research may then build on the “uncured” paper by applying the method in a similar fashion, resulting not only in the accumulation of biased findings but also in the spread of the incorrect practice. We believe that, despite the evidence of methods often being misapplied, there is already sufficient information available about how to correctly use the medicines; thus, our review is not intended to be yet another tutorial on dosages and prescriptions. Rather, in the next four sections, we seek to address a more fundamental question, which is how researchers can talk about the endogeneity disease in a way that allows others to understand whether the disease is present, what the specific strain of the disease is, what ways it can be treated, and what the most appropriate prognosis is.

The first section is devoted to understanding endogeneity as an important threat to valid research conclusions that arises from four distinct causes. Micro and macro researchers often use different terms to refer to the causes and associated concerns (Bliese et al., 2020). The lack of consistent and specific terminology presents a hurdle for understanding, not just for those new to the topic but also for more experienced scholars who still struggle to match specific endogeneity sources with appropriate methodological and statistical remedies (e.g., Certo et al., 2016; Clougherty et al., 2016; Semadeni et al., 2014). More specifically, using different terms both hinders communication about endogeneity, especially between micro and macro domains, and obscures links to methodological papers offering strategies to address various endogeneity issues. Thus, the first section of our review is intended to help researchers better understand endogeneity and present it in a way that improves communication between micro and macro researchers. We draw on Wooldridge (2010), who separates four causes of endogeneity (omitted variable, simultaneity, measurement error, and selection) to argue that endogeneity is really four different strains of the overarching problem, or disease, that each bias results in different ways.

The second section reviews how scholars discuss endogeneity. Building on the typology presented in the first section, we reviewed 435 papers in top management journals that discuss endogeneity trying to answer two questions: (1) Do the authors simply acknowledge endogeneity as a threat, or do they address it through design, analysis, or post hoc robustness tests? And (2) what specific cause of endogeneity are the authors concerned about and what method do they use if they address it? We diverge from prior reviews (e.g., Antonakis et al., 2010; Certo et al., 2016; Wolfolds & Siegel, 2019) that focused on whether particular methods are used correctly but which do not speak to the larger, holistic issue of how empirical researchers are interpreting and applying their understanding of endogeneity in their theorizing, design, and analysis. Our review shows that, independent of methods being used incorrectly, many studies are unclear in discussing endogeneity. For example, we find studies

often use a method to address endogeneity without explaining why endogeneity is a concern, justifying how the method used addresses that concern, or providing adequate statistical evidence that supports the use of the approach.

In the third section, we again use our typology to review the methodological literature and organize it into a map of causes and associated solutions. It is not our intent to provide detailed coverage of every method. Instead, we describe the reasoning behind each cause of endogeneity, explain various associated methods used to address each cause of endogeneity, and point to the appropriate sources for more detailed information. To do so, we identified and reviewed over 250 methodological sources and over 40 review articles about endogeneity. We organize key articles to help future scholars find methodological resources relevant to the specific cause in their research.² This section is intended to help researchers identify potential causes of endogeneity, determine appropriate remedies, and apply them correctly.

Our fourth section sets an agenda for future scholarship by offering advice to researchers and gatekeepers about how to identify, discuss, and report evidence related to endogeneity.

The Problem of Endogeneity and Why It Matters

Researchers are often interested in causal questions. Perhaps the clearest way to establish causality is with an ideal randomized trial where the causal effect of x (a predictor variable) on y (an outcome variable) is isolated through random assignment. That is, random assignment to different levels of the predictor variable x ensures that, with adequate sample sizes and when idealized conditions are met, exposure to the experimental effect (among those in the study) is uncorrelated with omitted factors (for a review, see Krause & Howard, 2003). Randomized trials are not without problems, but in principle, they can counteract various causes of endogeneity (as we detail below). Yet randomized trials are not always feasible or desirable, so researchers often use alternatives like archival data, quasi-experiments, or survey data where random assignment is not possible. In analyzing these kinds of data, the question is whether we can view an estimated coefficient as approximating the causal effect that might be determined in an ideal experiment. For causal inferences to be valid, assumptions of the analytical approach (e.g., ordinary least squares [OLS] regression, structural equation modeling [SEM]) must be met. Of concern here is the exogeneity assumption (i.e., no endogeneity)—that is, the residual in the model has an expected value of zero given any instance of the predictor variable so that there is no correlation between a predictor variable and residual (Wooldridge, 2010).

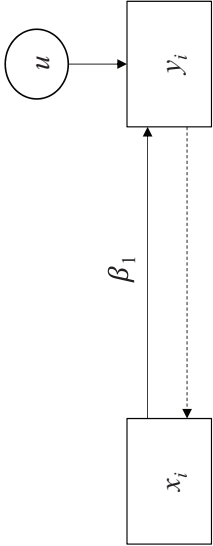
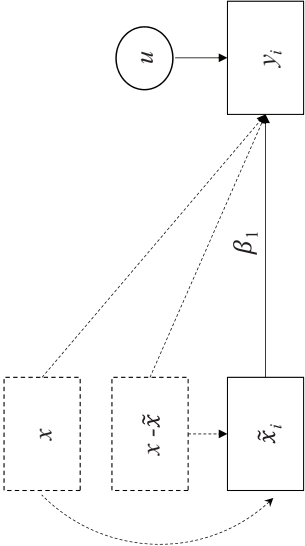
To aid presentation, we use an example equation with one outcome y and one predictor x where a is a constant (intercept) in the model, B is the estimated coefficient, and u is the residual: $y = a + Bx + u$. Endogeneity is formally defined as the observed predictor x correlating with the unobserved residual u (see Figure 1 embedded in Table 1). Wooldridge (2010) calls addressing u the most important component of any analysis because u contains myriad unobservable factors that can affect y . The difficulty in capturing and defining this relationship is that understanding u is inherently a theoretical exercise since u is defined by all the information *not* captured by x . Much of this review focuses on the many ways to address the specific forms of endogeneity, but we emphasize that defining the source of endogeneity and determining an appropriate remedy must always be accompanied by a theoretical rationale of variables or events that would cause x to be related to u . When researchers

Table 1
Depictions and Descriptions of Endogeneity

Figure	Description	Similar/Related Terms*
<p>Figure 1 – Endogeneity (General)</p> <p>$cov(x_i, u) \neq 0$</p>	<p>The residual u arises because factors that influence y are not included in the regression function. When the predictor, x, correlates with the residual, u, this causes bias in the estimate of β. The bias can be upward or downward depending on the unmodeled factors that correlate with x and predict y.</p>	
<p>Figure 2 – Omitted Variable</p>	<p>The omitted variable q affects y, so when q is not modelled it is included in residual u. When q is also correlated with x, then x is correlated with the residual u.</p>	<p>Omitted variable bias Missing variable Unmeasured variable problem Left out variable error Unobserved heterogeneity Confounding variable</p>

(continued)

Table 1. (continued)

Figure	Description	Similar/Related Terms*
<p>Figure 3 – Simultaneity</p> 	<p>In addition to x causing y, y also causes x. Because y causes x, both y and u will correlate with x. Time separation does not solve the problem unless u has zero autocorrelation (i.e., no correlation with itself over different time periods). x at time 0 is not caused by u at Time 1 because it cannot cause something in the past; however, u at Time 1 is frequently correlated with u at Time 0 and, as noted above, x at Time 0 is correlated with u at Time 0.</p>	<p>Reciprocal Reverse causality Feedback loop Joint determination Interdependence</p>
<p>Figure 4 – Measurement Error</p> 	<p>With measurement error, observe \tilde{x} instead of x. When the measurement error in x affects y, it becomes part of u, which correlates with \tilde{x}.</p>	<p>Errors-in-variables Observational error CMV (special case)</p>

(continued)

Table 1. (continued)

Figure	Description	Similar/Related Terms*
<p>Figure 5 – Selection: Selection into Sample</p> <p>The model captures the relationship between x and the observed value of y (y^*). y^* is affected by a selection process, s, that restricts the range of y^* by excluding observations from the sample. s is influenced by y and by other unobserved causes, w. The processes that affect y^* are part of the residual u. The value of x is correlated to u if x is related to y.</p>	Sample selection or bias Self-selection Range restriction	
<p>Figure 6 – Selection: Selection of Treatment</p> <p>The level of x is affected by w through selection, s. The selection process, s, is an indicator of participation—indicating the degree to which a particular entity (individual, firm) in the sample is exposed to treatment. If w is related to y, then it becomes part of u so x is correlated with u.</p>	Endogenous choice Treatment effect Self-selection	

Note: Solid lines represent modeled paths, dotted lines represent unmodeled paths. While we present just a single x and y variable, the subscript i denotes that the same considerations hold for multiple predictors x_1, x_2, \dots, x_i and outcomes y_1, y_2, \dots, y_i .

*See Dodge (2006) for a full glossary of similar and related terms.

cannot use random assignment in an ideal format to dismiss alternate explanations, they must present both logical or theoretical support alongside empirical support for the contention that x is unrelated (i.e., orthogonal or exogenous) to u .

Our typology draws from Wooldridge (2010), who separates the causes of endogeneity into four categories: omitted variable, simultaneity, measurement error, and selection. Table 1 depicts and describes these causes as well as links each to synonymous terms as a means of aiding communication about specific endogeneity issues. Figure 1 depicts the broad definition of endogeneity noted above— x correlating with the residual u . The conditions that give rise to this correlation are summarized by the four causes. The first cause of endogeneity is an *omitted variable*. Most studies have omitted variables, but bias is created when a variable not included in the model is related to both x and y . Figure 2 depicts an omitted variable q , which relates to both x and y . The second cause of endogeneity is *simultaneity*. The estimate of how x affects y is biased if y also affects x , as the omitted path from y to x in Figure 3 depicts. The third cause of endogeneity is *measurement error*. Bias is created if any error in measuring x , resulting in \tilde{x} rather than x , is correlated with y (Figure 4). The final cause of endogeneity is *selection*, which is divided into two subtypes. Bias can arise if selection into a sample is not random (which then affects the observation of the outcome denoted as y^* ; Figure 5) or if selection of treatment x is not randomly assigned (which then affects estimates of x on y ; Figure 6).

To further explain and align the four categories across micro and macro research, Table 2 lists the causes of endogeneity as well as examples of thought experiments that, if asked, may help identify whether or not a cause of endogeneity is present in a given study (we provide both micro and macro examples). To illustrate, omitted variable endogeneity may also be discussed as a missing or confounding variable, among other terms (see Table 1). When thinking about whether omitted variable endogeneity may affect results, one can conduct a thought experiment, asking what other predictors or constructs might be included in the residual u and also relate to the focal predictor x . A micro example of such an issue is if one proposes that job satisfaction leads to higher job performance without accounting for negative affect. Negative affect is likely correlated with both variables; when it is not included as a variable in the model, the effects will be captured by the residual u for job performance, which will then correlate with job satisfaction. A macro example is if one proposes that advertising intensity leads to higher sales without accounting for firms' industry. Again, the correlation of firms' industry with the predictor and residual is a source of endogeneity. This four-part typology of causes (omitted variable, simultaneity, measurement error, selection) is used to organize the next two sections.

How Scholars Discuss Endogeneity

To document how management research identifies, describes, and addresses endogeneity, we reviewed empirical articles across micro and macro domains. We first identified and coded articles in top tier journals in both broad-based (i.e., *Academy of Management Journal*, *Administrative Science Quarterly*, and *Journal of Management*) and specific domains (i.e., *Journal of Applied Psychology*, *Strategic Management Journal*) with the keywords endogeneity and endogenous over the 5 years preceding submission (2014–2018). The sole inclusion criterion was that the article discussed endogeneity in the context of designing or

Table 2
The Four Causes of Endogeneity

Endogeneity Cause	Thought Experiment(s)	Micro Example	Macro Example
Cause 1: Omitted Variable	What other predictors or constructs could be included in the residual? Are any of these also likely to be correlated with the predictor variable?	X = Job satisfaction Y = Job performance Negative affect is likely correlated with both variables.	X = Advertising intensity Y = Sales The industry a firm is in is likely correlated with both variables.
Cause 2: Simultaneity	Are there any feedback loops that connect the predictor variable and outcome variable? That is, is the relationship reciprocal?	X = Alcohol consumption Y = Job status Alcohol consumption may affect, and be affected by, job status.	X = R&D spending Y = Firm performance R&D spending may affect, and be affected, by firm performance.
Cause 3: Measurement Error	Is there any systematic error in either the predictor variable or outcome variable? Might this systematic error be correlated with the other variable (i.e., predictor with outcome)?	X = Job satisfaction Y = Job Performance If both variables are rated by the same individual at the same time, they may correlate highly.	X = Firm reputation Y = Stock price A survey on firm reputation might systematically overrate firms with a high stock price.
Cause 4: Selection (of treatment and/or into sample)	What attributes of the unit of analysis or environment it is in might "select" the level of the predictor or outcome variable (selection of treatment) or whether data exist for testing (selection into sample)? Are any of these attributes likely to be correlated with the outcome variable?	X = Person-job fit Y = Job performance Individuals will seek out jobs that are a good fit (selection of treatment) and with bad fit may have quit before job performance is measured (selection into sample)	X = Acquisition Y = Stock appreciation Firms that acquire might be in a stronger competitive position than firms that do not acquire (selection of treatment), but we can only gather data on firms where reports of acquisitions can be gathered (selection into sample).

Note: X refers to a predictor (regressor, independent) variable and Y to an outcome (dependent) variable.

interpreting a study. This excluded editorials, review articles, meta-analyses, and a large number of articles—mostly at the micro level—discussing “endogenous” latent variables in a SEM context without reference to endogeneity as a validity threat. This rendered a sample that included 435 articles.

Table 3
How Sampled Articles Discuss Endogeneity

Total	AMJ 85		ASQ 24		JAP 13		JOM 64		SMJ 249		Total 435	
Coding of all sampled articles	#	%	#	%	#	%	#	%	#	%	#	%
Acknowledged	6	(7.1)	2	(8.3)	1	(7.7)	9	(14.1)	33	(13.3)	51	(11.7)
Addressed	48	(56.5)	16	(66.7)	8	(61.5)	41	(64.1)	140	(56.2)	253	(58.2)
<i>Of addressed, results affected?</i>												
• Yes	1	(2.1)	2	(12.5)	3	(37.5)	5	(12.2)	10	(7.1)	21	(8.3)
• No	2	(4.2)	3	(18.8)	1	(12.5)	16	(39.0)	14	(10.0)	36	(14.2)
• Not stated	45	(93.8)	11	(68.8)	4	(50.0)	20	(48.8)	116	(82.9)	196	(77.5)
Robustness	31	(36.5)	6	(25.0)	4	(30.8)	14	(21.9)	76	(30.5)	131	(30.1)
<i>Of robustness, results affected?</i>												
• Yes	1	(3.2)	0	—	0	—	1	(7.1)	0	—	2	(1.5)
• No	24	(77.4)	6	(100.0)	4	(100.0)	12	(85.7)	76	(100.0)	122	(93.1)
• Not stated	6	(19.4)	0	—	0	—	1	(7.1)	0	—	7	(5.3)
Coding of addressed or robustness only												
Used an instrumental variable method	51	(60.0)	9	(37.5)	4	(30.8)	29	(45.3)	117	(47.0)	210	(48.3)
• Explained instrumental variable(s)	40	(78.4)	6	(66.7)	3	(75.0)	17	(58.6)	84	(71.8)	150	(71.4)

Note: AMJ = *Academy of Management Journal*; ASQ = *Administrative Science Quarterly*; JAP = *Journal of Applied Psychology*; JOM = *Journal of Management*; SMJ = *Strategic Management Journal*.

After identifying articles, we first coded if endogeneity was (1) *acknowledged* as a viable concern (usually as a study limitation), (2) *addressed* with design or analysis in the main analysis, or (3) tested through post hoc *robustness* tests. For articles that addressed endogeneity or reported robustness tests, we also coded whether (a) the results were affected once endogeneity was accounted for, (b) an instrumental variable method was used, and (c) the instrumental variables were explained (summarized in Table 3). Instrumental variables, often abbreviated as simply instrument(s), are exogenous variable(s) introduced in different analytical models to address certain types of endogeneity concerns (Semadeni et al., 2014). We elaborate on this definition, applicable methodologies, and assumptions related to instrumental variables later in the paper.

Our first finding echoes past reviews that there seems to be a disconnect in the literature: Empirical papers largely present endogeneity as an issue that has either been addressed or has not affected results. Of the 435 articles mentioning endogeneity, most ($k = 253$; 58.2%) use a method to address suspected concerns in the main analysis, while a smaller portion ($k = 131$; 30.1%) use robustness tests (i.e., report results and then assess if the results may be biased by endogeneity). Based on the robustness tests, only 1.5% ($k = 2$) of the articles offer evidence the results differ from the primary analysis after addressing endogeneity. Taken together, we find that published papers often present endogeneity as an issue that is important enough to be addressed as a methodological concern, but once addressed, the results tend to remain unchanged.

To ensure the representativeness of our sample and categorization, we performed two robustness checks. First, our focus is on both micro and macro research, but the initial search returned a disproportionate number of articles appearing in a macro-focused journal: *Strategic*

Management Journal. As such, to achieve a representative sample of articles for the other journals, we extended our search back to 1998 to coincide with Shaver's (1998) seminal article on endogeneity. We coded an additional 140 articles from our sampled journals but did not identify any unique approaches for addressing endogeneity used more than twice, suggesting our sample is representative of the approaches used in the field. Second, we recognize that articles may address specific causes of endogeneity but not use the term *endogeneity*, particularly in micro journals where we had the fewest number of studies ($k = 13$ in *Journal of Applied Psychology*). To confirm that our findings were applicable to both micro and macro research, we repeated our search using all of the similar and related terms listed in Table 1 (rather than endogenous/endogeneity) across all sampled journals in the years 2014–2018. We coded the largest subgroup of articles, which was from *Journal of Applied Psychology*, consisting of 636 different uses of these terms in 351 articles. After removing terms inconsistent with our use of the concept of endogeneity (e.g., use of interdependence to describe team characteristics rather than a methodological concern) and nonempirical work, the final sample included 251 unique uses of these terms in 141 articles (an average of 1.78 terms per paper). In our additional coding, 75 (30%) acknowledged endogeneity bias could affect results, 156 (62%) addressed endogeneity through design or analysis, and 20 (8%) used robustness tests for endogeneity bias.³ Taken together, these additional searches resulted in more articles, but not more solutions or unique approaches to addressing endogeneity, which increases confidence in our sample coding.

Contrary to the evidence in these published papers, multiple reviews (e.g., Antonakis et al., 2010; Certo et al., 2016; Clougherty et al., 2016; Semadeni et al., 2014) note how statistical techniques are often misapplied and/or not adequately explained and justified. We assessed whether papers using an instrumental variable method (the most commonly used approach; $k = 210$, 48.3%) adequately justified and explained how the approach dealt with a given endogeneity threat. We view it as good news that a majority of the articles did offer some explanation of the variables ($k = 150$, 71.4%), yet similar to findings in prior reviews, in many cases, explanations lacked enough information to determine the exact cause of endogeneity or the theoretical rationale behind the instrumental variable(s). So even though published papers often state that endogeneity is not a concern, vague explanations of analytical techniques and choices hinder readers' ability to assess those claims. These issues give rise to two significant problems. First, if the only published papers are those reporting endogeneity tests when results are unaffected, a rose-tinted literature skewed by publication bias may result. Second, misconceptions about endogeneity and unclear reporting standards may encourage particularly weak tests, creating the endogeneity equivalent of *p*-hacking where, for example, statistical techniques are applied not based on their appropriateness or quality but on their proclivity to leave results unchanged.

To further explore how well researchers are explaining specific endogeneity threats, we coded the 384 articles that addressed endogeneity ($k = 253$) or used robustness tests ($k = 131$) based on our typology: omitted variable(s), simultaneity, measurement error, and selection. In 26 articles, more than one cause of potential endogeneity was noted, so we coded each cause and method. Thus, the total number of endogeneity causes and methods, summarized in Table 4, is 411. The coding revealed two insights. First, of those articles identifying a specific cause of endogeneity, selection is most often noted ($k = 132$, 32.1%). This may not be surprising since several authors (e.g., Hamilton & Nickerson, 2003; Shaver, 1998)

Table 4
Summary of Identified Endogeneity Sources and Methods

	Omitted Variable	Simultaneity	Measurement Error	Selection	Unclear	Total
Experiment	2	2	0	4	2	10
Quasi-experiment	4	1	0	5	1	11
Design choices	2	2	0	1	3	8
Matching sample	4	2	0	10	4	20
Measurement	0	1	5	0	0	6
Control variables	10	1	0	3	5	19
Panel	11	11	1	2	13	38
Instrumental variable	41	42	0	94	48	225
Dynamic panel	6	16	0	3	12	37
Other	6	7	0	10	14	37
Total	86 (20.9%)	85 (20.7%)	6 (1.5%)	132 (32.1%)	102 (24.8%)	411

argue selection is a constant threat as it is hard to disentangle the outcomes of decisions from the drivers of those decisions. Second, the next largest subset of articles ($k = 102$; 24.8%) do not clearly identify a source of endogeneity but instead discussed it in more generic terms. These articles generally made broad statements along the lines of “our results may be affected by endogeneity” without specifying a concern over omitted variable(s), simultaneity, measurement error, and/or selection. This ambiguity regarding the cause(s) of endogeneity impairs readers’ ability to assess if the approach used in the study alleviated endogeneity or exasperated the problem (Semadeni et al., 2014).

Beyond our formal coding, we noted some practices that were hard to quantify but worth mentioning. First, we found more recent articles often discuss methodological details in an online appendix and/or note unpublished results are available from the author(s). We applaud this rigor but stress that addressing endogeneity is a meaningful part of the research and should not be seen as an exercise tangential to the main analysis. If addressing endogeneity is only seen as tangential, it becomes harder to accumulate knowledge with multiple studies as is advocated (e.g., Shaver, 2019; Wolfolds & Siegel, 2019); thus, we argue that addressing endogeneity must be seen as a meaningful part of the main research design and analysis. Second, we found many articles that cite prior empirical work to justify a choice of methodology. Methodological work can be daunting to wade into, and it is understandable to cite approaches and rationales of prior work published in the journals in which authors aspire to publish. This practice of citing prior empirical work to justify a method presents two challenges, however. One is that it is difficult to determine the exact percentage of papers that are appropriate or not, as it can be hard to follow the chain of citations. Another is that since multiple papers both identify methodological mistakes in prior empirical work and note that methodologies are frequently inadequately explained and justified (e.g., Antonakis et al., 2010; Certo et al., 2016; Semadeni et al., 2014), the practice of citing prior empirical work rather than then methodological source papers sometimes leads to a “telephone problem” where approaches are not justified in following a prior empirical paper and/or errors in prior empirical papers are repeated and magnified over time.

In addition to reviewing empirical papers, we also identified and reviewed over 250 methodological sources and 40 review articles offering advice on endogeneity. As with our coding of the empirical papers, this “review of the reviews” yielded insights that were hard to quantify but which return us to the fundamental divide in the literature: Why have efforts to address endogeneity been so problematic despite the increasing number of methodological and review articles published on the topic (many in our top journals)? Our tentative answer to that question builds on our observation of a “telephone problem” where some papers adopt practices from prior empirical research rather than recommendations from methodological articles. If one draws conclusions about endogeneity from empirical research only rather than from methodological research, it might be said that (1) endogeneity is a broad and ambiguous issue often encountered but rarely explained in detail—especially if prior empirical work can serve as precedent; (2) the ways endogeneity is addressed or tested for are too technical to be explained adequately in empirical work; and/or (3) while endogeneity is often discussed, it almost never affects the results. We emphasize that we do not think these are useful conclusions to draw, so we now turn to a review of the methodology literature as a way to clarify terminology, point researchers to good methodological advice, provide some recommendations for improving both research, and perhaps more importantly, communication about research practices.

Endogeneity: Causes and Associated Solutions in the Literature

Much of the early concern over endogeneity in management research originated in macro domains, likely due to the prevalence of nonexperimental studies where firm actions (e.g., entry mode, acquisitions) are not randomly assigned (Shaver, 1998). Concerns over endogeneity, as an explicit concern, have since expanded to nonexperimental studies in micro domains (e.g., Antonakis et al., 2010; MacKinnon & Pirlott, 2015; Maydeu-Olivares, Shi, & Fairchild, 2020; Sajons, 2020; Schmidt & Pohler, 2018), although related methodological issues have a somewhat longer history (e.g., CMV, omitted variables, etc.; see list of related terms in Table 1). Scholars in related fields (e.g., marketing, Shugan, 2004; accounting, Larcker & Rusticus, 2010) also recognize the threat of endogeneity and offer recommendations to address it. In reviewing the literature related to endogeneity, we find that knowledge is fragmented by the use of different terms and conceptualizations of endogeneity (Bliese et al., 2020), the different fields in which it is discussed and addressed, and the fact that efforts to address issues related to endogeneity often focus on a “particular subdimension of the greater endogeneity problem” without necessarily addressing other forms (Clougherty et al., 2016: 287).

To address this fragmentation and increase the accessibility of the methodological literature, we do three things. First, we provide an annotated bibliography of primary contributions to this literature so that researchers new to the topic can acquaint themselves with the material relevant to their research domain (see Online Appendix A). Second, we organize the most used methods according to the previously discussed typology of causes. For each method, we provide links to corresponding resources that offer further details. Third, we offer linkages to related terms (see Table 1), often used as synonyms, to facilitate communication about endogeneity (which, as we note in Section 4, can help guide authors and reviewers).

Our review of empirical papers revealed that many studies refer to “endogeneity” broadly rather than to specific causes. Focusing on the specific mechanism causing endogeneity rather than a general endogeneity problem is important for two reasons. First, many of the methods to address endogeneity only apply to a specific cause of endogeneity. Second, multiple causes of endogeneity may affect the same variable in a single study. We are encouraged by the fact that 26 articles in our sample address more than one cause of endogeneity, but we do not want to signal to authors or gatekeepers that every paper must address every possible cause of endogeneity. Instead, we emphasize there is no generic way to address general endogeneity concerns, but there is an extensive toolbox of methods that can be used to address specific causes of endogeneity. Much of the confusion and misapplication of methods in the literature noted by others (e.g., Antonakis et al., 2010; Certo et al., 2016; Clougherty et al., 2016; Semadeni et al., 2014) stems from addressing an endogeneity concern without specifying the exact cause. As a result, subsequent research may precisely replicate the method for addressing endogeneity but not address the problem because the study suffers from a different cause of endogeneity.

As we discuss the causes of endogeneity, we point out how multiple causes can affect a single estimated relationship. Endogeneity may occur between many predictor and outcome variables in a single analysis, but we primarily use one micro and one macro example with one predictor and one outcome to aid exposition: how job satisfaction (x) affects job performance (y) and how firm reputation (x) affects firm performance (y). As we delineate these causes, we also highlight associated solutions to address each source of endogeneity (summarized in Table 5).

Cause 1: Omitted Variables Endogeneity

Perhaps the most intuitive cause of endogeneity is omitted variable bias (found as a cause of endogeneity 20.9% of articles reviewed; see Table 4). As Figure 2 in Table 1 shows, for an omitted variable q to bias estimates, it must affect the outcome y (resulting in the new equation: $y = a + Bx + q + u$). If we know q affects y but we do not have q in our model, the unexplained part of our model is $q + u$ rather than just u . The exogeneity requirement is that x is uncorrelated with the unexplained part of the model; thus, an endogeneity problem from an omitted variable q exists if x correlates with q and q relates to y . For example, variables related to job performance and correlated to job satisfaction might be the level of rewards a worker receives or the worker's ability. Variables related to firm performance and correlated to firm reputation might be the firm's market share and/or capability in public relations or public perceptions of the industry. If any of these variables are omitted in regressing y on x , the coefficient B will be biased.

Perhaps the most direct way to address an omitted variable problem is to include it in the study, yet researchers cannot always do so. For example, there are many variables that can potentially relate to x and y , so the process of adding additional variables (or worrying about potential endogeneity from this cause) has to stop at some point. As Frank (2000: 149) notes, “the simple question, ‘yes, but have you controlled for xxx’ puts social scientists forever in a quandary” both for design (e.g., collecting a large number of alternative variables is limited by survey length) and analysis reasons (e.g., adding variables ad nauseum causes other issues; Bernerth & Aguinis, 2016; Spector & Brannick, 2011). Another reason is measuring

Table 5
Techniques That Offer Solutions to Help Remedy Endogeneity

Technique and Description	Requirements and Limitations	Selected Resources*
Avoiding/minimizing endogeneity threats through design		
Laboratory Experiment – Randomly divide participants into experimental and control groups. Introduce a change (i.e., manipulation) to the experimental group.	Must be able to manipulate the predictor variable and randomly assign groups. This may not be feasible or ethical. Findings may lack external validity and generalizability.	Fromkin & Streufert, 1976; Griffin & Kacmar, 1991; Shadish, Cook, & Campbell, 2002
Field Experiment – Done in a natural context for participants to increase external validity. Researcher manipulates predictor variable in experimental but not control group.	Lack of random assignment increases threat of alternative explanations.	Podsakoff & Podsakoff, 2019
Natural Experiments – A naturally occurring situation that creates treatment and control groups; the cause is generally not manipulated by the researcher.	Control groups and treatment groups may differ in systematic ways other than in regard to the treatment.	Campbell & Stanley, 2015; Chatterji, Findley, Jensen, Meier, & Nielson, 2016; Grant & Wall, 2009; Greenberg & Tomlinson, 2004; Harrison & List, 2004
Quasi-experiments – Various ways to establish causality by analyzing data before and after an intervention or unexpected exogenous event.	Many of these design techniques overlap with analytical methods described below.	Shadish, Cook, & Campbell, 2002
Omitted Variable		
Control Variables – Extraneous or confounding variables that are not of central interest to the researcher are included in the analysis to address omitted variable bias. If an omitted variable is not available, a <i>proxy variable</i> can sometimes be used.	Researchers are unlikely to be aware of all relevant confounding variables. Some omitted variables may not be available or observable. Arbitrary inclusion of control variables can also create bias.	Becker, 2005; Bernerth & Aguinis, 2016; Breaugh, 2008; Frost, 1979; McCallum, 1972; Pei, Pischke, & Schwandt, 2019; Spector & Brannick, 2011
Sensitivity Analysis – Estimating the magnitude of the bias created by potential violations of exogeneity assumption by analyzing how the inclusion of control variables affects coefficient estimates.	Sensitivity analysis is only meaningful if controls meet requirements discussed above.	Frank, 2000; Oster, 2019; Pan & Frank, 2003; Peel, 2014; Xu, Frank, Maroulis, & Rosenberg, 2019
Fixed Effects – Include individual or group effects to account for unobserved heterogeneity.	Heterogeneity must be constant over time or within group. Unable to estimate effects of variables that do not change over time.	Antonakis, Bastardo, & Rönkkö, 2019; Bliese, Schepker, Essman, & Ployart, 2020; Shaver, 2019
Instrumental Variables – Two-step or simultaneous equation techniques that address bias by either replacing the endogenous variable with a predicted value or including a calculated control variable.	Instrument must (1) cause variation in the endogenous variable and (2) be only indirectly related to outcome through the endogenous variable. Weak instruments can be worse than no instruments.	Semadeni, Withers, & Certo, 2014

(continued)

Table 5. (continued)

Technique and Description	Requirements and Limitations	Selected Resources*
<u><i>Instrumental Specification Tests</i></u> – Some of the assumptions of instrumental variables can be tested. If the instrumental variables are valid, the exogeneity can be tested.	Tests of overriding restrictions and strict exogeneity are all built on the assumption that there is at least one valid instrument.	Baum, Schaffer, & Stillman, 2003; Basmann, 1960; Hansen, 1982; Hausman, 1978; Sargan, 1958; Stock, Wright, & Yogo, 2002
<u><i>Instrumental Variable Estimators</i></u> – Instrumental variable models can be estimated in various ways including two-stage least squares (2SLS), three-stage least squares (3SLS), maximum likelihood (ML), and generalized method of moments (GMM).	The various estimation techniques differ in their efficiency and robustness to various assumptions. None of these estimators reduce the need for valid and justified instruments.	Angrist & Imbens, 1995 (2SLS); Antonakis, Bendahan, Jacquart, & Lalive, 2010 (2SLS); Blundell & Bond, 2000 (GMM); Hansen, 1982 (GMM); Newey & West, 1987 (GMM); Wooldridge, 1997 (2SLS)
<u><i>Lagged Variables as Instruments</i></u> – Using lagged values of the endogenous variable as an instrument.	Lagged variable must predict the endogenous variable and not be related to the dependent variable.	Reed, 2015
<u><i>Model-Implied Instrumental Variables</i></u> – Limited-information estimator for latent variable models that relies on existing observed variables to create instruments.	The identification is not “free”; additional assumptions are needed.	Bollen, 2019; Bollen & Bauer, 2004; Gates, Fisher, & Bollen, 2019
<u><i>Exotic Techniques</i></u> – Sometimes endogeneity can be addressed with assumptions of distributional form of variables and residual.	The identifying assumptions may be more difficult to sustain than assumptions needed for traditional instruments.	Bollen, 2012; Papies, Ebbes, & Van Heerde, 2017; Sande & Ghosh, 2018
Simultaneity		
Instrumental Variables – Methods described above can also address simultaneity.	In the presence of simultaneity, instrumental variables may be even more difficult to find.	<i>See above</i>
Lagging the Endogenous Variable – Use a lagged version of the endogenous variable.	May not address endogeneity if predictor or dependent variables are serially correlated.	Fair, 1970; Bellemare, Masaki, & Pepinsky, 2017
Dynamic Panel Techniques – Estimate a model of first differences. Use lagged first differences as instruments. Sometimes called GMM or Arellano-Bond estimators.	Assumes that endogeneity is caused by time-invariant heterogeneity. Residual in first-difference equation cannot be serially correlated.	Arellano & Bond, 1991; Ballinger, 2004; Bergh, 1993; Blundell & Bond, 1998
Using Exogenous Events – Quasi-experiments where intervention or exogenous event is used to establish direction of causality.	The key identifying assumption is that the event was not anticipated.	Angrist & Krueger, 1999; Angrist & Pischke, 2010
Measurement Error		
Model Measurement Error – Use a latent variable method (SEM) to account for measurement error.	In most cases the variance of the measurement error must be known and normally distributed.	Bound, Brown, Mathiowetz, 2001; Durbin, 1954; Fornell & Larcker, 1981; Griliches & Hausman, 1986; Hausman, 1977

(continued)

Table 5. (continued)

Technique and Description	Requirements and Limitations	Selected Resources*
Instrumental Estimation – Use one variable with measurement error as an instrument for another variable with measurement error. Sometimes called indicator variable method.	The systematic errors in the two variables must be uncorrelated with each other.	Griliches, 1977
Addressing CMV – Design and statistical techniques aimed at reducing CMV, which is a source of measurement-error-induced endogeneity	Direction and strength of bias depends on data collection strategy, type of analytical model, symmetrical effects of CMV on observed variables, and number of variables in the model Selection into Sample	Evans, 1985; Lindell & Whitney, 2001; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff, MacKenzie, & Podsakoff, 2012; Siemsen, Roth, & Oliveira, 2010
Heckman Selection Correction – Use a first-stage probit model of all possible observations to predict inclusion in the sample. The inverse Mill's ratio from this equation is used as a control in the second-stage model of the sample.	An instrument is preferred but not strictly required for the first-stage model. Only addresses bias caused by non-representativeness of the sample, not other forms of endogeneity. Selection of Treatment	Certo, Busenbark, Woo, & Semadeni, 2016; Clougherty, Duso, & Muck, 2016
Selection as Omitted Variable Bias	If the endogenous variable is continuous and “selected” by the subject or the context, then methods used to address omitted variable are applicable.	Bascle, 2008
Heckman Treatment Estimate – Use a first-stage probit model to predict “treatment.” The inverse Mill's ratio from this equation is used as a control in the second-stage model to estimate the treatment effect.	Some variations of this model are available, but all require an instrumental variable or other identifying assumption.	Bascle, 2008; Hamilton & Nickerson, 2003; Wolfolds & Siegel, 2019
Difference in Differences – Panel-data method applied to sets of group means in cases when certain groups are exposed to the treatment over time and others are not.	Endogeneity is only avoided if the treatment is exogenously chosen or treated and nontreated have parallel trends over time.	Athey & Imbens, 2006; Bertrand, Duflo, & Mullainathan, 2004
Regression Discontinuity – An effect is inferred if the regression line displays a discontinuity—a change in slope or intercept—at the cutoff between treatment and control.	Selection of treatment must be determined by a cutoff or threshold in a continuous variable such as a test score.	Hahn, Todd, & Van der Klaauw, 2001; Imbens & Lemieux, 2008; Lee & Lemieux, 2010; Thistlethwaite & Campbell, 1960
Synthetic Control Groups – Creating a control group by matching, coarsened exact matching, or propensity score matching.	Endogeneity is only avoided if the assumptions of selection or observables or ignorability of treatment apply.	Caliendo & Kopeinig, 2008; Dehejia & Wahba, 2002; Li, 2013; Rosenbaum & Rubin, 1983; Stuart, 2010

*Several econometrics textbooks cover the majority of these topics and have been excluded due to space constraints. We recommend Angrist & Pischke (2008), Kennedy (2008), and Wooldridge (2010) as additional sources for the majority of these topics. See Online Appendix A for a list of specific chapters and topics.

a given omitted variable is not always possible. For example, a worker's true ability or a firm's true capability may not be directly observable. If a researcher is concerned about a specific omitted variable (or multiple omitted variables) but unable to include the variable(s) in the study, there are multiple approaches to address this type of endogeneity, as we detail below.

Solutions to Omitted Variable Endogeneity

Solution 1: Design

Among the 86 studies that identified omitted variable(s) as the cause of endogeneity (see Table 4), eight addressed the issue with an experiment, quasi-experiment, or research design choice. Experimental trials can avoid omitted variable endogeneity by randomly assigning participants to treatment and control conditions; random assignment ensures that in idealized format (i.e., adequate sample sizes, effective manipulations, etc.; Krause & Howard, 2003), any omitted variable is evenly distributed across both conditions (thus, the predictor will not display systematic variation with the residual). To illustrate, if it were possible to randomly assign people (firms) to different levels of job satisfaction (firm reputation), we could reasonably assume all omitted variables (e.g., rewards, worker ability; market share, public relations ability) are evenly distributed across treatment groups. In turn, there is no systematic relationship between x and u ; thus, omitted variable endogeneity is not present with this design.

Random assignment may not always be possible and is not problem free (Krause & Howard, 2003), but at times a sample can be defined where an omitted variable does not vary significantly. If the concern is unobserved ability, a study design may use a cohort of workers (firms) promoted at the same time (having the same public relations event), as it may be argued that any unobserved effect of ability is not dissimilar across workers (firms) and thus does not affect the outcome. The "not dissimilar" logic also underlies the use of matched samples to derive "treatment" and comparison groups (see Dehejia & Wahba, 2002). Key to avoiding endogeneity through design is to anticipate the most important omitted variables in advance, measure what is possible, and design the sample to reduce the variance in variables that cannot be measured.

Solution 2: Control and proxy variables

Ten of the 86 articles that identified omitted variables as the cause of endogeneity noted using control variables to address the concern. If a control variable perfectly measures the omitted variable, that source of endogeneity is removed. However, if the variable is not available and cannot be ignored, another way to address this is to find a proxy for the omitted variable (thus replacing an unobserved variable q with a proxy \tilde{q} , so now $y = a + Bx + \tilde{q} + u$). For example, a proxy for worker ability might be education level, and a proxy for firm public relations capability might be the size of media market in which the firm's headquarters is. Inherently, the proxy variable solution implies \tilde{q} has some relationship with q , but measuring q with \tilde{q} includes some measurement error (i.e., $q = \gamma\tilde{q} + u_q$, where γ is the strength of the relationship and u_q is the part of q unexplained by \tilde{q}). Using a proxy to address an omitted variable is often called a "plug-in" solution, but "plugging-in" a proxy requires the proxy itself does not create a condition of endogeneity (i.e., a residual correlating with a predictor). Thus,

like all predictors of y , \tilde{q} should not correlate with the residual for the overall model u , and further, the error specific to measuring q with (i.e., u_q) \tilde{q} should also not correlate with any predictor in the model \tilde{q} (i.e., x). Given q cannot be measured (or there is no need to use the proxy \tilde{q}), scholars must use theory and prior research suggesting \tilde{q} relates to q to justify the proxy (the unobserved q might even be replaced with multiple proxies $\tilde{q}_1, \tilde{q}_2, \dots, \tilde{q}_n$, or as we detail further below, these multiple variables might be utilized as instrumental variables to predict q).

The distinction between controlling for an omitted variable and using a proxy variable depends on how well the measured variable reflects the conceptualized omitted variable, which may be hard to assess. Pei, Pischke, and Schwandt (2019) note the hazards of poorly measured control variables, while Spector and Brannick (2011) discuss how arbitrarily including extra controls can also create bias. Frank (2000) and Pan and Frank (2003) develop a procedure to estimate an impact threshold of a confounding variable (ITCV) that offers promise in this area. ITCV addresses how likely it is that an omitted variable is biasing results by calculating how correlated an omitted/confounding variable would have to be with both the outcome variable and focal predictor variable to change the original inference. The ITCV procedure has only appeared recently in management journals, so best practices are not yet established. For now, we offer a few observations. First, ITCV can be a way to gain added insight into whether the inclusion of an additional variable improves the model estimates. In addition to looking at how the inclusion of the variable affects coefficient estimates, one can use ITCV to see if the potential for omitted variable bias has been reduced. Second, ITCV may be a more principled way to decide when to stop including more control variables. It may be possible to imagine additional variables that are correlated with both the predictor variable and outcome variable, but ITCV gives some guidance on both what that variable would need to look like in terms of relationship to the outcome and predictor to alter inference and whether additional control variables would change the results. Finally, a note of caution: ITCV changes the focus from establishing the assumptions necessary to obtain an unbiased estimate of a coefficient to a focus on whether a result would still be statistically significant in the face of a confound (e.g., an omitted variable is included). Such a shift in focus may be problematic if ITCV becomes another tool in the p -hacking arsenal; thus, as with other approaches, it is necessary to appropriately justify the use of ITCV.

Solution 3: Fixed effects

Eleven of the 86 articles that identified omitted variables as the cause of endogeneity discussed the use of panel data, which implied the use of a fixed effect to address unobserved heterogeneity. If an omitted variable is not available or directly observable but theory or evidence suggests it is constant within a group or invariant over time, estimating a model with individual or group fixed effects can address the issue. Fixed effects add a constant c_i for each entity i in the analysis ($y_i = a + c_i + Bx_i + u$). For example, leadership style (perception of an industry) may be the same for all workers (firms) with the same supervisor (in the industry). If this is so, a fixed effect for supervisors (industries) would address this concern. Similarly, if a researcher had longitudinal data and there is theory or evidence to suggest an omitted variable of concern (e.g., worker ability, firm capability) does not change significantly over time, then an individual or firm fixed effect would address this concern.

A few caveats of fixed effects are notable. First, fixed effects do not fix all endogeneity concerns, but they do work in situations where the omitted variable is constant for all observations with the same fixed effect (Antonakis, Bastardo, & Rönkkö, 2019). Second, fixed effect analyses assess within effects not between effects (for a discussion, see Certo, Withers, & Semadeni, 2017). For example, fixed effects can explain how changes in job satisfaction (firm reputation) affect job performance (firm performance); they cannot explain why some workers (firms) perform differently than others. Bliese et al. (2020) offers a thorough review of the limits and potentials of fixed effects, while also noting how random effects models coupled with the group mean (i.e., the average of all workers or firms in a group) of the predictor variable offer three alternatives that allow for unbiased coefficients and testing of both within and between effects. First, the “hybrid” approach involves “demeaning” or group mean-centering the predictor x_{ij} (for each entity i [e.g., individual worker or firm] grouped by j [e.g., time, work group, industry, etc]) by subtracting the group mean \bar{x}_j . The demeaned predictor variable (i.e., $x_{ij} - \bar{x}_j$) can then be included with the group mean \bar{x}_j , resulting in $y = a + B_1 \bar{x}_j + B_2(x_{ij} - \bar{x}_j) + u$. Second, the group mean \bar{x}_j can be included along with the unaltered, or raw, predictor x_{ij} , resulting in $y = a + B_1 x_{ij} + B_2 \bar{x}_j + u$. Third is to simply use the demeaned value, resulting in $y = a + B_1(x_{ij} - \bar{x}_j) + u$.

Solution 4: Instrumental variables

Omitted variable endogeneity is often addressed with a method dependent on instrumental variables. There may be a concern about omitted variables, but measuring or proxying the variables is not viable. Instrumental variables offer an avenue for unbiased estimates in such cases, but their use requires assumptions based on theory (Wooldridge, 2010). Specifically, instrumental variables require an additional variable z that predicts the endogenous variable x in what is often called “the first stage” ($x = a_x + B_x z + u_x$) but is unrelated to the unexplained portion of the model u_y in what is often called “the second stage” ($y = a_y + B_y x + u_y$; subscripts indicate that coefficients and residuals differ in the two equations). Here, the instrumental variable z predicts job satisfaction x (firm reputation) but is not related to job performance y (firm performance) except via the effect on job satisfaction (firm reputation).

It can be hard to find acceptable instrumental variables as they must meet two conditions. First is relevance, which implies that an instrumental variable is related to the endogenous predictor(s) x . The assumption that the instrumental variable z affects x can be tested directly (Stock, Wright, & Yogo, 2002). Second is exogeneity, which implies that z is uncorrelated with the residual u of the outcome y , meaning the only effect z has on y is through x . This second assumption—that z is exogenous—cannot be tested directly; rather, a researcher must provide conceptual arguments for why the instrumental variable is uncorrelated with the residual in the second stage of the regression (Bascle, 2008). So-called overidentifying restriction tests (e.g., Sargan-Hansen or Sargan’s-J) can test the exogeneity condition but only if a researcher has more instruments than are needed. These tests are based on the assumption that the model is correctly identified (Kennedy, 2008). A Hausman test can determine if an instrumental variable estimation method is needed but only if the instrumental variable is valid (Semadeni et al., 2014).

Methods that rely on instrumental variables require both (1) additional theoretical justification because the exogeneity condition cannot be tested directly and (2) empirical justification to establish the strength of the instruments. Weak instruments are variables that are poor

predictors of the endogenous variable. Using weak instruments can be a case where the cure for endogeneity is worse than the disease (Semadeni et al., 2014), leading to bias in both estimates and standard errors (and the resulting confidence intervals). Often, the bias gets worse if additional weak instruments are added (Bascle, 2008; Conley, Hansen, & Rossi, 2012; Stock et al., 2002). Given the hazards associated with instrumental variables, it is concerning that 28.6% of the articles we reviewed did not explain what instrumental variables were used (see Table 3).

Considering the oft-noted difficulty in identifying instrumental variables (e.g., Larcker & Rusticus, 2010; Semadeni et al., 2014), it may be natural to ask if there are sound strategies to identify them. There are indeed strategies available, but none of the strategies eliminate the need for theory, and in fact, some of the strategies require even more restrictive and difficult-to-justify assumptions than traditional instrumental variables. One strategy to identify instrumental variables is to look for random processes that affect the endogenous variable x . In the micro example, a context where an aspect of the job related to job satisfaction is randomly assigned (e.g., a firm may give better parking spots in a random drawing). In the macro example, social media events involving firms that go viral might serve as an instrumental variable (if the events' publicity is theoretically or empirically shown to be random and unrelated to firm performance).

Another strategy to identify instrumental variables is to use prior period data (often called "lagged variables"). Prior job satisfaction (firm reputation) might affect current job satisfaction (firm reputation) but not directly relate to current job performance (firm performance), yet the lagged variables must still meet the requirements of an instrumental variable. It is usually not hard to argue that a lagged variable of x is related to its current value, but it is likely harder to argue the lagged variable is not related to the residual. For example, a researcher may have concerns about an omitted variable like mental (innovation) capability in testing the relationship between job satisfaction (firm reputation) and job (firm) performance. Prior mental (innovation) capability might be a proposed instrumental variable for current mental (innovation) capability, but a researcher would need to argue the lagged value is not related to the residual in current job (firm) performance. Using deeper lags increases the likelihood that the instrumental variable is unrelated to the residual in the outcome variable but likely also decreases the strength of the relationship to the endogenous variable. Estimation techniques like Arellano and Bond (1991) build on the assumption that taking first differences (subtracting the lagged value from current value of a variable) eliminates unobserved heterogeneity and serial correlation of the residual. If these assumptions hold, the lagged first differences can serve as valid instrumental variables.

Several emerging methodologies, many from outside the management literature, have been proposed for instrumental variable approaches (e.g., Maydeu-Olivares et al., 2020). One is using instrumental variables in SEM contexts through model-implied instrumental variables (MIIVs; Bollen, 1996; 2019). Typical instrumental variables are external to the model—what Bollen (2012) calls *auxiliary variables*. In contrast, MIIVs are derived from variables already in the model, so there is no need to add variables to the model. In general, if an SEM is identified, enough MIIVs can be derived to estimate each equation. There are additional methods based on assumptions about the distributional form of variables and various residuals such as Gaussian copula (Papies, Ebbes, & Van Heerde, 2017), simulated maximum likelihood estimator (Villas-Boas & Winer, 1999), Garen's two-step-model (Zaefarian,

Kadile, Henneberg, & Leischnig, 2017), polychoric instrumental variables (Bollen, 2012), and De Blander's estimator (Sande & Ghosh, 2018). While these methods may avoid theoretical specification of an instrumental variable, they rely on strong assumptions that only apply to specific situations.

Instrumental variables models can be estimated with various techniques like two-stage least squares (2SLS), three-stage least squares (3SLS), maximum likelihood, generalized method of moments (GMM), and SEM. Variations of these methods like two-stage predictor substitution and two-stage residual inclusion (Terza, Basu, & Rathouz, 2008) are also proposed. While comparing these methods is beyond the scope of our paper, important to note is that any technique still requires identifying the cause of endogeneity and justifying the instrumental variable(s). The validity of any approach depends on assumptions built into that approach, and thus, it is important to explicitly specify the assumptions required. Studies should explain the relevance of the instrumental variable(s) and estimation technique, show "first-stage" or similar analyses, and provide relevant tests for any approach used (Semadeni et al., 2014).

Cause 2: Simultaneity

A second mechanism that can cause endogeneity is simultaneity, which is sometimes also labeled reverse causality. Simultaneity was identified as the cause of endogeneity in 20.7% of the works we reviewed (Table 4). So far, our focus has been on how a predictor variable x affects an outcome y . When the reverse is also true, so y affects x , simultaneity exists (see Figure 3 in Table 1). For example, testing how job satisfaction (firm reputation) affects job (firm) performance may be intended, but job (firm) performance may also affect job satisfaction (firm reputation). Thus, one way to view simultaneity is as a type of omitted variable problem: Prior performance may be an omitted variable in any study when it is correlated with both current performance as an outcome y and the predictor variable x (e.g., job satisfaction, firm reputation). However, with simultaneity between a predictor x and outcome y , there are no controls that can be added to the model to fix the problem, so control variable and proxy variable solutions noted above are not applicable but instrumental variables are (as we further detail below; e.g., Bhawe, 2014).

The most common context where researchers address simultaneity bias is in longitudinal or panel data, where there are repeated measures on multiple units (e.g., individuals, firms). While such models have long been common for macro researchers, the adoption of longitudinal models for micro researchers has increased substantially in the last decades (e.g., growth models, experience sampling methodologies, etc.; Bliese et al., 2020; Fisher & To, 2012). Part of the motivation for adopting such strategies in micro studies is a desire to take steps toward causal claims that are not possible in single-measurement data. Drawing from Kenny's (1979) prescription that justification of causality comes from establishing three conditions (a relationship— x is related to y ; temporal precedence— x must precede y in time; and nonspuriousness—no third variables cause both x and y), if the effects of prior levels of the outcome variable can be statistically controlled, this can bring us closer to establishing temporal precedence and potentially narrow the range of variables that may create spurious relationships.

Some argue that, in some cases, simultaneity (or reverse causality) does not create endogeneity if the variables do not affect each other *at the same time*—that is, if there is a time lag

(or temporal spacing) in the study design. Here, if past x (time $t - 1$) affects current y (time t) and current y (time t) affects future x (time $t + 1$), then one can control for previous events. However, this argument does not necessarily eliminate the endogeneity concern (Bellemare, Masaki, & Pepinsky, 2017). A main assumption of regression-based analyses is independence of residuals; that is, u_1 at $t - 1$ and u_2 at t are expected to have a correlation of 0. Yet in longitudinal data, both the variables and the residual for adjacent observations are likely correlated (referred to as autocorrelation, serial correlation, or serial dependency; Dodge, 2006; Wooldridge, 2010).

Variables can be autocorrelated for many reasons. For example, prior performance (of a worker or a firm) is a good predictor of current performance since many performance predictors are similar over time. Performance may also be self-reinforcing (a good performing worker/firm receives feedback, resources, and experience that enables future performance as in the Matthew effect). Residuals can also be autocorrelated for several other reasons. One case may be if variables are left out of the model in the present period (time t) and were also left out of the model in the last period (time $t - 1$) and those variables do not change over time. For example, failing to measure depression in the relationship between job satisfaction and job performance or failing to measure the strength of the economy (e.g., a recession) in the relationship between firm reputation and firm performance. Another case may be when the value of the last period's variable (time $t - 1$) is directly related to the present period's (time t) outcome variable in a cyclical fashion. For example, data collected on an hourly basis may reflect within-day cycles (e.g., ups and downs in performance), while data collected less frequently can reflect seasonal, quarterly, or yearly cycles. In both these cases, the outcome variable is autocorrelated, and since the cause of the autocorrelation is not included in the model, the residual is autocorrelated.

Solution 1: Design. Experimental trials address simultaneity by manipulating a predictor variable. When a researcher can assign or otherwise manipulate levels of x in a treatment group and not in a control group in an ideal experimental setting, the variation in y can be attributed to the manipulation and not the simultaneous effects of y on x (assuming adequate sample size and the mitigation of all threats to internal validity). Some quasi-experimental designs offer a solution if researchers can design questions around exogenous events (i.e., events that occur naturally that have nothing to do with the proposed model) to mimic a true experiment. For example, consider a firm receiving unexpected recognition for being the best place to work. The subsequent media attention might increase worker satisfaction (firm reputation), so a comparison of job (firm) performance before and after the event may eliminate the simultaneity bias. True experiments of this kind are often not feasible and exogenous events may be few and far between, so researchers often must use analytical techniques to make causal inferences.

Longitudinal designs also offer a solution if the observed data are not autocorrelated; thus, sound diagnostics should be utilized for any analyses. Though beyond our scope, West and Hepworth (1991: 626) offer a nice primer, both noting how ignoring autocorrelation (and thus assuming there is no serial dependency between x and y) "leads to biased standard errors for all significance tests and biased estimates" and comparing the merits of various specification tests. As with all techniques, we again emphasize justifying their application (Wooldridge, 2010).

Solution 2: Instrumental variables. The solution to the simultaneity problems using instrumental variables is essentially the same as the solution to omitted variables endogeneity. If one considers simultaneity as two simultaneous equations, then $y = a_y + B_y x + u_y$ at the same time that $x = a_x + B_x y + u_x$ (where subscripts indicate each equation has separate coefficients and residuals). At least one exogenous instrumental variable is needed to estimate the coefficient for each endogenous variable. Important here (as noted above) are (1) theoretical grounding (i.e., the instrumental variable must be related to the endogenous variable but unrelated to the unobserved residual) and (2) appropriate analyses (i.e., conducting and reporting appropriate tests).

Cause 3: Measurement Error Endogeneity

Measurement error is a third mechanism that creates endogeneity but is rarely identified as such (1.5%, see Table 4), yet rather than indicating a lack of a problem, the rarity may be due to the fact that, as Kennedy (2008: 160) speculates, most econometric models work best with the assumption of zero measurement error. As Figure 4 of Table 1 shows and consistent with Classical Test Theory, when x is imperfectly measured as \tilde{x} , measurement error can be expressed as $x - \tilde{x}$ (i.e., true score minus observed score equals error). If the measurement error is related to y , then it will be captured by the residual u , thus creating an endogeneity bias as \tilde{x} and u are correlated. In the case of a single estimator, measurement error will attenuate estimates of B , but when there are multiple predictors in the model, the sizes and directions of the biases are not easily derived (Angrist & Pischke, 2008). Imagine gathering self-reported job satisfaction data via a survey. Some people fill out the survey at the start of the day while others stay after hours to do so. Reported job satisfaction may systematically differ from actual job satisfaction based on when the survey is completed. For example, high performers may be more likely to handle nuisance surveys at the end of the day. A similar issue extends to measures of firm reputation: There may be systematic measurement error if larger firms have higher ratings regardless of real perceptions of the firm. In both cases, if the cause of measurement error is known, then the cause is an omitted variable that can be handled with the methods noted above (Antonakis et al., 2010).

Viewing CMV as a special case of measurement error may help clarify this point. When not modeled, the variance associated with a referent (e.g., all self-report) or study format (e.g., all scales measured on a 5-point Likert-type scale) may bias the coefficients; that is, a common method may affect responses (e.g., answering high or low, avoiding extremes) that bias the estimates of predictor x on outcome y . The parallel between CMV and endogeneity is seen in the analogous ways each are addressed statistically (Schaller, Patil, & Malhotra, 2015).

Macro researchers may feel they do not have measurement error problems since they use surveys less frequently and instead rely on objective data, yet ratio variables (like R&D intensity and return on assets) are often used and there is growing awareness that these variables face their own measurement challenges (Wiseman, 2009). Recently, Certo, Busenbark, Kalm, and LePine (2020) framed these concerns in terms of endogeneity by arguing that a ratio predictor variable is necessarily endogenous when the outcome variable is also a ratio with the same denominator.

Solution 1: Design. A main design principle to avoid measurement error endogeneity is using measures free of systematic bias. If primary data are collected, common ways to do this

are using validated measurement instruments (Greco, O'Boyle, Cockburn, & Yuan, 2018) and survey designs that reduce spurious correlations (e.g., vary scale anchors, separate measures by time; Podsakoff, MacKenzie, & Podsakoff, 2012). If archival data are used, similar tenets apply; measures should have sound validity (i.e., measure what it intends to measure and not something else) and steps should be taken to reduce design-induced errors. While archival data may limit researchers' ability to address certain aspects of design, one advantage is it often allows for using multiple existing measures. Multiple measures can offer evidence that the results are not due to error in a given measure, assuming the measures exhibit convergent validity (e.g., Bromiley, Rau, & Zhang, 2017; Hill, Kern, & White, 2012, 2014).

Experimental design, while desirable in many ways, does not rule out measurement error endogeneity. For example, if a researcher manipulates the effect of job satisfaction (firm reputation) on job performance (firm performance) in an experiment, the outcome y may be poorly measured such as if performance is operationalized as performance on a single task or through error associated with supervisor or external ratings. If the unmeasured portion of y —that is u —is related to the manipulated predictor variable x , then endogeneity is still present.

Solution 2: Account for measurement error. Methods exist to address measurement error by directly modeling it like SEM, but these techniques are generally only applied in latent variable models (more common in micro, but applicable to macro; Bergh et al., 2016; Shook, Ketchen, Hult, & Kacmar, 2004). One benefit of SEM and related approaches is that researchers can model correlations between residuals among both the indicators and latent variables; however, methodologists have emphasized the need to craft strong *a priori* reasons for doing so as it is easy to capitalize on chance (e.g., Cole, Ciesla, & Steiger, 2007; Cortina, 2002; Landis, Edwards, & Cortina, 2009). Another technique is a marker variable (Williams & O'Boyle, 2015), which calls for using a theoretically unrelated variable measured with the same or similar scale, valence, referent, etc. Since its relationship with the predictor x and outcome y are assumed to be zero, any observed covariation is assumed to be a function of CMV. As a marker variable is exogenous, the method variance is then addressed, or “covaried out.” This functionally is the same as an instrumental variable's requirement of being exogenous. As noted above, using multiple measures to address limitations in a measure may offer evidence that estimated relationships are robust to measurement error if the measures converge.

Cause 4: Selection

The final mechanism that can create endogeneity is selection (identified as the cause of endogeneity in 32.1% of articles; Table 4), which occurs through two separate mechanisms. First, bias is created when observations are not randomly sampled but instead “selected” through choices of the researcher or participants (Cause 4a). For example, data are only available from workers who responded to a survey (e.g., conscientious or neurotic ones, only those with a job have an observable job satisfaction or job performance) or certain firms (e.g., not all firms appear on a stock market but stock price is used to measure firm performance). We label this “selection into sample” (Cause 4a). For many micro researchers, selection effects may be understood within the framework of indirect range restriction (Beatty, Barratt, Berry, & Sackett, 2014; Hunter, Schmidt, & Le, 2006). As an example of hiring, job applicants may

be hired or “selected” based on a variety of characteristics including previous experience, relevant skills, and fit with the organization. If a researcher is interested in whether person-organization fit predicts job performance, the analysis is limited because applicants with lower levels of expected fit were screened out and the resulting coefficient between fit and performance may be biased.

Second, bias is created when the level of the endogenous predictor variable is “selected” (Cause 4b). For example, participation in training is not random but determined by a supervisor, and various firm strategies that affect performance are not randomly determined but chosen by the firm. We call these various examples “selection of treatment.” When the level of the endogenous variable is not randomly created, the potential bias means that the estimated coefficient cannot generalize to a larger population (i.e., it is an artifact of selection).

Cause 4a: Selection Into Sample

Solution 1: Design. Randomly assigning study participants can counteract endogeneity concerns from selection into the sample, but such designs are not failproof (Krause & Howard, 2003). For example, even if participants are randomly selected into training in a firm, the workers have all been selected into that firm, so effects of training on performance is indicative of only job-holding individuals who have been selected into that firm rather than the full population of potential workers (as is a common criticism of using undergraduate students to represent working adults, for example; the effects of the two groups may not be equal). Likewise, attrition from randomized studies may be nonrandom. If individuals self-select into or out of a treatment or control in a nonrandom fashion, then endogeneity may still be a concern.

Solution 2: Heckman selection model. If selection into a sample cannot be randomized, another approach is needed. For example, consider Heckman’s (1976) comparison of female workers’ hours and wages to their male counterparts. An endogeneity concern in this study is that choosing to work or not is not random (similarly, workers reporting job satisfaction and firms having publicly available financial data may not represent a random sample). Many factors may lead women to choose to work (firms to choose to be publicly listed), meaning the observed data do not fully represent women (firms), only women choosing to work (firms choosing to be publicly listed). If unmeasured factors (e.g., family or personal factors for workers, industry or financial factors for firms) affect the binary choice to act and also influence the focal outcome (e.g., wages, job or firm performance), an endogeneity concern exists that cannot be simply addressed by including the unmeasured factors (there are no data on wages of women choosing not to work or firms choosing to be private, so these variables are unobservable for a portion of the population). Thus, we can estimate how job satisfaction (firm reputation) affects job (firm) performance only for those workers (or firms) we can observe. It is important to distinguish this cause from other causes of endogeneity where the sample is representative of the population and the full range of the outcome is available (e.g., we can observe wages and job performance for union and nonunion members or firm performance of publicly traded and privately held firms; what is missing are the factors leading to the self-selection of the level of the predictor).

The Heckman selection model is akin to an instrumental variable method, addressing an omitted variable bias arising from a specific sample selection issue (Certo et al., 2016; Clougherty et al., 2016). In a Heckman model, a first-stage probit model estimates the likelihood of entering the sampling condition, and a transformation of the predicted value in the first stage (the inverse Mills ratio, or IMR) is derived to represent the selection hazard of entering the sample. Using the IMR in a second-stage model of interest provides an estimate of the selection hazard, yielding consistent estimates of the predictor on the outcome. Like instrumental variable methods, a third variable w (referred to as an exclusion restriction) is needed, and this variable w should affect the probability of being in the sample (i.e., be related in the first-stage probit) but be “excluded,” hence the name, from the second-stage model based upon theoretical logic for why w does not affect the outcome (Wooldridge, 2010). Here again, then, researchers must rely on theory and also should assess the underlying assumptions; Shaver, (1998), Hamilton and Nickerson (2003), Certo et al. (2016), and Clougherty et al. (2016) offer thorough discussions.

Cause 4b: Selection of Treatment

Solution 1: Design. As with selection into sample endogeneity, randomly assigning the participants to treatment conditions can counteract selection of treatment concerns in the ideal situation; yet again, the design is not failproof. In particular, it is often impractical to assign a meaningful number of participants to varying levels of the treatment while determining what the meaningful levels are (e.g., how much training); in turn, conclusions drawn may apply only to those levels (e.g., an hour of training versus five) rather than more broadly (e.g., more training, generally). As such, identifying designs where different treatments or levels of treatment either are possible, or occur naturally, can help address selection of treatment endogeneity concerns.

Solution 2a: Omitted variable techniques. We first consider when there is a treatment that is not randomly assigned and, therefore, selected. If the endogenous predictor variable that is selected is not dichotomous and instead is continuous (e.g., years of education or advertising spending) or ordinal (e.g., a worker having an associate, bachelor, or master degree or a firm choosing an acquisition, joint venture, or greenfield), this is similar to omitted variable endogeneity and all solutions discussed above are appropriate.

Solution 2b: Heckman treatment model. If the treatment is a dichotomous variable (like participating in training or making an acquisition), then 2SLS and related instrumental variable methods are inappropriate, and instead, a method such as a Heckman treatment effect should be used (again, see Certo et al., 2016; Clougherty et al., 2016; Hamilton & Nickerson, 2003; Shaver, 1998). The Heckman treatment model is similar to the Heckman selection model except that the first-stage model is a prediction of treatment rather than a prediction of inclusion in the sample. A second complication is created by a dichotomous selection of treatment in which the estimated coefficient is a “treatment effect” with many possible interpretations. For example, studies of how training affects job performance may want to determine: How much the average worker would benefit from training, whether those receiving training benefited, or how would workers that did not have training benefit if they had

it? Depending on the mechanisms that determine who is trained, these different forms of the “treatment effect” might all be different (Blundell & Dias, 2009). Further complications are posed by categorical variables (e.g., different trainings, strategic choices), thus requiring unique estimation models and care with a “treatment effect” given categories (e.g., Bollen & Maydeu-Olivares, 2007).

Solution 2c: Estimating average treatment effects. Selected, dichotomous treatment is common in many fields. Many methods have been developed to address this by estimating what is called the average treatment effect (e.g., Angrist & Imbens, 1995; Wooldridge, 1997) or how much the average treated participant benefits versus the average nontreated participant. These approaches can be split into two categories: (1) Difference-in-Differences approaches calculate how much participants (workers, firms) improve after treatment and compare the improvement to how much nonparticipants improved in the same period (e.g., Athey & Imbens, 2006; Bertrand, Duflo, & Mullainathan, 2004), and (2) Synthetic control group approaches (e.g., matched sample, propensity score methods, coarsened exact matching) compare treated entities to nontreated entities that are similar on observables or have similar likelihood of treatment (e.g., Caliendo & Kopeinig, 2008; Rosenbaum & Rubin, 1983). In general, methods estimating an average treatment effect do not address endogeneity; rather, they rely on assumptions (labeled ignorability of treatment, selection on observables, or the conditional independent assumption) that endogeneity is not a concern based on the logic that sampled entities may differ on a treated variable but are otherwise about equal (or at least not dissimilar); thus, the claim of unbiased estimates rests on the assumption that unmeasured variables affect all sampled groups equally (Dehejia & Wahba, 2002; Li, 2013). This is not to say these methods are not important, as they address important problems other than endogeneity. Like all techniques, they require theoretical and empirical justification regarding solving a given problem and are not “catch all” cures.

Solution 3: Regression discontinuity designs (RDDs). One final way to address dichotomous selection is an RDD (Lee & Lemieux, 2010), which can be considered a quasi-experimental approach (Angrist & Pischke, 2010). The basic idea of RDD is that there may be existing environmental conditions, which create an arbitrary threshold or cutoff point that can approximate random assignment. The observations just below and just above the threshold should be approximately equal on all omitted variables (similar to random assignment), yet they are categorized by researchers as being in different treatment groups based on falling above or below the arbitrary threshold. Returning to our training example, if workers are selected into training based on poor performance ratings, where only those with a rating of 2.5 or below on a 5-point scale are sent to training, a researcher may consider that workers with ratings of 2.4 who qualify for training and workers with rating of 2.6 who do not qualify for training are functionally the same in terms of performance. Thus, one could test the effect of training in the trained versus untrained population as in a true experiment. Complications in RDD can arise in selecting the number of units to include around the treatment effect (e.g., from 2.4 and 2.6 or 2.3 and 2.7?) and issues related to contamination from those receiving the treatment with those not receiving the treatment (Imbens & Lemieux, 2008; Thistlethwaite & Campbell, 1960).

The Possibility of Multiple Causes of Endogeneity

In delineating the causes of endogeneity, we highlight how one study may have multiple endogeneity problems (e.g., how job satisfaction affects job performance or how firm reputation affects firm performance). There are two ways in which multiple endogeneity issues may arise in a single study. First, a variable can be endogenous due to multiple causes (i.e., omitted variables, measurement error, simultaneity, and selection). Notably, addressing one cause does not address other causes (e.g., simultaneity in job satisfaction and job performance, omitted variable bias in firm reputation on firm performance). Second, a study can have multiple endogenous variables. Imagine a study on job satisfaction also looking at how training affects job performance (or a study of firm reputation also investigating how acquiring affects firm performance). Effects of both predictors are likely subject to different omitted variables, but addressing the endogeneity of one variable does not address endogeneity from the other. To make matters worse, endogeneity in one variable biases *all* coefficient estimates, not just a coefficient for the endogenous variable (Wooldridge, 2010). As such, scholars must turn to theory and analyses to identify endogeneity for all predictors and address each cause for each variable. York, Vedula, and Lenox (2018) is an example, arguing economic incentives, social movement pressure, and market intermediaries all affect the adoption of green building practices. The paper addresses possible simultaneity by estimating a model in which each variable is predicted by a separate exogenous variable. Certo et al. (2016) also provides guidance on addressing multiple sources of endogeneity.

Summary of causes. Table 1 outlines and depicts the causes of endogeneity, while Table 5 maps those causes to associated solutions to remedy them and provides key source material. Although we do not include the entirety of the copious endogeneity discussion happening across related fields, we contend our review and summary accurately reflect the relevant endogeneity discussion as it pertains to management research at all levels of analysis and content areas.

Recommendations to Bridge the Methodology-Practice Gap

Our review aims to enhance understanding of endogeneity - an issue that poses serious implications for interpretation of study outcomes - by organizing the vast literature on the sources of bias and methodological solutions (e.g., Antonakis et al., 2019; Bhawe, 2014; Maynard, Luciano, D'Innocenzo, Mathieu, & Dean, 2014). As we outlined above, definitional and terminology differences across this literature hinder direct comparisons on the term endogeneity alone. In this paper, we first reviewed how researchers discuss and address endogeneity and then mapped the extensive methodological literature on how to treat the associated problems. Doing so has reaffirmed prior reviews documenting divergence between best practices and actual practices and has also shown that this divergence is not caused by a lack of relevant methodological resources. Building on our review, we offer a set of recommendations for both authors and gatekeepers that may help reduce the disconnect between the methodological and empirical literatures.

To better illustrate our recommendations, we again use the metaphor of endogeneity as a disease and extend it to physicians treating patients. If endogeneity were a disease, we would want those treating it to (1) offer a clear diagnosis, (2) justify the technique used in treatment,

and (3) be transparent in prognosis of the result. Our review of empirical work, like others (e.g., Antonakis et al., 2010; Certo et al., 2016; Semadeni et al., 2014; Wolfolds & Siegel, 2019), shows actual practices often deviate from best practices: Diagnoses of endogeneity are not clearly connected to causes, treatments for endogeneity are not clearly justified, and the concluding prognoses are usually that endogeneity has been addressed or cured. Our recommendations leverage the desired practices, but they cannot be achieved by researchers alone. Reviewers and editors must adopt them as well. Finally, just as there is no such thing as unequivocal perfect health, there is no such thing as the perfect study; thus, authors and gatekeepers must accept that tradeoffs are often needed. We elaborate on each of these points below, while Table 6 summarizes our recommendations and serves as a guide for authors and gatekeepers regarding endogeneity.

Clear Diagnosis of Endogeneity

Without a clear diagnosis of the cause of endogeneity, wrong treatments may be given, which either will not address the actual cause or may even exacerbate the problem (Semadeni et al., 2014). Thus, we first recommend exercising greater care in diagnosis—that is, in establishing if and why a specific cause of endogeneity may exist. Here, we offer three specific action items.

First, to identify if and why a specific cause of endogeneity exists, thinking through possible causes of endogeneity is a key step in study design. To this end, Table 2 provides a set of thought experiments for each cause of endogeneity that can be used prior to data collection and analyses. Identifying a possible cause of endogeneity can allow scholars to design around endogeneity threats rather than trying to analyze through them alone. Researchers may also realize there are multiple potential causes of endogeneity, each one requiring a separate diagnosis (see Table 1). Thorough and specific diagnosis is essential not only to prescribing a treatment but also to clarifying the specific causation implied by the hypotheses and theoretical models.

Second, the typology of causes of endogeneity in Table 1 can help authors and reviewers to “speak the same language” when communicating about endogeneity. Specifically, we suggest any endogeneity concern be stated in terms of a specific cause (omitted variable, simultaneity, measurement error, selection into sample, or selection of treatment) with as much specificity and theoretical rationale as possible. That is, comments akin to “we addressed endogeneity” should be replaced by the specific cause (e.g., simultaneity) of endogeneity and rationale for why that cause may be present in the specific relationships of the study. Like our first suggestion, this requires identifying possible endogeneity causes (and, as we note below, also requires authors to diagnose, with justified techniques, if a specific cause of endogeneity is in fact present).

The recommendation of a clear diagnosis has implications for gatekeepers as well. This means that endogeneity cannot be cast as a lurking “boogieman.” Authors share frustration that endogeneity both often seems to be a cudgel for gatekeepers to strike down any paper that by design or analysis is not problem free and, relatedly, that such cudgeling may move us toward irrelevance as we cannot advance knowledge in any way for fear the advancement is not perfect (Frank, 2000; Shugan, 2004). We recommend critiques and any solutions given in the review process should also speak to a specific cause of endogeneity bias rather than a

Table 6
Recommendations for Authors and Gatekeepers Regarding Endogeneity

	Practices to Avoid	Practices to Encourage
Clear diagnosis		
Authors	<ul style="list-style-type: none"> Discussing endogeneity as a generic concern or an abstract threat to the study 	<ul style="list-style-type: none"> Identifying specific causes of endogeneity and link to related terms (see Table 1) Assess potential bias specific to each hypothesized relationship (see Table 2) Acknowledging that endogeneity concerns are present in virtually all social science research
Gatekeepers	<ul style="list-style-type: none"> Rejecting a study because of a vague "endogeneity concern" Expecting that every possible endogeneity concern can be addressed within a single study 	<ul style="list-style-type: none"> Providing theoretic/empirical rationale for potential bias when asking that authors address a specific cause of endogeneity Educating researchers less familiar with endogeneity to consider its effects on study findings
Justification of technique used		
Authors	<ul style="list-style-type: none"> Choosing a methodology because it was used in a previously published study Claiming that any one methodology is capable of addressing all endogeneity issues 	<ul style="list-style-type: none"> Explaining how each specific cause of endogeneity is addressed by a chosen technique (see Table 5 for a catalogue of approaches) Citing methodological source material Stating the assumptions of the chosen technique
Gatekeepers	<ul style="list-style-type: none"> Accepting a poorly explained methodology Suggesting a methodology to address endogeneity without describing the need, purpose, and proper interpretation of findings, especially for authors who may be unfamiliar with the topic 	<ul style="list-style-type: none"> Expecting that the authors justify and explain how the chosen methodology applies to this study Guiding authors less familiar with endogeneity to design and analysis techniques appropriate for their study (including those developed in other fields and content areas)
Transparency in prognosis of results		
Authors	<ul style="list-style-type: none"> Claiming that endogeneity concerns are addressed in unreported results 	<ul style="list-style-type: none"> Reporting (a) Naïve results — without any correction; (b) results from the specific approach; (c) all associated analyses (e.g., first-stage models) and (d) specification tests. Utilize online supplements if needed.
Gatekeepers	<ul style="list-style-type: none"> Claiming that all endogeneity concerns have been eliminated Expecting or encouraging claims that endogeneity has been eliminated 	<ul style="list-style-type: none"> Acknowledging that conclusions are contingent on the limitations and assumptions underlying the methodology Fully describe the need, purpose, and proper interpretation of findings Encourage conclusions about results that are contingent on the limitations and assumptions of the method

general concern about endogeneity (Shaver, 2019). Focused statements such as “although you utilize an instrumental variable approach, you have not clearly defined the source or type of endogeneity that this approach addresses” or “it is possible that your sample suffers from selection into sample endogeneity because of . . . you may consider using a Heckman selection model to address this type of endogeneity” provide clearer guidance to authors about potential causes and also help inform possible solutions. At the same time, such specificity limits the likelihood that authors choose a method that may address endogeneity “in general” (e.g., instrumental variables) but that also leaves results unchanged (i.e., *p*-hacking). To this end, gatekeepers can also help ensure papers provide clear diagnostics and, in doing so, help establish norms about specifying the form of endogeneity and linking to related terminology. Tables 1, 2, 5, and 6 may also serve as guides to reviewers. The specific causes, diagrams, terminology, and thought experiments can inform comments with greater diagnostic precision that point to appropriate techniques, which we elaborate on below. We thus hope our review facilitates a more productive conversation between authors and reviewers focused on specific causes rather than vague concerns.

Third, and building on the earlier points, clear diagnosis and associated communication about it can be enhanced by linking related terms, especially as it applies to communication across micro and macro domains. As others have identified, different terminology is at times understandable as research traditions evolve but nonetheless “produces confusion . . . that impede(s) the ability of members of a research community to communicate with each other or to accumulate knowledge” (Suddaby, 2010: 352; see also Pfeffer, 1993). Thus, we also suggest authors and reviewers make explicit linkages to related terms (e.g., noting how omitted variable endogeneity is also called left out variables, missing variables, etc.) to aid communication with those more familiar with alternative terms. There are at least two benefits of this common lexicon: (1) It expands the methodological toolbox for researchers so that they can apply novel methodologies to a particular endogeneity concern, and (2) it enables clearer understanding across domains, as micro and macro researchers, or even authors and gatekeepers therein, can more easily interpret methodological approaches and findings from disparate realms.

Justify the Technique Used to Address Endogeneity

As scientific researchers, we operate using a common compendium of knowledge (here, the methodological source material). Citing how prior empirical papers addressed endogeneity cannot supplant specific justification of design and analyses grounded in the methodological literature. There is value in seeing how peers have addressed endogeneity, but relying solely on prior empirical studies rather than the source material is inadequate for two reasons. First, prior empirical studies may be flawed (as evidenced by the numerous methodological reviews) or even out of date as the literature on ways to address endogeneity is quickly evolving. Second, every methodological choice must be made for a specific research context. An instrumental variable used in one study does not mean that it is appropriate in another study with a different predictor variable, different outcome variable, or different cause of endogeneity, for example. When researchers properly justify and explain their methodological choices, they build awareness of the compendium of knowledge. We offer four interrelated steps to justify a technique used to address endogeneity. Further, a summary of techniques,

requirements and limitations, and selected methodological sources appear in Table 5 and the annotated bibliography in Online Appendix A as resources to help authors and reviewers with the recommendations below.

First, given that using a technique to address endogeneity that is not correct can be worse than the problem itself (Semadeni et al., 2014), we first need to justify that the issue needs to be addressed—here, our diagnostics from Step 1 are important, but analyses can help establish if the specific problem diagnosed is present. Second, once a specific endogeneity cause is established, we suggest greater detail in justifying whether the technique utilized to address it is appropriate, building on relevant methodological source material (Table 5 may be helpful in linking specific causes of endogeneity to appropriate remedies and methodological source material). Justification cannot be by mere citation alone—as seen in statements such as “following Paper X . . .”. Instead, it is important to establish methodological justification for why a remedy is appropriate for addressing the specific source of endogeneity in the context of the focal study building on source material. Doing so can help avoid the “telephone problem” identified in our review, where papers follow prior empirical work rather than the methodological article, as well as misapplication of techniques. Third, justification should also be explicit about the assumptions of the technique. Clarifying the assumptions are worth specific note, as it is possible that a technique could be well-justified but not offer readers clarity regarding the assumptions on which the technique is based. Fourth, clear justification requires greater reporting of information relevant to endogeneity, including appropriate detail of analyses, so others are able to tell both if the methodology was needed in the first place and adequate for the purpose and context it was used for. Specifically, then, the naïve results—without any correction—should be presented alongside the results from the approach that attempted to address the specific form of endogeneity as well as any associated analyses (e.g., first-stage models) and specification tests (see Semadeni et al., 2014).

The recommendation for better justification of methods used to address endogeneity will require the cooperation of reviewers and editors. When gatekeepers observe a study using precedent (e.g., following Paper X . . .) to justify the methods used, they should advise authors to instead clearly justify the technique following the aforementioned steps. To this end, Table 6 may serve as a guide for explanation of the methodology. At the same time, additional reporting of the results showing how much results were affected by methods to address endogeneity may also require more journal space. If space constraints do not allow, we encourage gatekeepers to ask for the information during the review process and encourage researchers to use online supplemental materials. We further encourage journals to maintain these for the purpose of posterity and access rather than asking individual scholars to find ways to share them widely. Of course, creating space for the discussion of one topic may be thought to come at the expense of another, and we realize that detailed explanations and associated tables that are recommended (e.g., producing the first-stage model) are often cut for length reasons. Beyond our recommendation of supplements that can be stored online, we hope that our review spurs conversations among gatekeepers on changing norms to help advance our knowledge. Numerous authors from different fields (Angrist & Pischke, 2010; Antonakis et al, 2010; Shaver, 2019) have argued that establishing causality is a demanding process, so maybe the justification of a chosen method deserves as much attention as the justification of hypotheses.

Transparency in the Prognosis of Results

We also recommend greater transparency in the resulting prognosis. As mentioned earlier, endogeneity is a complex problem for which there is no blanket solution. Multiple studies that build on each other may be needed to address the various endogeneity issues. Including an additional control variable does not address all omitted variable bias and certainly does not address simultaneity. In a research paper, the discussion of results is where the prognosis is given. Rather than blanket statements such as “we controlled for endogeneity” or “endogeneity was found not be a problem in a previous publication,” we instead encourage statements such as “these conclusions are robust to the alternative explanation that Z is the causal mechanism of both X and Y” or “our design reduced concerns over simultaneity.” Precise claims about the conclusions that can be drawn from the analysis help future researchers to correctly build on the findings and highlight areas where future research could further clarify the results. As with the aforementioned suggestions, gatekeepers can help here by taking steps to establish this norm.

A Final Recommendation

Beyond our best practices that build on the medical example of needing clear diagnosis, justification of the treatment, and transparency in the prognosis, we have a final recommendation that applies equally to our field and medicine. This recommendation, stated directly, is “authors and gatekeepers should realize it is impossible for any one study to fully mitigate all endogeneity concerns.” In other words, no study is perfect, and if we hold that ideal, we may not advance understanding in a systematic way and thus miss important knowledge. This recommendation may be difficult, as norms in management research have tended to coalesce around more unequivocal statements of findings (which will also affect whether more transparent prognoses become the norm as well). Thus, we repeat others in suggesting that journals allow an accumulated body of knowledge to emerge on a specific research question through multiple studies (e.g., Shaver, 2019; Wolfolds & Siegel, 2019). Only through repeated studies of a research question using different designs and analyses can we gain confidence that a specific endogeneity threat is fully addressed. Even the idealized experiment requires triangulation with different designs, samples, and operationalization of variables before causal claims free of any validity threat can be made. In micro research, it has long been common to have separate studies that (1) introduce a new construct (perhaps in a theoretical article only), (2) establish a valid measure of the construct, (3) identify antecedents, (4) consequences, (5) mediators, and (6) moderators of the construct. Similarly, progress on some research questions will require multiple studies that address different aspects of endogeneity. One study may identify valid instrumental variables to account for simultaneity, while another may use a given exogenous event to rule out viable alternatives.


Conclusion

Based on our review of empirical research, the process of building knowledge through many studies is hindered by the lack clarity related to endogeneity in terms of diagnosing the cause, justifying the techniques used, and reporting of results. We are optimistic future research will allow building knowledge in a body of research if the practices of researchers and gatekeepers

in addressing endogeneity provide (1) clear diagnosis, (2) justification of techniques used to address endogeneity, and (3) transparency in the prognosis. To this end, this paper offers tools that can help achieve these goals. Moreover, we are hopeful our field can cumulatively agree no study is perfect and therefore take steps toward building systematic knowledge.

ORCID iDs

Aaron D. Hill  <https://orcid.org/0000-0002-9737-1718>

Ernest H. O'Boyle  <https://orcid.org/0000-0002-9365-1069>

Notes

1. Endogeneity in reference to bias created by violating the exogeneity assumption is related but conceptually distinct from the concept of an endogenous variable as used in a SEM framework. In SEM, an endogenous variable is one that is "caused" by another variable in the model (i.e., has an arrow pointed to it) and is accompanied by estimation of a disturbance or error term.
2. The full list of sources is available from the first author while a selection of the most relevant sources appears in Table 5 and the annotated bibliography in Online Appendix A.
3. The results from the alternate search terms are not included in the count for Tables 3 or 4. Results of this coding and further analysis are available by request.

References

- Angrist, J. D., & Imbens, G. W. 1995. *Identification and estimation of local average treatment effects*. National Bureau of Economic Research.
- Angrist, J. D., & Krueger, A. B. 1999. Empirical strategies in labor economics. In Eds. A. Ashenfelter and D. Card *Handbook of labor economics*: 1277-1366. Elsevier.
- Angrist, J. D., & Pischke, J.-S. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, Princeton, NJ.
- Angrist, J. D., & Pischke, J.-S. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2): 3-30.
- Antonakis, J., Bastardo, N., & Rönkkö, M. 2019. On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, 1094428119877457.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. 2010. On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21: 1086-1120.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. 2014. Causality and endogeneity: Problems and solutions. *The Oxford Handbook of Leadership and Organizations*, 93.
- Arellano, M., & Bond, S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58: 277-297.
- Athey, S., & Imbens, G. W. 2006. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74: 431-497.
- Ballinger, G. A. 2004. Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, 7: 127-150.
- Bascle, G. 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, 6: 285-327.
- Basman, R. L. 1960. On finite sample distributions of generalized classical linear identifiability test statistics. *Journal of the American Statistical Association*, 55: 650-659.
- Baum, C. F., Schaffer, M. E., & Stillman, S. 2003. Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, 3: 1-31.
- Beatty, A. S., Barratt, C. L., Berry, C. M., & Sackett, P. R. 2014. Testing the generalizability of indirect range restriction corrections. *Journal of Applied Psychology*, 99: 587.

- Becker, T. E. 2005. Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, 8: 274-289.
- Bellemare, M. F., Masaki, T., & Pepinsky, T. B. 2017. Lagged explanatory variables and the estimation of causal effects. *Journal of Politics*, 79: 949-963.
- Bergh, D. D. 1993. Don't "waste" your time! The effects of time series errors in management research: The case of ownership concentration and research and development spending. *Journal of Management*, 19: 897-914.
- Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P. R., Lau, C. L. L., & Joo, H. 2016. Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal*, 37: 477-497.
- Bernerth, J. B., & Aguinis, H. 2016. A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, 69: 229-283.
- Bertrand, M., Duflo, E., & Mullainathan, S. 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119: 249-275.
- Bhave, D. P. 2014. The invisible eye? Electronic performance monitoring and employee job performance. *Personnel Psychology*, 67: 605-635.
- Bliese, P. D., Schepker, D. J., Essman, S. M., & Ployhart, R. E. 2020. Bridging methodological divides between macro-and microresearch: Endogeneity and methods for panel data. *Journal of Management*, 46: 70-99.
- Blundell, R., & Bond, S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87: 115-143.
- Blundell, R., & Bond, S. 2000. GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews*, 19: 321-340.
- Blundell, R., & Dias, M.C. 2009. Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44: 565-640.
- Bollen, K. A. 1996. An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, 61: 109-121.
- Bollen, K. A. 2012. Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38: 37-72.
- Bollen, K. A. 2019. Model Implied Instrumental Variables (MIIVs): An alternative orientation to structural equation modeling. *Multivariate Behavioral Research*, 54: 31-46.
- Bollen, K. A., & Bauer, D. J. 2004. Automating the selection of model-implied instrumental variables. *Sociological Methods & Research*, 32: 425-452.
- Bollen, K. A., & Maydeu-Olivares, A. 2007. A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, 72: 309.
- Bound, J., Brown, C., & Mathiowetz, N. 2001. Measurement error in survey data. In *Handbook of Econometrics*: 3705-3843: Elsevier.
- Breaugh, J. A. 2008. Important considerations in using statistical procedures to control for nuisance variables in non-experimental studies. *Human Resource Management Review*, 18: 282-293.
- Bromiley, P., Rau, D., & Zhang, Y. 2017. Is R & D risky? *Strategic Management Journal*, 38: 876-891.
- Caliendo, M., & Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22: 31-72.
- Campbell, D. T., & Stanley, J. C. 2015. *Experimental and quasi-experimental designs for research*. Ravenio Books.
- Certo, S. T., Busenbark, J. R., Kalm, M., & LePine, J. A. 2020. Divided we fall: How ratios undermine research in strategic management. *Organizational Research Methods*, 23(2): 211-237.
- Certo, S. T., Busenbark, J. R., Woo, H. S., & Semadeni, M. 2020. Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, 37: 2639-2657.
- Certo, S. T., Withers, M. C., & Semadeni, M. 2017. A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal*, 38: 1536-1556.
- Chatterji, A. K., Findley, M., Jensen, N. M., Meier, S., & Nielson, D. 2016. Field experiments in strategy research. *Strategic Management Journal*, 37: 116-132.
- Clougherty, J. A., Duso, T., & Muck, J. 2016. Correcting for self-selection based endogeneity in management research: Review, recommendations and simulations. *Organizational Research Methods*, 19: 286-347.
- Cole, D., Ciesla, J., & Steiger, J. 2007. The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12: 381-398.

- Conley, T. G., Hansen, C. B., & Rossi, P. E. 2012. Plausibly exogenous. *Review of Economic Statistics*, 94: 260-272.
- Cortina, J. 2002. Big things have small beginnings: An assortment of 'minor' methodological misunderstandings. *Journal of Management*, 28: 339-262.
- Dehejia, R., & Wahba, S. 2002. Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics*, 84: 151-161.
- Dodge, Y. 2006. *The Oxford dictionary of statistical terms*. Oxford University Press.
- Durbin, J. 1954. Errors in variables. *Revue de l'institut International de Statistique*, 22: 23-32.
- Evans, M. G. 1985. A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis. *Organizational Behavior and Human Decision Processes*, 36: 305-323.
- Fair, R. C. 1970. The estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors. *Econometrica: Journal of the Econometric Society*, 38(3): 507-516.
- Fisher, C. D., & To, M. L. 2012. Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior*, 33: 865-877.
- Fornell, C., & Larcker, D. F. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18: 39-50.
- Frank, K. A. 2000. Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research*, 29: 147-194.
- Fromkin, H. L., & Streufert, S. 1976. Laboratory experimentation. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*: 415-465. Chicago: Rand McNally.
- Frost, P. 1979. Proxy variables and specification bias. *The Review of Economics and Statistics*, 61: 323-325.
- Gates, K. M., Fisher, Z. F., & Bollen, K. A. 2019. Latent variable GIMME using model implied instrumental variables (MIIVs). *Psychological Methods*, 25: 227-242.
- Grant, A. M., & Wall, T. D. 2009. The neglected science and art of quasi-experimentation: Why-to, when-to, and how-to advice for organizational researchers. *Organizational Research Methods*, 12: 653-686.
- Greco, L. M., O'Boyle, E. H., Cockburn, B. S., & Yuan, Z. 2018. Meta-analysis of coefficient alpha: A reliability generalization study. *Journal of Management Studies*, 55: 583-618.
- Greenberg, J., & Tomlinson, E. C. 2004. Situated experiments in organizations: Transplanting the lab to the field. *Journal of Management*, 30: 703-724.
- Griffin, R., & Kacmar, K. M. 1991. Laboratory research in management: Misconceptions and missed opportunities. *Journal of Organizational Behavior*, 12: 301-311.
- Griliches, Z. 1977. Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, 45: 1-22.
- Griliches, Z., & Hausman, J. A. 1986. Errors in variables in panel data. *Journal of Econometrics*, 31: 93-118.
- Hahn, J., Todd, P., & Van der Klaauw, W. 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69: 201-209.
- Hamilton, B. H., & Nickerson, J. A. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization*, 1: 51-78.
- Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 50: 1029-1054.
- Harrison, G. W., & List, J. A. 2004. Field experiments. *Journal of Economic literature*, 42: 1009-1055.
- Hausman, J. A. 1977. Errors in variables in simultaneous equation models. *Journal of Econometrics*, 5: 389-401.
- Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 46: 1251-1271.
- Heckman, J. J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5: 475-492.
- Hill, A. D., Kern, D. A., & White, M. A. 2012. Building understanding in strategy research: The importance of employing consistent terminology and convergent measures. *Strategic Organization*, 10: 187-200.
- Hill, A. D., Kern, D. A., & White, M. A. 2014. Are we overconfident in executive overconfidence research? An examination of the convergent and content validity of extant unobtrusive measures. *Journal of Business Research*, 67: 1414-1420.
- Hunter, J. E., Schmidt, F. L., & Le, H. 2006. Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91: 594-612.

- Imbens, G. W., & Lemieux, T. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142: 615-635.
- Kennedy, P. 2008. *A guide to econometrics*. Malden, MA: Blackwell.
- Kenny, D. A. 1979. *Correlation and causality*. New York, NY: Wiley.
- Krause, M. S., & Howard, K. I. 2003. What random assignment does and does not do. *Journal of Clinical Psychology*, 59: 751-766.
- Landis, R., Edwards, B. D., & Cortina, J. 2009. Correlated residuals among items in the estimation of measurement models. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends*: 195-214. New York, NY: Routledge.
- Larcker, D. F., & Rusticus, T. O. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, 49: 186-205.
- Lee, D. S., & Lemieux, T. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48: 281-355.
- Li, M. 2013. Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods*, 16: 188-226.
- Lindell, M. K., & Whitney, D. J. 2001. Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, 86: 114.
- MacKinnon, D. P., & Pirlott, A. G. 2015. Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, 19: 30-43.
- Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, 25: 243-258.
- Maynard, M. T., Luciano, M. M., D'Innocenzo, L., Mathieu, J. E., & Dean, M. D. 2014. Modeling time-lagged reciprocal psychological empowerment-performance relationships. *Journal of Applied Psychology*, 99: 1244.
- McCallum, B. T. 1972. Relative asymptotic bias from errors of omission and measurement. *Econometrica*, 40: 757-758.
- Newey, W. K., & West, K. D. 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28: 777-787.
- Oster, E. 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37: 187-204.
- Pan, W., & Frank, K. A. 2003. A probability index of the robustness of a causal inference. *Journal of Educational and Behavioral Statistics*, 28: 315-337.
- Papies, D., Ebbes, P., & Van Heerde, H. J. 2017. Addressing endogeneity in marketing models. In P. S. H. Leeftang, J. E. Wieringa, T. H. A. Bijmolt, & K. H. Pauwels (Eds.), *Advanced methods for modelling marketing*: 581-627. Switzerland: Springer.
- Peel, M. J. 2014. Addressing unobserved endogeneity bias in accounting studies: Control and sensitivity methods by variable type. *Accounting and Business Research*, 44: 545-571.
- Pfeffer, J. 1993. Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, 18: 599-620.
- Pei, Z., Pischke, J. S., & Schwandt, H. 2019. Poorly measured confounders are more useful on the left than on the right. *Journal of Business & Economic Statistics*, 37: 205-216.
- 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*, 88: 879.
- 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: 539-569.
- Podsakoff, P. M., & Podsakoff, N. P. 2019. Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *The Leadership Quarterly*, 30: 11-33.
- Reed, W. R. 2015. On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics*, 77: 897-905.
- Rosenbaum, P. R., & Rubin, D. B. 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B* 45: 212-218.
- Sajons, G. 2020. Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *The Leadership Quarterly*, 31: 101348.

- Sande, J. B., & Ghosh, M. 2018. Endogeneity in survey research. *International Journal of Research in Marketing*, 35: 185-204.
- Sargan, J. D. 1958. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, 26: 393-415.
- Schaller, T. K., Patil, A., & Malhotra, N. K. 2015. Alternative techniques for assessing common method variance: An analysis of the theory of planned behavior research. *Organizational Research Methods*, 18: 177-206.
- Schmidt, J. A., & Pohler, D. M. 2018. Making stronger causal inferences: Accounting for selection bias in associations between high performance work systems, leadership, and employee and customer satisfaction. *Journal of Applied Psychology*, 103: 1001-1018.
- Semadeni, M., Withers, M. C., & Certo, S. T. 2014. The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal*, 35: 1070-1079.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shaver, J. M. 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science*, 44: 571-585.
- Shaver, J. M. 2019. Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management*. Advance online publication. doi:10.1177/0149206319846272.
- Shook, C. L., Ketchen, D. J., Jr., Hult, G. T. M., & Kacmar, K. M. 2004. An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal*, 25: 397-404.
- Shugan, S. M. 2004. Endogeneity in marketing decision models. *Marketing Science*, 23: 1-3.
- Siemsen, E., Roth, A., & Oliveira, P. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods*, 13: 456-476.
- Spector, P. E., & Brannick, M. T. 2011. Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, 14: 287-305.
- Stock, J. H., Wright, J. H., & Yogo, M. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20: 518-529.
- Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 25: 1.
- Suddaby, R. 2010. Editor's comments: Construct clarity in theories of management and organization. *Academy of Management Review*, 35: 346-357.
- Terza, J. V., Basu, A., & Rathouz, P. J. 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27: 531-543.
- Thistlethwaite, D. L., & Campbell, D. T. 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51: 309.
- Villas-Boas, J. M., & Winer, R. S. 1999. Endogeneity in brand choice models. *Management Science*, 45: 1324-1338.
- West, S. G., & Hepworth, J. T. 1991. Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, 59: 609-662.
- Williams, L. J., & O'Boyle, E. H. 2015. Ideal, nonideal, and no-marker variables: The confirmatory factor analysis (CFA) marker technique works when it matters. *Journal of Applied Psychology*, 100: 1579.
- Wiseman, R. M. 2009. On the use and misuse of ratios in strategic management research. *Research Methodology in Strategy and Management*, 5: 75-110.
- Wolfford, S. E., & Siegel, J. 2019. Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal*, 40: 432-462.
- Wooldridge, J. M. 1997. On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters*, 56: 129-133.
- Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT Press.
- Xu, R., Frank, K. A., Maroulis, S. J., & Rosenberg, J. M. 2019. konfound: Command to quantify robustness of causal inferences. *The Stata Journal*, 19: 523-550.
- York, J. G., Vedula, S., & Lenox, M. J. 2018. It's not easy building green: The impact of public policy, private actors, and regional logics on voluntary standards adoption. *Academy of Management Journal*, 61: 1492-1523.
- Zafarian, G., Kadile, V., Henneberg, S. C., & Leischnig, A. 2017. Endogeneity bias in marketing research: Problem, causes and remedies. *Industrial Marketing Management*, 65: 39-46.