Journal of Operations Management 37 (2015) v–viii

Contents lists available at ScienceDirect

Journal of Operations Management

journal homepage: www.elsevier.com/locate/jom

Editorial

Notes from the Editors: Redefining some methodological criteria for the journal☆

There are many things in which Operations Management (OM) researchers can take pride. Since the inception of empirical OM, we have rigorously incorporated measurement reliability and validity into our analyses. In many respects, the OM literature is a few steps ahead of its sister disciplines — incorporating measurement error into analyses is perhaps the best example. We have also made considerable progress in terms of theory development, whether by way of case research or purely conceptual and theoretical analysis. Finally, recent developments in the area of problem solving and design science demonstrate that OM scholars are genuinely interested in solving actual managerial problems and remaining practically relevant. These are all reasons to celebrate the progress in empirical OM.

But there are a number of blind spots, many of which continue to be reasons for rejections in the manuscript review process. The purpose of this editorial is to describe some of these issues. Specifically, there are a number of misunderstandings about some of the key methods used in manuscripts submitted to us. There are also some outdated practices that we want to discourage authors from using in their manuscripts. These issues are discussed in this editorial, in a roughly descending order of importance.

1. It is time to take causality seriously

We all know correlation does not establish causality. It is high time we do something about this. We constantly get manuscripts — based on cross-sectional surveys in particular — where the authors make causal claims. We no longer send to the review process manuscripts that uncritically interpret a cross-sectional correlation of X and Y as support of a causal claim, or more mildly, that the variance of X is driving the variance of Y. This applies to both econometric and structural equation models.

The problem with assuming that the variance of X drives the variance of Y is well documented. Ignoring the problem often results in over-permissive tests of substantive hypotheses: we see evidence for our hypotheses even when there is not any.

We now require all authors to take steps — theoretical or empirical, preferably both — to address *the problem of endogeneity*. This is now a standard practice in most top-tier management journals, and it is time for JOM, as a premier operations management journal, to follow suit. The literature on endogeneity is massive, going back almost a hundred years. Roberts and Whited (2013) offer a comprehensive summary of the key issues in the context of corporate finance research. All the issues discussed are directly applicable to OM research as well.

In a nutshell, the problem of endogeneity is this: when a researcher is using non-experimental data to test the hypothesis that X has an effect on Y, it is possible that the variance of X is not exogenous but endogenous to the model. The end result is that the model is misspecified. This in fact applies not just to cross-sectional but even longitudinal research. Even if X is measured at *t-1* and Y at *t*, there could be an unobserved variable Z that affects X and *t-1* and Y at *t*.

In a recent manuscript submitted to us, authors hypothesized that organizational integration drives employee commitment. Integration was assumed exogenous to commitment. This is a very problematic assumption, because we have many reasons to believe commitment could easily drive integration, making the variance of organizational integration indeed endogenous to the model. The consequence of endogeneity is asymptotic bias in parameter estimation.

We must come to terms with the fact that plausible claims about the direction and magnitude of an effect cannot rest on an analysis that completely ignores endogeneity. If our inferences are to be biased, they need to be biased toward being *conservative*. The problem of endogeneity often has just the opposite effect, it inflates our results. We are not aware of any scientific principles that warrant the use of over-permissive inference.

1.1. What can you do about this?

Examination of endogeneity starts with a simple question: What is the source of the variance in the exogenous variables in my model? So far JOM authors have been allowed simply *to declare* that these sources are exogenous to the model. Authors must take steps toward either *demonstrating exogeneity* or *correcting for endogeneity*. Both approaches have the common denominator: they call for addressing assumptions that have thus far gone untested.

Endogeneity can probably never be completely eliminated from empirical analysis, and it is well known that many "solutions" create more problems than they solve (Murray, 2006). But there are no

☆ We thank John Antonakis, Mikko Rönkkö, Aleda Roth, Fabrizio Salvador, and Suzanne de Treville for useful comments on an earlier draft of this editorial. All errors and omissions are obviously our own.

http://dx.doi.org/10.1016/j.colsurfb.2014.12.044

good reasons to avoid tackling the issue, at least theoretically. If the problem of endogeneity cannot be addressed empirically by testing for it or using instrumental variables or an experimental research design to mitigate it (Roberts and Whited, 2013), we expect at least a theoretical treatment of the topic in all JOM submissions where the general claim that one variable induces variance in another is made. When arguing that the variance of X gives rise to the variance of Y (causally or otherwise), we expect to see a plausible argument that the direction is indeed from X to Y, not vice versa, or perhaps caused by an omitted variable. Measurement error can also cause an endogeneity problem: if X and Y have a common measurement error source, X will unavoidably correlate with the error term of Y. Finally, sample selection bias may lead to problems very similar to that of endogeneity (Heckman, 1979).

While there is definitely a time and a place for cross-sectional research, we strongly encourage cross-sectional researchers to rethink their research designs. We all know how difficult it is to get longitudinal data, but prospective JOM authors must push themselves on this issue and try to fix at least some of the problems of past research by getting out of their comfort zone. If we want to know the magnitude of the effect X has on Y, cross-sectional data is almost guaranteed not to give us a valid estimate. Not only the principles of scientific rigor but also those of practical relevance demand that we get the magnitude right.

2. Know which rules are worth following

Many authors continue to build their arguments on the premise that application of statistical inference boils down to rule following. One of the most commonly found "rules of thumb" in manuscripts submitted to us is the claim that a measure is internally consistent if Cronbach's alpha exceeds .70. Nunnally's (1994) book *Psychometric Theory* is typically cited as the source. But to attribute the rule to Nunnally is a tell-tale sign one has not actually read Nunnally, because if anything, he claimed just the opposite: the criteria for adequate reliability always depend on the context. Lance et al. (2006) unambiguously debunk the ".70 rule." The technical details of the argument can be found in the works cited in this editorial; there is no need to reproduce them here.

If a manuscript submitted to us makes extensive use of unsubstantiated, non-inferential "rules of thumb," we will desk reject the manuscript. We say *non-inferential*, because it is crucial to make a distinction between rules that directly link to an inferential test and those that do not. Model fit in structural equation modeling is a good example. Consider two common tools for assessing model fit: the omnibus chi-square test and the Comparative Fit Index (CFI). The chi-square test is an inferential procedure: if the statistic is statistically significant, the model does not fit the data in the sense that the observed and predicted covariance matrices do not match. This is solid inference and methodologically acceptable reasoning. But the claim that a CFI > 0.95 means the model fits the data is not. This is because CFI is a descriptive index, not a test statistic with a commensurate inferential test. A high CFI value simply means the focal model fits the data better than the baseline model. What is the baseline model? It is typically the model where all measured variables are assumed uncorrelated. Using such a baseline model is dubious, because we already know it provides horrible fit for the data. All the CFI thus tells you is that your model fits the data better than a model that does not fit the data at all. It is difficult to see the insight in this conclusion. Lance et al. (2006) discuss the issue in detail, and Tanaka (1993) provides structural equation modelers with a great overview of SEM model fit.

2.1. What can you do about this?

Again, tackling the issue starts with an elementary statistical question: in applying any statistical test, what are the hypotheses being tested? Consider internal consistency of measurement:

H_0: the measurement instrument is not internally consistent
H_1: the measurement instrument is internally consistent

Using the "alpha > .70″ rule can help reject the null, but this is not an inferential test, it is merely a social convention that has no methodological basis. What is more, applying the rule misconstrues what methodological texts have actually said.

Consider second SEM model fit under the conventional formulation (Bollen, 1989):

H_0: the model fits the data
H_1: the model does not fit the data

The chi-square omnibus test fares much better than the "alpha > .70″ rule. The chi-square test is a valid inferential procedure (it is a *test* that produces a *p*-value). A rigorous modeler would consider the fact that the null means the model fits the data, which means low statistical power works to the advantage of the model, not against it. This leads to an over-permissive test, and sometimes this can present a problem. A skillful researcher is able to examine whether or not this is cause for concern. Of course, it is possible to reformulate the null and the alternative hypotheses such that over-permissiveness is not a problem.

As far as author requirements, authors of JOM submissions must exhibit an understanding of which rules have a basis in formal statistical inference and which do not. Here, a very simple litmus test works very well: Does the procedure I am using produce a test statistic (with a *p*-value) or not? At the very minimum, we expect authors to know which rules are simply "urban legends." This is crucial, because many of the cutoff criteria cited by OM researchers have been thoroughly discredited in the methods literature (Cortina, 2002; Lance, 2011; Lance et al., 2006; Lance and Vandenberg, 2008; Spector and Brannick, 2011). Prospective JOM authors must make themselves aware of this important literature. Citing "an urban legend" will likely lead to desk rejection of the manuscript.

Instead of relying on "rules of thumb," we encourage authors *to contextualize* their measurement. Indeed, this is what methodological authorities such as Nunnally actually recommend (e.g., Nunnally and Bernstein, 1994, p. 249). Suppose you are interested in estimating a regression model with two explanatory variables ($x1$ and $x2$) and a dependent variable (y), and you are assessing measurement reliability. By contextualization we mean asking the question: How does measurement error in my variables affect estimation? It is well known that measurement error in an independent variable is more problematic than in the dependent variable. One can think of measurement error as one of the components of the regression error term, therefore, measurement error in the dependent variable is implicitly already modeled. The statistical consequence of measurement error in the dependent variables is loss of efficiency, which typically does not create problems, particularly if the sample is of reasonable size. Measurement error in the independent variables, in turn, likely causes asymptotic bias to estimates. Although there are no hard and fast rules on the consequence, the resultant bias is roughly proportional to the amount of measurement error (Kennedy, 2008; Maddala, 1988). Increasing sample size does not fix the problem, because bias is asymptotic. How many authors citing the "alpha > .70 rule" for an independent variable realize that they are implicitly admitting that an asymptotic bias of up to 30 percent in a parameter estimate is acceptable? How much sense