Book Review

Counterfactuals and Causal Inference: Methods and Principles for Social Research, Stephen L. Morgan & Christopher Winship. New York: Cambridge University Press, 2007, 319 pages, $28.99 (Softcover)

Reviewed by John Antonakis and Rafael Lalive

Faculty of Business and Economics

University of Lausanne

Putting Causality Into Structural Equation Modeling

> "Most quantitative empirical analyses are motivated by the desire to estimate the causal effect of an independent variable on a dependent variable. Although the randomized experiment is the most powerful design for this task, in most social science research done outside of psychology, experimental designs are infeasible" (Winship & Morgan, 1999, p. 659).

The above quote from earlier work by Winship and Morgan, which was instrumental in setting the groundwork for their book, captures the essence of our review of Morgan and Winship's (2007) book: It is about causality in non-experimental settings. In a similar vein, our review "began" a few years ago, too. The first author of this review (John) was one of the members of the professorial selection committee of the second author (Rafael); John found Rafael's job talk intriguing. Rafael used a methodological approach that apparently produced causal estimates in the context of a non-experimental setting. Coming from a background in psychology, John was puzzled and surprised by Rafael's results and the confidence with which Rafael made causal claims. Rafael, an economist, argued that the non-experimental method he used mimicked an experiment even though Rafael had not randomly assigned anyone to treatment or control conditions.

How could Rafael defy the logic of random assignment yet claim to produce causal estimates? John, shaking his head in disbelief, chanted what was at that time his mantra: "correlation is not causation, correlation is not causation…." Yet, after Rafael's patient explanations during the job talk and after John took the time to understand the method Rafael used, John quickly changed his tune as he became convinced that Rafael was right. The topic of causality has since been a major discussion between John and Rafael.

What helped John better understand causality in nonexperimental settings was the counterfactual account of causality (Morgan & Winship, 2007; Winship & Morgan, 1999). Since our first discussions, we have both bought this book and Rafael has actually assigned it as compulsory reading in an advanced econometrics course (so we already voted on this book with our Swiss Francs!). Although John still uses experiments (e.g., Antonakis & Dalgas, 2009), part of his current work is far from experimental and yet it is at the doorstep of causality. He, too, now includes healthy doses of counterfactual-type thinking in the structural equation modeling course he teaches.

For those interested to know, Rafael used a regression discontinuity design (see Lalive, 2008); he went to publish this work in a special issue on regression discontinuity in a top economics journal (see Cook, 2008). Describing the workings of this estimator is beyond the scope of the review (and be it known that this estimator has impressed John so much that two of his graduate students are writing papers using a regression discontinuity design!). However, what is really important to understand before discussing the Morgan and Winship book is a persistent problem that most structural equation modelers doing nonexperimental research face: This problem is that of *Endogeneity*.

Many modelers have never heard of this term; those who have, particularly if they work outside economics, do not understand the consequences of endogeneity's insipient effects. They know of endogenous variables and simply assume them to be modeled as consequences of other, exogenous variables. However, if these other exogenous variables are not truly exogenous (i.e., vary randomly, independent of other causes) the modeler will estimate a model that cannot be interpreted.

The importance of this book will quickly become evident once readers understand the problem of endogeneity and the counterfactual account of causality. We thus take the time to first provide a summary overview of Morgan and Winship's counterfactual framework. Thereafter, we provide a summary of the book's chapters and evaluate it. As will be evident in our review, we believe that the Morgan and Winship book is an important and useful book that social science researchers and particularly structural-equation modelers who undertake nonexperimental research, should read. This book should appeal to such researchers, whether they are seasoned or aspiring professionals.

Summary of the book: Background to the counterfactual

A nice way of motivating the contribution of this book is via the following quotation:

> "In the counterfactual modelling tradition, attention is focused on estimating various average causal effects, by analysis of the values y_i, for groups of individuals defined by specific characteristics. To do so effectively, *the process by which individuals of different types are exposed to the cause of interest must be modelled* [italics ours]. Doing so involves introducing defendable assumptions that allow for the estimation of the average unobservable counterfactual values for specific groups of individuals. If the assumptions are defendable, and a suitable method for constructing an average contrast from the data is

> chose, then an average difference in the values $y_i$of can be given a causal interpretation" (Morgan and Winship, 2007, p. 6).

The above quotation is key to understanding the problems of selection and endogeneity. We briefly explain what is meant by these terms while we summarize key ideas and examples from the book, focusing particularly on the Morgan and Winship "counterfactual" framework. In fact, thinking in terms of counterfactuals is, for us, essential to understanding causality; we thus take the time to thoroughly explain the counterfactual account of causality to readers.

In a structural equation or regression model, variables can be endogenous or exogenous. Endogenous variables are determined by other variables (or the error term) in the system of equations. There are also exogenous variables, that is, variables that vary independently of other causes in the model. We will discuss the simple cases of simultaneous equation models with observed variables. However, what we discuss, and the recommendations made in this book, are equally relevant for structural equation models with latent variables as well.

Knowing whether a variable is exogenous or endogenous, and then modeling the system of equations correctly, is the most important factor for determining whether or not model estimates will make any sense. By "sense" we mean that the estimates are consistent, that is, that they reflect the true (causal) relation between a supposed cause, *x*, and an effect, *y*. A consistent estimate converges to the true estimate with an increasing sample size. However, an inconsistent estimate does not have this desirable property. Morgan and Winship are rightly concerned with consistency of estimation and they hammer at this point *again and again* in the book. Efficiency (i.e., having smaller estimations of the variance) is, of course, important too; however, as Morgan and Winship mention, there is no point is producing efficient estimates when they are biased. Unfortunately, many applied researchers in social sciences are unaware of the problem of consistency. It is our hope that this book, and our review, which strongly endorses this book, will help to correct this sad state of affairs.

To understand the problem of consistency we discuss omitted variable bias, a topic first introduced in the book in Chapter 1 (see page 11) and then discussed in more detail in Chapter 5. Assume the following basic and correct specification:

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{1}$$

As the book suggests, an estimator, whether maximum likelihood (ML) or ordinary least squares (OLS) assumes that *x* is orthogonal to the error term *e*. The error term includes all sources of variance in *y* that are not accounted for by *x*. Orthogonality with the error term is guaranteed in experimental research: There is no variable that could be modeled that would correlate with the treatment, *x*, and also correlate with *y*. In the case of nonexperimental research, the problem the modeler faces is the possibility of omitted variables, which would make *x* correlate with the *e* term. What does this correlation mean, precisely, and what are its consequences? This book seeks to answer this question and also to provide solutions to this problem.

In the context of an experiment with two groups (as captured by the dummy variable *x* above), the individuals who have been assigned to the treatment group (coded 1) and control group (coded 0) have been assigned using a random process. A key condition of the OLS or ML estimator, the orthogonality of *x* with the error term, is thus satisfied. Because of this assignment process, the individuals constituting the two groups are, on average, approximately equal on all observed or unobserved characteristics. Random assignment ensures this outcome because each individual has the same probability (.50) of being assigned to the treatment or control group. If the sample is sufficiently large, and given the variation in characteristics of samples of individuals, we should observe, on average, that the two groups are interchangeable (within sampling error). Now here comes the key idea of the book: *The counterfactual*. Because the sample of individuals in each group are approximately equal at pretreatment, we can observe the counterfactual: That is, we can observe what the treatment group, on the average, would have received on *y* had it not been treated, and we can equally observe, on the average, what the control group would have had on *y* had it been treated. Of course, as Morgan and Winship mention, the counterfactual is not directly observed in the sense that we do not observe what *y* of individual *i* would have been had she not received the treatment. The counterfactual is constructed at the level of analysis at which the treatment is administered, that is, the group, which is possible given that the groups are approximately equivalent. Thus, the causal effect is simply the difference in the means of *y* for the two groups, $\bar{y}_1 - \bar{y}_0$, and the reliability of this difference can be statistically estimated.

Morgan and Winship note that if, however, the treatment has not been assigned randomly and if this nonrandom selection process is not explicitly modeled, then *x* will correlate with *e*. The groups are not interchangeable anymore; the counterfactual cannot be observed. For example, suppose that some process affected how general intelligence was distributed between the two groups and suppose that individuals in the treatment group are on average more intelligent than those in the control group. Suppose also that being more intelligent predicts *y* and that the treatment had an effect. In this situation, the slope of *x* cannot be correctly estimated because it includes the effect of the treatment and that of IQ. Thus, with IQ omitted from the model, the slope of *x* will be biased to the extent that *x* correlates with the omitted cause and the cause correlates with *y*. Just how bad is this bias?

One will not know unless one includes all omitted causes (or if one uses some other procedure, as detailed in the book). The book looks at different ways in which endogeneity can raise its ugly head. Morgan and Winship's methodological *tour de force* highlights methods that are useful for reconstructing the counterfactual in the case where the regressors correlate with the error term. We summarize each of the book's chapters next.

Summary of book chapters

The book begins with Part I (Ch. 1 & 2) by first introducing the counterfactual framework, which we have made explicit above, and provides tangible examples throughout the text to demonstrate why counterfactuals are necessary for causality. In all sections where Morgan and Winship introduce important statistical concepts in the context of the counterfactual framework, they are very generous in giving credit to those who made major contributions to providing the scaffolding for this framework, including Donald Rubin, James Heckman, Donald Campbell, and others. The historical overview they provide, and how they synthesize it to explain modern thinking on the counterfactual framework is commendable.

In Part II, Morgan and Winship introduce causal methods designed to address simple problems with causal reasoning. The key assumption that makes these methods work is that the cause is randomly allocated to units that are the same in terms of some observed characteristic. If this characteristic takes on few values, the method of conditioning serves to identify causal effects (Ch. 3). The prime example from the literature is a randomized class size experiment that was conducted among 79 schools in the state of Tennessee in the early 1980s (Krueger & Whitmore, 2001). Under the experimental protocol each school was required to open at least one small class, one regular size class with teacher aide, and one regular size class serving to estimate the counterfactual. Schools differed with respect to the number of small size classes they opened depending on grade enrollment and school finances. Simply comparing students taught in small classes to students taught in regular size classes across the entire experiment is therefore misleading. The method of conditioning asks researchers to contrast students taught in small classes to students taught in regular size classes at the same school. This within school contrast identifies the effect of reducing class size for each school. Conditioning does not work, however, if the set of conditioning variables contains several continuous characteristics, for instance age and work experience. Attempts to condition are faced with the problem of the "curse of dimensionality." Chapter 4 discusses the method of matching on the propensity scores -- the probability of receiving treatment (along with other methods of matching). Matching on the propensity score addresses the curse because it compresses the full set of conditioning variables into a one dimensional index: The probability of receiving the treatment. Units that have the same probability of treatment but different exposure to treatment can be contrasted to identify the effect of treatment: This was the key contribution of a groundbreaking paper by Rosenbaum and Rubin (1983). Naturally, confounding variables can also be adjusted for by using multiple linear regression, the topic of Chapter 6. This chapter discusses how to specify the regression to measure the average causal effects of treatments.

Part III is where most of the meat of the book is. Chapter 6 introduces the problem of selection, which refers to the problem we discussed when introducing endogeneity. This chapter gently leads on the Chapter 7, which introduces "instrumental" variables. An instrumental variable is an exogenous variable that correlates with the problematic predictor q as per Equation 3 above. The instrument also correlates with y; however, the instruments only effects y via q. Because the instrument is exogenous, the predicted value of q has a unique property: It does not correlate with the error terms of the equations. Morgan and Winship explain in some detail the importance of instrumental-variable estimation techniques, which are the workhorse of econometrics. Chapter 8 is a continuation of the previous chapter, where Morgan and Winship make the case that explicit theory must be used to explain how an instrumental variable generates the supposed causal effect. Chapter 9 deals with time-series data and regression discontinuity designs.

Part IV is the concluding chapter. Morgan and Winship provide a discussion about the future of the counterfactual framework. Even though they are ardent proponents of it, they take a very sober, honest, and for us too much of a modest perspective of their contribution to causal reasoning. Their honesty, however, truly makes the readers see for themselves why estimate consistency should be the "α" and the "ψ" of research (we leave the "ω" for efficiency, which is important, too).

Strengths and Weaknesses of the Book

The book is very clearly written. The authors complement equations with intuitive path diagrams and provide ample and tangible examples throughout. The book nicely highlights the importance of causality and knowing how to demonstrate it by specifying the estimated model correctly so that the modeled variables do not correlate with omitted causes. We really appreciate how Morgan and Winship drummed the fact that a simple description of a relation between two variables is not very useful for research or society; in fact, we believe that such research should simply not be published. Unfortunately, we see the opposite happening in many social science journals. Researchers are still not "getting it!"

We just completed a major review paper and found that research in management and applied psychology is rife with endogeneity (Antonakis et al., in press). Such reviews have been done before having similar conclusions (Halaby, 2004; Hamilton & Nickerson, 2003). We really hope that those researchers and teachers who have not yet considered the problems of endogeneity will take a look at the Morgan and Winship book; as an hors d'oeuvre, they might want to start with one of the papers we have referenced above.

This book is a nice complement to a couple of other books that SEM researchers should have in their libraries. Along with Morgan and Winship, the essential basics include Angrist and Pischke (2008) and Shipley (2000); for those who are not fainted-hearted Pearl's (2009) book might also be of interest (Morgan and Winship refer to Pearl's work regularly). Together, these books go beyond books such as that by Bollen (1989), which is an outstanding book addressing technical aspects of estimation, but which pays scant attention to the essential basics of causality and the assumptions behind estimation procedures.

Of course, all objective reviews interrupt the dithyrambic eulogies with a bit of nitpicking. Although we are very positive about this book, we think that the book (or at least future versions of it) could cover a bit more ground. For example, Morgan and Winship do not discuss difference-in-differences models (Angrist & Krueger, 1999), which would be natural extension to their discussion on time-series models. They do not give as much attention as we would have like to regression discontinuity designs, which are the closest thing to randomized experiments (Cook, Shadish, & Wong, 2008; Shadish, Cook, & Campbell, 2002). They make little mention of panel models and the problems of random versus fixed effects models. In fact, we were quite surprised that they make no mention of the venerable Hausman (1978) endogeneity test. We would also have appreciated a more in-depth discussion of Heckman (1979) selection models. Also, they also do not discuss overidentification and how the correctness of a model is tested (see Basmann, 1960; Hansen, 1982; Sargan, 1958); this latter topic is crucial and it is interesting to note that unlike in other social sciences, there is no debate in econometrics about the utility of the chi-square test of model fit.

Finally, in some places of the book Morgan and Winship often jump into some rather sophisticated statistical concepts without explaining some basics; the authors assume that readers have the necessary background to understand certain advanced concepts and in the process we think that they will lose some readers who might not have advanced statistical training. For instance, in the case of omitted variable bias, why, from an algebraic perspective, are coefficients biased? A thorough explanation of this endogeneity problem would make the book more accessible to a larger audience. For example, assume the following model, which is the true model (Antonakis et al., in press):

$$y_i = \beta_0 + \beta_1 q_i + \beta_2 z_i + e_i \tag{2}$$

Now, assume that instead of the above model, one estimates a model where *z* is excluded:

$$y_i = \varphi_0 + \varphi_1 q_i + v_i \tag{3}$$

Because this model omits *z*, *q* may correlate with the error term *v* (which will be the case if *q* and *z* are correlated and *z* is a cause of *y*). In this case, instead of obtaining the unbiased estimate β_1 one obtains φ_1. Morgan and Winship do not take the time to explain why, specifically, these two estimates might be quite different. We feel it would have been useful to show some of the basic steps to demonstrate the problem at hand. To show how, we express *z* as a function of *q* and its unique cause *u*, and we omit the intercept for simplicity:

$$z_i = \gamma_1 q_i + u_i \tag{4}$$

We then substitute (4) into (2):

$$y_i = \beta_0 + \beta_1 x_i + \beta_2(\gamma_1 x_i + \ u_i) + e_i. \tag{5a}$$

Multiplying out gives*:*

$$y_i = \beta_0 + \beta_1 x_i + (\underbrace{\beta_2\gamma_1 x_i + \ \beta_2 u_i + e_i}_{v_i}) \tag{5b}$$

As is evident, *v* now correlates with *x*. Another way of looking at the problem is to rearrange the equation as a function of *x*:

$$y_i = \beta_0 + (\beta_1 + \beta_2\gamma_1)x_i + \ (\beta_2 u_i + e_i) \tag{5c}$$

It is now clear that the effect of *x* on *y*, as estimated by the slope β_1, was consistently estimated in (2); however, it is inconsistently estimated in (3) because as indicated in (5c), the slope will include the correlation of q with *z* (i.e., γ_1). When important causes have been omitted from the model, one does not estimate $\beta_1$as per (3), but something else (φ_1). This something else, $\varphi_1$could be higher or lower than the correct value, which of course depends on the signs of β_2 and γ_1. Only in the case of $\beta_2 = 0$ or $\gamma_1 = 0$ does v_i reduce to e_i, suggesting that omitting *z* the model does not affect the estimate of q. Such kinds of intuitive explanations could have made the book more accessible to readers.

To conclude, we are very confident that this book, as well as similar lines of research focusing on causal issues in nonexperimental settings, will be the future of social sciences research. One needs to reconstruct the counterfactual before causal effects can be correctly identified. Researchers should constantly be thinking in terms of: "*What would the treatment group have received on y had it not been treated*?" As for the importance of counterfactuals, we chuckled contemplating the following question, which we would like to pose to our future graduate students 25 years from now, just before we retire as professors: "*What would social sciences have been like had the counterfactual framework not been developed*?"

References

Angrist, J. D., & Krueger, A. B. (1999). Empirical strategies in labor economics. In O. C. Ashenfelter & D. Card (Eds.), Handbook of labor economics (Vol. 3, Part 1, pp. 1277-1366). Amsterdam: Elsevier.

Angrist, J. D., & Pischke, J.-S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton: Princeton University Press.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (in press). On making causal claims: A review and recommendations. The Leadership Quarterly, 21.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! Science, 323, 1183.

Basmann, R. L. (1960). On finite sample distributions of generalized classical linear identifiability test statistics. Journal of the American Statistical Association 55, 650-659.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Cook, T. D. (2008). "Waiting for Life to Arrive": A history of the regression-discontinuity design in psychology, statistics and economics. Journal of Econometrics, 142, 636-654.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. Journal of Policy Analysis and Management, 27, 724-750.

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. Annual Review of Sociology, 30, 507-544.

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, 50, 1029-1054.

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46, 1251-1271.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica, 47, 153-161.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college test-taking and middle school test results: Evidence from Project STAR. Economic Journal, Royal Economic Society, 111, 1-28.

Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. Journal of Econometrics, 142, 785-806.

Morgan, S. L., & Winship, C. (2007). Counterfactuals and causal inference: Methods and principles for social research. New York: Cambridge University Press.

Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). New York: Cambridge University Press.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. Econometrica, 26, 393-415.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

Shipley, B. (2000). Cause and correlation in biology: A user's guide to path analysis, structural equations, and causal inference. New York: Cambridge University Press.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. Annual Review of Sociology, 25, 659-706.