....New Books in Review

EDITOR: *George R. Franke*

ASSOCIATE EDITORS: *Naveen Donthu*

Meryl P. Gardner

Expanded versions of these reviews can be found on the AMA's Web site at http://marketingpower.com/jmr/.

CAUSALITY: MODELS, REASONING, AND INFERENCE, Judea Pearl, Cambridge, UK: Cambridge University Press, 2000, 384 pages, $39.95.

CAUSATION, PREDICTION, AND SEARCH, 2d ed., Peter Spirtes, Clark Glymour, and Richard Scheines, Cambridge, MA: MIT Press, 2000, 543 pages, $55.00.

> In the social sciences, there is a great deal of talk about the importance of "theory" in constructing causal explanations of bodies of data.... In many of these cases the necessity of theory is badly exaggerated.
> —Spirtes, Glymour, and Scheines 2000, pp. 97–98

> It is a capital mistake to theorize before you have all the evidence. It biases the judgment.
> —Sherlock Holmes in Sir Arthur Conan Doyle's (1887) *A Study in Scarlet.*

Users of structural equation modeling (SEM), a method that has also been known as causal modeling, have learned to tread lightly around the subject of causality. After all, the empirical raw material of SEM is typically a covariance matrix derived from nonexperimental data, and research dogma indicates that this is insufficient for making statements about cause-and-effect relations. At least as far back as Robert Ling's (1982) scathing review of David A. Kenny's (1979) book, *Correlation and Causality* (a classic text that is still well worth reading), users of SEM methods have found themselves on the defensive, careful not to claim too much. This, however, has produced something of a paradox. The models estimated with SEM clearly depict variable A as having an effect on variable B and distinguish between covariance relations and directional paths—that is, causal effects. Thus, SEM users propose structures that are causal but tend to disavow the causal element when they evaluate their results. Especially for the practitioner, the causal component is likely to be the point of the whole exercise: What a manager wants to know is, "If I do X, how will that change Y?"

Furthermore, reasonable people use causal language and reach causal conclusions all the time. The government releases economic statistics, the stock market subsequently moves, and the observer concludes that the new information moved the market. Rain falls, water drips through a hole in the roof, and an observer makes the connection. People do this without the aid of either experiments or sophisticated data analysis. True, sometimes the observers are wrong—for example, at one point medical researchers thought exposure to aluminum was a cause of Alzheimer's syndrome, but now they consider the linkage spurious. Nevertheless, people proceed through life making causal inferences from nonexperimental data.

Pearl's *Causality* and Spirtes, Glymour, and Scheines's (SGS's) *Causation, Prediction, and Search* (2d ed.) urge researchers to resolve the paradox by dropping the pretense and acknowledging the causal content of their models. A → B means more than that A is correlated with B. It means that B changes in response to changes in A, and the lack of an arrow in the opposite direction means that A does not change in response to changes in B. Returning to the path analysis roots of SEM, these authors argue that justification for causal interpretation of model parameters follows from satisfying basic criteria. These criteria have as much to do with proper sampling design as they do with modeling. However, these authors do not argue that their approach to inferred causation reveals "truth." Instead, as Pearl (pp. 47–48) writes, "It identifies the mechanisms we can plausibly infer from nonexperimental data; moreover, it guarantees that any alternative mechanism will be less trustworthy than the one inferred because the alternative would require more contrived, hindsighted adjustment of parameters (i.e., functions) to fit the data."

These books summarize intertwined research programs. Within the data mining and artificial intelligence community, Pearl and SGS are associated with machine learning and "Bayes networks," which look like structural equation models but are encountered in the context of discovery. The overall focus of the research has been on understanding how causal inferences are made and how they can be made reliably. One aim is to determine just how much analysts may discover from a data set using nothing more than a set of algorithms and common sense. In the data mining and knowledge discovery in databases literature, the issue arises because the organization has an excess of data, an adequate supply of computing power, and a shortage of analyst time.

Many SEM users have faced a similar dilemma. The researcher gains access to an attractive secondary data set, develops a theoretical model, finds variables in the data set that can serve tolerably as indicators of the latent variables, runs the analysis—and comes nowhere close to an acceptable fit. The analyst resorts to ex post modifications and

Journal of Marketing Research
Vol. XXXIX (February 2002), 137–140

indices of "approximate fit" and tries to emerge with something of value but is left uneasy by the contortions required to escape with even that much.

Perhaps, as the great detective suggests, the problem lies in imposing too much structure too soon. Instead of the deductive and confirmatory approach hardwired into SEM thinking, Pearl and SGS call for listening to the data to uncover the causal structure that generated it. Within the SEM community, such blatantly inductive behavior is likely to be dismissed as unreliable and virtually worthless, for a variety of reasons. Are these authors insensitive to or ignorant of those arguments? No. But instead of merely rejecting inductive analysis because of its limitations, these authors explore those limitations in detail, identifying what may be inferred from the data within those limitations. They further argue that some of those limitations are not as intractable as some may believe.

For example, one reason SEM users hesitate to make causal claims is the well-known problem of equivalent models (Stelzl 1986). If one model fits a set of data, there are likely to be other models that achieve equivalent fit—they imply exactly the same population covariance matrix—but lead to different substantive conclusions. The simplest example is a two-variable model. Given that A → B fits the data, so too will B → A. The first model suggests manipulating A to affect B, whereas the second indicates that manipulating A will have no effect on B. For SEM researchers, the logical implication is that they must be cautious in making causal statements about A. Pearl and SGS recognize the phenomenon of equivalent models, but instead of stopping there, they describe axiomatic, algorithmic procedures for uncovering the set of all equivalent models. Pearl in particular emphasizes finding paths or effects that are common to all members of this set, to which common features he is willing to ascribe a causal interpretation. (In Pearl's usage, the "identification" problem is not a matter of parameter estimability but one of finding these common features; SGS deal with this issue under the heading of "consistency.") Similarly, these authors recognize that statements about causal relations among variables are circumscribed by the set of variables in the model. Their response is to circumscribe their conclusions similarly and move on.

This body of research relies on a small core set of tools, axioms, and algorithms. One primary tool is graph theory. Graphs provide a language for representing causal claims. Graphical models consist of variables and links (or the absence of links) among them. Both Pearl and SGS focus on observed variables rather than the common factor latent variables of SEM, though SGS, in particular, devote a substantial amount of attention to discovery algorithms for latent variable measurement models. Pearl discusses "latent variables," but for the most part these are variables that have been excluded from the data set, rather than variables that are intrinsically unobservable. Links between variables may be directed or undirected. Undirected links, which represent residual correlation, are undesirable as an end state; as Pearl (p. 44) notes: "correlations that are not explained by common causes are considered spurious, and models containing such correlations are considered incomplete." The goal of the algorithms discussed in these books is to evolve from undirected to directed models, and the ideal goal is to achieve a directed acyclic graph that faithfully reproduces the empirical evidence but does not imply more constraints than the data can support. Axioms such as "faithfulness" and "minimality" guide the design of algorithms that are used to uncover causal structure. Although a directed acyclic graph offers the clearest causal interpretation, the graphical language used in this field accommodates models that retain some uncertainty.

Pearl and SGS favor acyclic (or recursive) graphs on the grounds of both principle and practicality. They fundamentally reject reciprocal causation or feedback loops as reflecting underlying causal structure. Instead, they treat this phenomenon as a by-product—a problem in sampling or measurement. Furthermore, it is a requirement of a causal model that each predictor variable can be set to a specific value—Pearl uses a special operator, do(x), to represent this manipulation—which leads to a predictable outcome. Reciprocal causation is inconsistent with this hypothetical manipulation.

As Pearl and SGS develop their models, their algorithms evaluate partial correlations and vanishing tetrads. As in SEM, the absence of a direct link between variables indicates a zero partial correlation between them, possibly conditional on intervening variables. (The conditional independence of correlated variables, given their "parent" or antecedent variables in the model, is called the Markov condition and is an essential requirement for a graph to be considered a causal model.) Within the SEM literature, Stelzl (1986) built his rule for locating equivalent models (see also Lee and Hershberger 1990) around the preservation of patterns of zero partial correlations.

Another key concept is the "vanishing tetrad difference," which SGS trace to Spearman. A tetrad difference, τ_{ABCD}, is a function of the correlations or covariances among four variables, A, B, C, and D:

$$\tau_{ABCD} = \sigma_{AB} \times \sigma_{CD} - \sigma_{AC} \times \sigma_{BD}.$$

A tetrad difference is "vanishing" when it is equal to 0. One way to evaluate a model or a data set is by searching for implied or actual vanishing tetrad differences. For example, imagine that the four variables all load exclusively on a single common factor and have uncorrelated errors. Such a model implies several vanishing tetrad differences, including τ_{ABCD}. Vanishing tetrad differences were an important tool in the early days of path analysis but were gradually displaced by maximum likelihood methods, though confirmatory tetrad analysis (Bollen and Ting 1993; see also Ting 1995) can be helpful when maximum likelihood methods fail. The SGS algorithms seek out vanishing tetrad differences in the data and then develop models that would produce these phenomena. One feature common to both vanishing tetrad differences and partial correlations is that neither requires the estimation of model parameters—both can be evaluated using nothing more than the input covariance matrix and candidate model structures. This means that these tools can be used to develop or evaluate models despite conditions such as underidentification that thwart parameter estimation.

What, then, is the role of theory in model development if researchers can use the algorithms described here to uncover causal models for groups of variables? Within the broader data mining community, the role of "prior knowledge" in modeling remains controversial. Groth (1998) argues that a

baggage of "assumptions" will only interfere with the discovery process. In contrast, Frawley, Piatetsky-Shapiro, and Matheus (1991, p. 19) cite prior domain knowledge as a key input to the discover process, declaring that "The best chance for discovery is with things we almost but do not quite know already." Even barring direct application of theory in model selection, there are many decisions that lead up to model selection that themselves may be guided by theory (Rigdon and Bacon 1997). Similarly, SGS endorse the role of theory in research design—only by employing some prior knowledge of the system can researchers expect to include all relevant causal variables in the data set.

Sampling issues are critical in this field. As both books explain, a sample that includes a mixture of representatives from different populations is likely to produce a covariance matrix that points to a saturated model. This will be true even if each population is correctly modeled with the same parsimonious causal structure, as long as the parameter values differ across populations. As a highlight of their review of confounding conditions, both books devote some attention to Simpson's paradox, "the phenomenon whereby an event C increases the probability of E in a given population p and at the same time decreases the probability of E in every subpopulation of p" (Pearl, p. 174). For example, a marketer distributes a coupon that increases the likelihood of choosing the marketer's brand across the whole market but reduces that likelihood within both urban and nonurban buyers. Simpson's paradox results from a spurious association between the effect and the treatment, which might occur if urban buyers, who are already more inclined to choose the marketer's brand, also have a greater likelihood of receiving the coupon. Pearl and SGS use their discussion of Simpson's paradox, which is something of an old chestnut in this field, to highlight the difference between evidence of probability and evidence of causality.

The two books are similar in several other ways. They both spend some time addressing their own field, which can be frustrating to the outsider. Both are easier to digest if the reader has a prior background in Bayesian methods. Both have problems with typos. (Errata and amplifying material for Pearl's book are available from his site, http://bayes.cs.ucla.edu/BOOK-2K/index.html.) And both can be used in conjunction with material the authors have provided online. Pearl's site provides slides and homework assignments specifically related to his book, and SGS have developed online materials for a course on causal and statistical reasoning (http://www.phil.cmu.edu/projects/csr/). Compared with Pearl's book, though, the one by SGS more clearly describes a course of instruction in the authors' area of interest. Pearl's book reads more like a collection of essays, which overlap but do not follow in linear fashion from introduction to advanced concepts. Pearl notes that the organization is somewhat chronological, recapping his team's pursuit of different topics in this area.

These books have other substantial differences because of their authors' various interests. Pearl's book is more "theoretical" (pardon the expression), focusing almost exclusively on population results and logical algebra. Pearl's style is also more brash—he labels the evolution of SEM's struggle with causality "bizarre"—whereas SGS's seems more measured. The greater presence of sample data and statistical issues in SGS's book is not surprising given their association with TETRAD, a computer package that encompasses the discovery algorithms discussed in this field. (Current information about TETRAD and additional materials about the authors and their research program are available at http://hss.cmu.edu/philosophy/TETRAD/tetrad.html.) A portion of SGS's book is devoted to describing applications of their software, recapping the superior performance of its simultaneous search procedure in simulation studies involving true but unknown underlying structures compared with the performance of the stepwise/hill-climbing exploratory options included in standard SEM packages. However, their book is not a manual by any means.

CONCLUSION

These two books use different language to describe the same concepts and phenomena, and readers are well advised to pick either one volume or the other, rather than to attempt a synthesis. I suspect that instructors will find SGS's book more useful, given its organization, empirical examples, glossary, and separate chapter of proofs. Applied researchers may also favor SGS's book because of its connection with the TETRAD software. Pearl's book is more focused on recent developments—the chosen references, for example, are more weighted toward publications from the 1990s—and so may be more helpful to readers seeking to catch up on recent events. Those looking for insight into Pearl himself—his perspectives and contributions—need look no further. And anyone looking for a lively and literate introduction to this field should take a look at Pearl's epilogue, which he delivered in a 1996 UCLA Faculty Research Lectureship Program lecture and which is available online (http://singapore.cs.ucla.edu/LECTURE/lecture_sec1.htm).

With an increasingly narrow focus on deductive and confirmatory methods and model fitting, the SEM discipline remains open to criticisms leveled by Freedman (1985) and others that SEM users practice ex post model modification to achieve a semblance of compliance with their self-promulgated success criteria and that this "success" amounts to very little in practical terms, because it is focused on "modeling the data" instead of explaining real-world phenomena. These books suggest ways to address both problems. First, SEM users might simply admit that their theory is not strong enough to be paired with such a purely deductive technique. Instead of this Procrustean analysis, researchers might gain more by doing the best data collection they can and then letting the data set describe its own features. For that matter, SEM users might as well admit that their measures do not really stand up to the intense scrutiny of confirmatory factor analysis. Researchers might learn more by applying the newer, more sophisticated forms of exploratory factor analysis, such as Rozeboom's (1991) HYBALL package (available in compressed format at http://web.psych.ualberta.ca/~rozeboom/), which includes the HYBLOCK module for analysis of block-structured data (Rozeboom 1998).

Second, SEM users might do themselves a favor by admitting that they are indeed trying to model causal relations. If they are, and they count themselves successful in their research, then the next logical step is to put their causal findings to work in the real world. If you have new causal knowledge, you have the power to solve problems. Currently, it is hard to find the results of SEM analysis at work in the world of marketing practice. Is that an accident, or does it expose the sterility of mainstream SEM?

REFERENCES

Bollen, Kenneth A. and Kwok-fai Ting (1993), "Confirmatory Tetrad Analysis," in *Sociological Methodology*, Peter V. Marsden, ed. Washington, DC: American Sociological Association, 147–75.

Frawley, William J., Gregory Piatetsky-Shapiro, and Christopher J. Matheus (1991), "Knowledge Discovery in Databases: An Overview," in *Knowledge Discovery in Databases*, Gregory Piatetsky-Shapiro and William J. Frawley, eds. Menlo Park, CA: AAAI Press, 1–27.

Freedman, David A. (1985), "Statistics and the Scientific Method," in *Cohort Analysis in Social Research: Beyond the Identification Problem*, William M. Mason and Stephen E. Fienberg, eds. New York: Springer-Verlag, 343–66.

Groth, Robert (1998), *Data Mining: A Hands-On Approach for Business Professionals*. Upper Saddle River, NJ: Prentice Hall PTR.

Kenny, David A. (1979), *Correlation and Causality*. New York: John Wiley & Sons.

Lee, Soonmook and Scott Hershberger (1990), "A Simple Rule for Generating Equivalent Models in Covariance Structure Modeling," *Multivariate Behavioral Research*, 25 (July), 313–34.

Ling, Robert F. (1982), "Review of 'Correlation and Causation [sic],'" *Journal of the American Statistical Association*, 77 (June), 489–91.

Rigdon, Edward E. and Lynd D. Bacon (1997), "Data Warehousing and Data Mining: Opportunities, Challenges, and Implications for Marketing Management and Research," working paper, Department of Marketing, Georgia State University.

Rozeboom, William W. (1991), "HYBALL: A Method for Subspace-Constrained Oblique Factor Rotation," *Multivariate Behavioral Research*, 26 (April), 163–77.

——— (1998), "HYBLOCK: A Routine for Exploratory Factoring of Block-Structured Data," working paper, Department of Psychology, University of Alberta.

Stelzl, Ingeborg (1986), "Changing the Causal Hypothesis Without Changing the Fit: Some Rules for Generating Equivalent Models," *Multivariate Behavioral Research*, 21 (July), 309–31.

Ting, Kwok-fai (1995), "Confirmatory Tetrad Analysis in SAS," *Structural Equation Modeling*, 2 (April), 163–71.

EDWARD E. RIGDON
Georgia State University