Duty to Elaborate One's Causal Theory: Toward a New Norm for Empirical Reporting in Strategy

Gwendolyn Lee
University of Florida

Michael D. Ryall
Florida Atlantic University

June 6, 2024

Abstract

Since its inception, Strategy has been and, into the foreseeable future, will continue to be a field dominated by empirical studies. Even its theoretical development relies primarily upon inductive versus deductive methods. Interestingly, the 2024 Strategy Summit comes at the same moment we are seeing a crisis in replicability across the social sciences. It is therefore crucial, as we cast our gaze toward the future of the field and map out our overarching research priorities, that we give careful thought to extending and refining our norms for empirical methods and reporting. This paper highlights the major advances in causal identification methods that began in Computer Science roughly three decades ago. Presently going under the "Structural Causal Modeling" (SCM) moniker, these methods provide a rich foundation for the design, evaluation, and interpretation of empirical work aiming to estimate the causal effects of certain variables of interest ("treatments") on others ("outcomes"). This essentially covers all studies in Strategy. Very recently, a trickle of papers calling attention to these methods have seen publication in leading Management journals. The purpose of this short discussion is to highlight some of the subtleties of empirical analysis when the data is generated by complex causal systems. We close with a call to action to promote the use of SCM methods as a standard part of the Strategy research toolkit.

Status quo: elementary causal thinking

Casual empiricism leads to the observation – in our judgment, without controversy – that the lion's share of quantitative analysis in the field of Strategy (and Management more generally) takes a standard, ordinary least squares (OLS) approach to studying phenomena of interest. By this, we mean empirical designs that have in mind estimating the causal effect of one or more treatment variables upon an outcome variable using standard OLS regression techniques.

To take an arbitrary, well-cited (2,573, Google Scholar) example, Zott and Amit (2007) propose the following hypothesis: "The more novelty-centered an entrepreneurial firm's business model design, the higher the firm's performance." Like most hypotheses in Strategy, this is a causal conjecture. We would describe this conjecture as "elementary" in the sense that the causal theory presented is simply, $n \to \pi$, where n is some measure of novelty and π is some measure of firm performance. In this case, imagine a novelty variable in indicator form ($n = 1$ for novel business model and zero otherwise; in the actual study n is continuous). Then, in an ideal world, the causal effect of n on π could be determined using a randomized experiment: find a large population of firms with similar features operating under similar circumstances; randomly assign half to adopt a novel business model and half a conventional business model; then, measure the differences in performance between the two groups.

As readers will be quick to point out, random assignment of business models to real firms is rarely a feasible option. However, *if* the differences in features and settings between firms that adopt novel business models and those that do not are, on average, zero, then the causal effect of n on π can be accurately estimated by running a regression according to the model in familiar OLS form:

$$\pi_i = \alpha + \beta_n n_i + e_i. \tag{1}$$

Under these conditions (no average feature differences across novel versus conventional business models, large sample, indicator n), β_n is an accurate estimate of the effect of n on

π.

Readers will again be quick to point out that, true though this may be, the assumption that features across the two business model adoption categories are almost certainly *not* the same on average. As Zott and Amit (2007) point out, for example, the level of competitive intensity in a firm's industry will tend to reduce performance – regardless of the business model adopted. If differences in competitive intensity averaged out between the two groups, this would not be a problem. But suppose this is not the case, i.e., suppose there are differences in competitive intensity between the adopters of novel business models versus those who adopt conventional models. For example, firms in competitive industries may be more likely to adopt novel business models precisely to increase their added value and, as a result, increase profit.

This is an example of the well-known issue of selection bias. As described, competitive intensity is a common cause that influences both the treatment and outcome variables. Therefore, the analysis calls for the inclusion of a *control variable* that measures competitive intensity. Keeping the example simple, the analyst could introduce another indicator variable for competition, c, where $c = 1$ when competition is strong and zero when it is weak. Doing so breaks the sample into two groups according to whether a firm is in a strongly vs weakly competitive industry. Running the expanded regression

$$\pi_i = \alpha + \beta_n n_i + \beta_c c_i + e_i, \tag{2}$$

results in a comparison of the within-the-same-competitive-categories treatment effect (n on π), with β_n representing the observation-weighted average effect across the two categories (strong vs weak competition). If the differences of all other features of the firms in the novel versus conventional business model are, on average, zero, then β_n will be an unbiased (accurate) estimate of the expected treatment effect for the population as a whole. Moreover, conditioning on c_i provides the expected effects within each competitive category. By not controlling for industry competitiveness, OLS Model (1) results in a biased estimate of β_n. This is known as *omitted variable bias* (OVB).

We close this section with several concluding points. First, the norm for our discipline is to state an elementary causal hypothesis verbally and then test it empirically, typically (though not always) using a standard OLS design. Second, whether implicitly or explicitly stated, the goal is to create apples-to-apples (or, ceteris paribus) comparisons of direct treatment effects by controlling for those features which are correlated with meaningful differences across the treatment subpopulations. In other words, the idea is to mimic, as closely as possible, a randomized experiment by placing firms (or whatever agents of study) into categories such that the only within-category differences that remain are random, mean zero. This leads to a dominant emphasis in Strategy studies on including control variables in order to avoid OVB. Today's norm is to think of control variables as refinements to the estimation process, motivated by the idea that greater refinement leads to more accurate estimation by reducing the possibility of selection bias.

The new frontier: SCMs

Important findings in the young but burgeoning field of structural causal modeling (SCM), which has been developing for roughly the past three decades in computer science, have begun to make inroads into the social sciences. The literature on SCMs provides a rigorous and transparent foundation for the design and interpretation of empirical studies. SCMs, which can be thought of as nonparametric versions of structural equation models (Pearl, 2009, p. 30), have generated a rich body of theory with respect to conducting empirical analyses in the context of data generated by causal systems. For interested readers, additional texts in SCM methods include Pearl (2014, 2009, 2020), Spirtes et al. (2001), Ryall and Bramson (2013), and Morgan and Winship (2007). Recent papers that introduce these methods into Management include Lee and Bettis (2023, on causal mediation) and Frake et al. (2024, on collider bias).

The central tool at the heart of the SCM approach is the directed acyclic graph (DAG). A DAG summarizes the causal relationships between variables of interest in the system under study. The nodes of the graph represent the variables in the system – such as the type of business model adopted by a firm, whether the firm's industry is competitive or not, and so on. Directed arcs between the nodes indicate direct causal influences. In these models, "influence" is taken to mean that the actual values taken by a variable's parents (e.g., the business model is "innovative") determines the probability distribution on its own values (e.g., the probability that the firm's actual performance is "high"). DAGs provide a compact, intuitive and very flexible tool for elaborating the causal structure of interest.

In our simple example adapted from Zott and Amit (2007), the DAG representing the data-generating system would be:

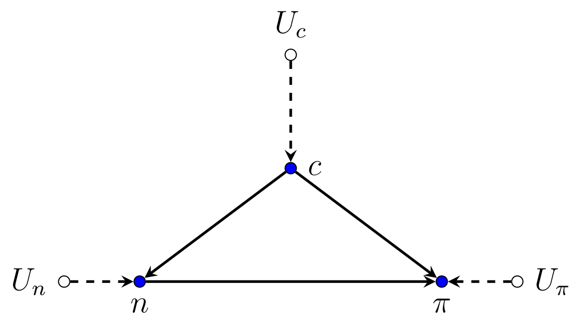

Figure 1: A simple DAG adapted from the business model study

This diagram highlights, in a nice qualitative form, the causal relations between the variables. It also highlights the need to control for industry competitiveness: competitiveness is a confounder – it simultaneously influences the choice of business model and firm performance. This effect must be controlled for to obtain an unbiased estimate of the direct influence of business model choice on firm performance. The remaining nodes represent unobserved factors. These are illustrated with empty nodes and dashed lines. As depicted, this illustrates a setting in which the only unobserved factors are those that cause independent random variation in the observed variables.

A DAG becomes a SCM when the stochastic relationships implied by the directed arcs are quantified. Using the language of randomized experiments, DAGs classify variables into six types: pre-treatment, treatment, confounder, collider, mediator, and outcome. In Figure 1, the U_is are unobserved pretreatment variables, n is the treatment, c is a confounder, and π is the outcome.

Of course, causal systems in the real world are much more elaborate than the simple version shown in Figure 1.[1] The complexity of the real world – especially in business strategy contexts – means that rarely, if ever, is the causal story as simple as, "A direct influence relation between treatment and outcome exists, potentially confounded by features that can be controlled for by including them as independent variables." This complexity leads to a number of subtle issues when designing empirical studies aimed at estimating causal effects. These issues present both dangers and opportunities for the introduction and correction, respectively, of biases in the estimation of causal effects (see the excellent discussion in Cinelli et al., 2022). For example, as we show below, bias can also be introduced by the inclusion of variables meeting certain structural conditions in a DAG.

Thus, a single-minded focus on OVB may well lead to the introduction of "bad controls," resulting in bias that could have been avoided had the variables simply been excluded. The notion that increasing the number of controls induces an increasingly refined estimate of the causal effect of interest is not correct. Moreover, while some variables can be added with with a neutral effect on estimation, good empirical design dictates they be skipped to save the cost of collecting them and to avoid the reduced degrees of freedom associated with adding more independent variables to the analysis.

At the same time, a correct understanding of the causal structure that underlies the data can also lead to insightful approaches to eliminating OVB. It is not always necessary to resort to advanced identification methods, such as instrumental variables, diff-in-diff, or discontinuity analyses to resolve problematic situations. Moreover, working through one's causal theory, à la the SCM approach provides insight into how these more advanced methods work and when they are actually called for.

The OVB problem is well understood. Therefore, let us turn toward some illustrative examples of the kinds of novel insights that arise when SCM methodologies are adopted. In these examples, we simply assert certain results without elaborating the technical details behind them (which are fully discussed in the works cited above). Our goal here is to advocate for the adoption of these methods by pointing out some interesting insights that, without them, would be missed in a standard OLS approach.

[1] Including the actual setting described by Zott and Amit (2007).

Subtleties highlighted by SCMs

Unobserved but neutral(ized) confounder Consider the previous example but with the following twist: the effect of business model type on performance is mediated by the spirit of the firm's employees. The causal story is that once senior leadership decides upon a business model (novel vs conventional), they then select employees whose mix (m) of personality profiles fits that model best. These employees then do the work and, in so doing, directly influence the firm's performance. Furthermore, assume that the industry competitiveness is not easily measurable. The following diagram summarizes this causal story (with noise variables removed to reduce clutter):

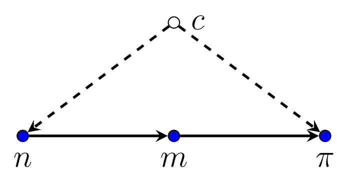

Figure 2: Using m to neutralize the OVB due to c

Here, we see that c is an unobserved common cause of n and π. Immediately, the OVB alarm bells begin to ring. Perhaps surprisingly, the OVB problem induced by confounder c can be neutralized without actually measuring c. In this situation, a two-stage OLS regression will create an unbiased estimate of the effect of n on π: Stage I: regress m on n; Stage II: using the Stage I estimates, regress π on m.

Additionally, in both of the following situations, controlling for m neutralizes the OVB due to c as well:

(a) m controls for c OVB

(b) m controls for c OVB

Figure 3: Two situations where controlling for m neutralizes the OVB due to c

What are the causal stories here? In Figure 3a, we might imagine that competitive intensity influences the available mix of job candidates in the general population and, hence, in the company. Then, the mix of personality profiles in the company may affect the choice of business model. On the other hand, Figure 3b, could represent a similar situation only under which the mix of personality profiles in the company directly affect performance rather than choice of business model. The point is that, even without the ability to measure industry competitiveness directly and even without advanced identification methods, including a measurable variable m that is consistent with the causal story can solve the OVB problem.

Unfortunately, deciding how and when to use variables such as m to solve the OVB problem is, itself, a challenge. For instance, the variable m in Figure 2 is a mediator, which is a variable that is influenced by the treatment and subsequently influences the outcome. Including m as a control variable in a regression analysis introduces *overcontrol bias*, because controlling for a mediator blocks the very effect we want to estimate (see Lee and Bettis, 2023, for an application of SCMs for causal mediation in strategy research). Thus, m should not be used as a control variable in Figure 2. The key to solving the OVB problem, as mentioned earlier, is running a two-stage regression (an insight derived from the DAG).

"Controls" that introduce bias We put scare quotes around *controls* to indicate variables that, when included in a regression with the aim of eliminating OVB actually introduce some other bias instead (and, hence, should not be included). Generally, these are referred to as "bad controls." Chief among these is the *collider*. Colliders are variables that are caused by or share a common cause with both the outcome and the treatment. Including such variables leads to *collider bias*.

Let us return to Zott and Amit (2007) who hypothesize (p. 188) that firm size is, "a proxy for the focal firm's bargaining power, relative to rival firms and other business model stakeholders. All other things being equal, the larger the focal firm, the greater its potential for value creation as well as its bargaining power, and, hence, the better its performance." Firm size was measured as the logarithm of the number of employees. The potential problem is that the number of employees could be an outcome of business model novelty and performance. This would be the case if firms adjust hiring policies to fit business model types and if better performance enables firms to hire employees by exercising bargaining power vis-à-vis stakeholders. Figure 4 illustrates this setting, where e is the employee variable (intuitively, we see why e is referred to as a "collider").

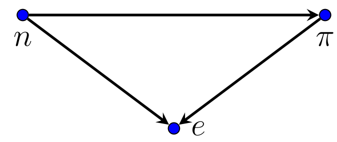

Figure 4: Collider bias

Another example of a collider situation would be one that includes winning the Boston Consulting Group's award of most innovative companies (https://www.bcg.com/publications/most-innovative-companies-historical-rankings), where business model innovation is one of the award criteria. Both novelty and performance likely affect a firm's probability of winning the innovation award.

The problem is that conditioning, controlling, selecting, or stratifying on a collider variable distorts the relationship between its parent variables (Cinelli et al., 2022; Elwert and Winship, 2014; Griffith et al., 2020; Schneider, 2020). For instance, when examining the effect of corporate social responsibility (CSR) on a firm's financial performance, collider bias occurs when a firm's inclusion in Forbes' Most Admired Companies list is specified as a control variable. This is a bad control because both CSR and financial performance likely affect a firm's probability of being included on the list. Controlling for the probability of inclusion on the list, which is a collider caused by both the outcome and the treatment, creates a spurious correlation between CSR and financial performance.

A literature review by Frake et al. (2024) shows that papers published in leading strategy and management journals discuss collider bias at roughly half the rate of top economics papers, and four times less than those in leading sociology journals. Collider bias manifests not only as a problematic control in standard OLS studies, but also intersects with many research design problems, including matching, fixed-effect analysis, selection on the dependent variable, selection into archival dataset, sample exclusion criteria, attrition, and non-response.

Summing up, the "more controls = greater accuracy" mindset is one that continues to be dominant in our field. Yet, informed by the now extensive body of work in the SCM literature, a small but growing number of papers have begun to highlight the problems with this mindset in the context of research in business strategy. Properly designed studies can find ways around variables that are impossible to measure (e.g., Figure 2). At the same time, there is such a thing as a bad control. Using colliders as controls is a problem as is the unsophisticated use of mediators. Even studies that incorporate sophisticated identification techniques must be aware of these issues.

Conclusion: a call to action

Apparently, there is no getting around beginning one's empirical study with a causal theory – at best summarized in the form of a DAG. Without such a theory, the empirical analyst cannot make informed decisions with respect to his or her research design. The researcher must posit a theory (or theories) up front and proceed to conduct the analysis accordingly. Without seeing such a theory, it is impossible to know what, exactly, the authors think their study is accomplishing.

SCMs are particularly helpful in observational studies, when actual experiments cannot be conducted. Developing a causal theory in DAG form as part of an empirical analysis accomplishes several objectives. First, it communicates the theory in a single, easy to understand diagram. Second, in doing so, it permits readers to decide for themselves the extent to which that theory is compelling (or even believable). Third, and most importantly, it provides justification for the design of the empirical analysis – including which variables were included, which were excluded, and whether multi-stage regressions were conducted as remedial steps taken to deal with OVB problems. In addition, the DAG implies a minimal set of variables that must be controlled for to obtain an unbiased estimate. As Hünermund and Louw (2023) demonstrate, it is not necessary to include all causal influence factors of the outcome variable in a regression. In many cases it may be easier to control the treatment assignment mechanism than attempt to include a long list of variables that affect the outcome.

Knowing the effect of including a particular variable as a control is helpful when comparing one's work against a prior study or must respond to a request by a reviewer once the paper is under submission. Furthermore, SCM methods help researchers assess the robustness of their findings in the presence of unobserved confounders (see Lee and Bettis, 2023, for a discussion about sensitivity tests). Cinelli and Hazlett (2019) discuss how to conduct sensitivity analyses for quantifying the necessary strength of a hypothetical unobserved confounder that could nullify the findings. If there is no suitable data in which all necessary and sufficient control variables are observed, SCMs can be used to improve research design by screening instrumental variables, assessing collider bias in fixed-effects models, and evaluating collider bias in matching variables.

Finally, it is worth noting that the SCM community has developed several statistical tests to assess goodness of fit of the data to the causal theory (i.e., the DAG, which may include nodes for omitted variables). Even without employing sophisticated identification techniques, such as instrumental variables, discontinuity designs, diff-in-diff designs, and so on, the SCM approach provides fertile ground for the testing of causal theories and estimation of causal effects. Moreover, these techniques are intuitive and accessible by a broad set of scholars.

In closing, it must be pointed out that every empirical researcher is, by the very nature of their work, a causal theorist – the only question is whether that theory has been explicitly examined. Creating a norm in which causal theories are discussed and elaborated using the SCM standards now prevalent in computer science would impose a crucial level of self-reflection during the research stage. It would also lead to better studies with more robust findings. Indeed, we conjecture that it may even contribute to resolving the present crisis in the replicability of findings in social science.

References

Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1): 39–67, December 2019. ISSN 1467-9868. doi: 10.1111/rssb.12348.

Carlos Cinelli, Andrew Forney, and Judea Pearl. A crash course in good and bad controls. *Sociological Methods and Research*, page 004912412210995, May 2022. ISSN 1552-8294. doi: 10.1177/00491241221099552.

Felix Elwert and Christopher Winship. Endogenous selection bias: The problem of conditioning of a collider variable. *Annual Review of Sociology*, 40:31–53, 2014.

Justin Frake, Andreas Hagemann, and Jose Uribe. Collider bias in strategy and management research: An illustration using women ceo's effect on other women's career outcomes. *Strategic Management Journal*, March 2024. ISSN 1097-0266. doi: 10.1002/smj.3588.

Gareth J Griffith, Tim T Morris, Matthew J Tudball, Annie Herbert, Giulia Mancano, Lindsey Pike, Gemma C Sharp, Jonathan Sterne, Tom M Palmer, George Davey Smith, et al. Collider bias undermines our understanding of covid-19 disease risk and severity. *Nature communications*, 11(1):5749, 2020.

Paul Hünermund and Beyers Louw. On the nuisance of control variables in causal regression analysis. *Organizational Research Methods*, December 2023. ISSN 1552-7425. doi: 10.1177/10944281231219274.

Gwendolyn K. Lee and Rich Bettis. Structural causal modeling of managerial interventions: What if managers had not intervened by doing this? *Strategy Science*, 8(1):24–43, March 2023. ISSN 2333-2077. doi: 10.1287/stsc.2022.0169.

Stephen Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, Cambridge, 2007.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Elsevier, 2014.

Judea Pearl. *The book of why.* Basic Books, New York, first trade paperback edition edition, 2020. ISBN 9780465097609. Literaturverzeichnis: Seite 377-404.

Michael D. Ryall and Aaron L. Bramson. *Inference and Intervention: Causal Models for Business Analysis.* Routledge, 2013. ISBN 978-0-415-65760-0. URL `http://books.google.com/books?id=7Q6CAAAAQBAJ&pgis=1`.

Eric B. Schneider. Collider bias in economic history research. *Explorations in Economic History*, 78:101356, October 2020. ISSN 0014-4983. doi: 10.1016/j.eeh.2020.101356.

P Spirtes, C N Glymour, and R Scheines. *Causation, prediction, and search.* The MIT Press, 2001.

Christoph Zott and Raphael Amit. Business model design and the performance of entrepreneurial firms. *Organization Science*, 18(2):181–199, April 2007. ISSN 1526-5455. doi: 10.1287/orsc.1060.0232.