

Causal Inference in Accounting Research

IAN D. GOW,* DAVID F. LARCKER,[†] AND PETER C. REISS[†]

ABSTRACT

This paper examines the approaches accounting researchers adopt to draw causal inferences using observational (or nonexperimental) data. The vast majority of accounting research papers draw causal inferences notwithstanding the well-known difficulties in doing so. While some recent papers seek to use quasi-experimental methods to improve causal inferences, these methods also make strong assumptions that are not always fully appreciated. We believe that accounting research would benefit from more in-depth descriptive research, including a greater focus on the study of causal mechanisms (or causal pathways) and increased emphasis on the structural modeling of the phenomena of interest. We argue these changes offer a practical path forward for rigorous accounting research.

JEL codes: C18; C190; C51; M40; M41

Keywords: Causal inference; accounting research; quasi-experimental methods; structural modeling

1. Introduction

There is perhaps no more controversial practice in social and biomedical research than drawing inferences from observational data. Despite ...

*Harvard Business School; [†]Rock Center for Corporate Governance, Stanford Graduate School of Business.

Accepted by Philip Berger. We are grateful to our discussants, Christian Hansen and Miguel Minutti-Meza, and participants at the 2015 JAR Conference for helpful feedback. We also thank seminar participants at London Business School, Karthik Balakrishnan, Robert Kaplan, Christian Leuz, Alexander Ljungqvist, Eugene Soltes, Daniel Taylor, Robert Verrecchia, Charles Wang, and Anastasia Zakolyukina for comments.

problems, observational data are widely available in many scientific fields and are routinely used to draw inferences about the causal impact of interventions. The key issue, therefore, is not whether such studies should be done, but how they may be done well. (Berk [1999, p.95])

Most empirical research in accounting relies on observational (or non-experimental) data. This paper evaluates the different approaches accounting researchers adopt to draw causal inferences from observational data.¹ Our discussion draws on developments in fields such as statistics, econometrics, and epidemiology. The goal of this paper is to identify areas for improvement and suggest how empirical accounting research can improve inferences drawn from observational data.

The importance of causal inference in accounting research is clear from the research questions that accounting researchers seek to answer. Most long-standing questions in accounting research are causal: Does conservatism affect the terms of loan contracts? Do higher quality earnings reports lead to lower information asymmetry? Did International Financial Reporting Standards cause an increase in liquidity in the jurisdictions that adopted them? Do managerial incentives lead to managerial misstatements in financial reports? The accounting researchers focus on causal inference, which is consistent with the view that “the most interesting research in social science is about questions of cause and effect” (Angrist and Pischke [2008, p. 3]). Simply documenting descriptive correlations provides little basis for understanding what would happen should circumstances change, whereas using data to make inferences that support or refute broader theories could facilitate these kinds of predictions.

To provide insights into what is actually done in empirical accounting research, we examined all papers published in three leading accounting journals in 2014. While accounting researchers are aware of problems that can arise from the use of observational data to draw causal inferences, we found that most papers still seek to draw such inferences. Making causal inferences requires strong assumptions about the causal relations among variables. For example, estimating the causal effect of X on Y requires that the researcher has controlled for variables that could confound estimates of such effects. Section 2 provides an overview of causal inference using causal diagrams as a framework for thinking about the subtle issues involved. We believe that these diagrams are also very useful for communicating the cause-and-effect logic underlying regression analyses that use observational data. Nonetheless, difficulties identifying, measuring, and controlling for all possible confounding variables have led many to question causal inferences drawn from observational data.

Recently, some social scientists have held out hope that better research designs and statistical methods can increase the credibility of causal

¹ Thus, our focus is on what Bloomfield, Nelson, and Soltes [2016] call “archival studies.” Floyd and List [2016] discuss opportunities for researchers to use experiments in accounting research.

inferences. For example, Angrist and Pischke [2010] suggest that “empirical microeconomics has experienced a credibility revolution, with a consequent increase in policy relevance and scientific impact.” Angrist and Pischke [2010, p. 26] argue that such “improvement has come mostly from better research designs, either by virtue of outright experimentation or through the well-founded and careful implementation of quasi-experimental methods.” Our survey of research published in 2014 finds 5 studies claiming to study natural experiments (or “exogenous shocks”) and 10 studies using instrumental variables (IVs). Although these numbers suggest that quasi-experimental methods are infrequently used in accounting research, we believe their use will increase in the future.²

Section 3 evaluates the use of quasi-experimental methods in accounting research. Quasi-experimental methods produce credible estimates of causal effects only under very strong maintained assumptions about the model and data. For example, variations in treatments are rarely random, the list of controls rarely exhaustive, and instruments do not always satisfy the necessary inclusion and exclusion restrictions. We explain some of these concerns using causal diagrams. In general, it appears that the assumptions required to apply quasi-experimental methods are unlikely to be satisfied by observational data in most empirical accounting research settings.

Ultimately, we believe that accounting research needs to recognize the stringent assumptions that need to be maintained to apply statistical methods to derive estimates of causal effects for observational data. Statistical methods alone cannot solve the inference issues that arise in observational data. The second part of the paper (sections 4 and 5) identifies approaches that can provide a plausible framework for guiding future accounting research. Specifically:

- There should be an increased emphasis on the study of causal mechanisms, that is, the “pathways” through which claimed causal effects are propagated. We believe that evidence on the actions and beliefs of individuals and institutions can bolster causal claims based on associations, even absent compelling estimates of the causal effects. We also suggest that more careful modeling of phenomena, using structural modeling or causal diagrams, can help to identify plausible mechanisms that warrant further study.
- Causal diagrams are a useful tool for conveying the key elements of a structural model and can also act as a middle-level stand-in when structural modeling of a phenomenon is infeasible.³

² We use the term “quasi-experimental” methods to refer to those methods that have a plausible claim to “as if” random assignment to treatment conditions. The term “as if” is used by Dunning [2012] to acknowledge the fact that assignment is not random in such settings, but is claimed to be *as if* random assignment had occurred.

³ “Middle-level” here refers to the placement of causal diagrams between relatively informal verbal reasoning and the rigors of a structural model.

- There should be an increased use of structural modeling methods. Structural models provide a more complete characterization of the behavior and institutions that underlie a phenomenon of interest. We acknowledge that while structural models need not be a correct characterization, they have the advantage of making what is assumed explicit. This gives other researchers a rigorous way to assess the model and understand what would happen if features of the model change.
- There are many important questions in accounting that have not yet been addressed by formal models. In these settings, it is important to conduct sophisticated descriptive research aimed at understanding the phenomena of interest so as to develop clearer cause-and-effect models. In our view, many hypotheses that are tested with observational data are only loosely tied to the accounting institutions and business phenomena of interest. We hope that a larger number of richer descriptive studies will provide insights that the theorists can use to build models that empiricists can actually “take to data.”

2. *Causal Inference: An Overview*

2.1 CAUSAL INFERENCE IN ACCOUNTING RESEARCH

To get a sense for the importance of causal questions in accounting research, we examined all papers published in 2014 in the *Journal of Accounting Research*, *The Accounting Review*, and the *Journal of Accounting and Economics*. We counted 139 papers, of which 125 are original research papers. Another 14 papers survey or discuss other papers. We classify each of the 125 research papers into one of the following four categories: “Theoretical” (7), “Experimental” (12), “Field” (3), or “Archival Data” (103). For our next discussion, we collect the field and archival data papers into a single “Observational” category.

For each nontheoretical paper, we determine whether the primary or secondary research questions are “causal.” Often the title reveals a causal question, with words such as “effect of . . .” or “impact of . . .” (e.g., Clor-Proell and Maines [2014], Cohen et al. [2014]). In other cases, the abstracts reveal that authors have causal inferences as a goal. For example, de Franco et al. [2014] inquires “how the tone of sell-side debt analysts’ discussions about debt-equity conflict events *affects* the informativeness of debt analysts’ reports in debt markets.”

We recognize that some authors might disagree with our characterizations. For example, a researcher might argue that a paper that claimed that “theory predicts X is associated Y and, consistent with that theory, we show X is associated with Y ” is merely a descriptive paper that does not make causal inferences. However, theories are invariably causal in that they posit how exogenous variation in certain variables leads to changes in other variables. Further, by stating that “consistent with . . . theory, X is associated with Y ,” the clear purpose is to argue that the evidence tilts the scale, however

slightly, in the direction of believing the theory is a valid description of the real world: In other words, a causal inference is drawn.⁴

Of the 106 original papers using observational data, we coded 91 as seeking to draw causal inferences.⁵ Of the remaining empirical papers, we coded seven papers as having a goal of “description” (including two of the three field papers). For example, Soltes [2014] uses data collected from one firm to describe analysts’ private interactions with management. Understanding how these interactions take place is key to understanding whether and how they transmit information to the market. We coded five papers as having a goal of “prediction.” For example, Czerney, Schmidt, and Thompson [2014] examine whether the inclusion of “explanatory language” in unqualified audit reports can be used to predict the detection of financial misstatements in the future. We coded three papers as having a goal of “measurement.” For example, Cready, Kumas, and Subasi [2014] examine whether inferences about traders based on trade size are reliable and suggest improvements for the measurement of variables used by accounting researchers.

In summary, we find that most original research papers use observational data and about 90% of these papers seek to draw causal inferences. The most common estimation methods used in these studies include ordinary least-squares (OLS) regression, difference-in-differences (DD) estimates, and propensity-score matching (PSM). While it is widely understood that OLS regressions that use observational data produce unbiased estimates of causal effects only under very strong assumptions, the credibility of these assumptions is rarely explicitly addressed.⁶

2.2. CAUSAL INFERENCE: A BRIEF OVERVIEW OF RECENT DEVELOPMENTS

In recent decades, the definition and logic of causality has been revisited by researchers in fields as diverse as epidemiology, sociology, statistics, and computer science. Rubin [1947, p. 7] and Holland [1986] formalized ideas from the potential-outcome framework of Neyman [1923], leading to the so-called “Rubin causal model.” Other fields have used path analysis, as initially studied by geneticist Sewell Wright (Wright [1921]), as an organizing framework. In economics and econometrics, early proponents of structural models were quite clear about how causal statements must be

⁴ Papers that seek to estimate a causal effect of X on Y are a subset of papers we classify as causal. A paper that argues that Z is a common cause of X and Y and claims to find evidence of this is still making causal inferences (i.e., Z causes X and Z causes Y). However, we do not find this kind of reasoning to be common in our survey.

⁵ While we exclude research papers using experimental methods, all these papers also seek to draw causal inferences.

⁶ There are settings where DD and fixed-effect estimators may deliver causal estimates. For example, if assignment to treatment is random, then it is possible for a DD estimate using pre- and posttreatment data to yield unbiased estimates of causal effects. But, in this case, it is the detailed understanding of the research setting, not the method per se, that makes these estimates credible.

tied to theoretical economic models. As discussed by Heckman and Pinto [2015], Haavelmo [1943, p. 4] promoted structural models “based on a system of structural equations that define causal relationships among a set of variables.” Goldberger [1972, p. 979] promoted a similar notion: “By structural equation models, I refer to stochastic models in which each equation represents a causal link, rather than a mere empirical association . . . Generally speaking the structural parameters do not coincide with coefficients of regressions among observable variables, but the model does impose constraints on those regression coefficients.” Goldberger [1972] focuses on linking such approaches to the path analysis of Wright.

An important point worth emphasizing is that the model-based causal reasoning is distinct from statistical reasoning. Suppose we observe data on x and y and make the strong assumption that we know causality is one-way. How do we distinguish between whether X causes Y or Y causes X ? Statistics can help us determine whether X and Y are correlated, but correlations do not establish causality. Only with assumptions about causal relations between X , Y , and other variables (i.e., a theory) can we infer causality. While theories may be informed by evidence (e.g., prior research may suggest a given theory is more or less plausible), they also encode our understanding of causal mechanisms (e.g., barometers do not cause rain).

Computer and decision scientists, as well as researchers in other disciplines, have recently sought to develop an analytical framework for thinking about causal models and their connection to probability statements (e.g., Pearl [2009a]). Pearl’s framework, which he calls the structural causal model, uses causal diagrams to describe causal relationships. These diagrams encode causal assumptions and visually communicate how a causal inference is being drawn from a given research design. Given a *correctly specified* causal diagram, these criteria can be used to verify conditioning strategies, IV designs, and mechanism-based causal inferences.⁷

We use figure 1 to illustrate the basic ideas of causal diagrams and how they can be used to facilitate causal inference. Figure 1 depicts three variants of a simple causal graph. Each graph depicts potential relationships among the three (observable) variables. In each case, we are interested in understanding how the presence of a variable Z impacts the estimation of the causal effect of X on Y . The only difference between the three graphs is the direction of the arrows linking either X and Z , or Y and Z . The boxes (or “nodes”) represent random variables and the arrows (or “edges”) connecting boxes represent hypothesized causal relations, with each arrow pointing from a cause to a variable assumed to be affected by it.

Pearl [2009b] shows that, if we are interested in assessing the causal effect of X on Y , we may be able to do so by conditioning on a set of variables,

⁷ While Pearl [2009a, p. 248] defines an instrument in terms of causal diagrams, additional assumptions (e.g., linearity) are often needed to estimate causal effects using an instrument (Angrist, Imbens, and Rubin [1996]).

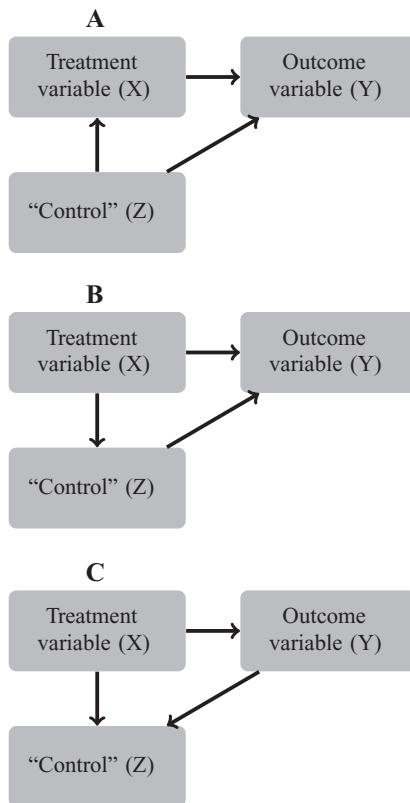


FIG. 1.—Three basic causal diagrams. (A) Z is a confounder, (B) Z is mediator, and (C) Z is a collider.

Z , that satisfies certain criteria. These criteria imply that very different conditioning strategies are needed for each of the causal diagrams (see the appendix for a more formal discussion).

While conditioning on variables is much like the standard notion of “controlling for” such variables in a regression, there are critical differences. First, conditioning means estimating effects for each distinct level of the set of variables in Z . This nonparametric concept of conditioning on Z is more demanding than simply including Z as another regressor in a linear regression model.⁸ Second, the inclusion of a variable in Z may not be an appropriate conditioning strategy. Indeed, it can be that the inclusion of Z results in biased estimates of causal effects.

Each of the three graphs in figure 1 provides an alternative view of the causal effect of X on Y . Figure 1(A) is straightforward. It shows that we need

⁸ Including variables in a linear regression framework “controls for” only under strict assumptions, such as linearity in the relations between X , Y , and Z .

to condition on Z in order to estimate the causal effect of X on Y . Note that the notion of “condition on” again is more general than just including Z in a parametric (linear) model.⁹ The need to condition on Z arises because Z is what is known as a *confounder*.

Figure 1(B) is a bit different. Here, Z is a *mediator* of the effect of X on Y . No conditioning is required in this setting to estimate the total effect of X on Y . If we condition on X and Z , then we obtain a different estimate, one that includes the indirect effect of X on Z .

Finally, in figure 1(C), we have Z acting as what is referred to as a “collider” variable (Glymour and Greenland [2008], Pearl [2009a]).¹⁰ Again, not only do we not need to condition on Z , but that we *should* not condition on Z to get an estimate of the total effect of X on Y . While in epidemiology, the issue of “collider bias . . . can be just as severe as confounding” (Glymour and Greenland [2008, p. 186]), collider bias appears to receive less attention in accounting research than confounding. Many intuitive examples of collider bias involve selection or stratification. Admission to a college could be a function of combined test scores (T) and interview performance (I) exceeding a threshold, that is, $T + I \geq C$. Even if T and I are unrelated unconditionally, a regression of T on I conditioned on admission to college is likely to show a negative relation between these two variables.

2.3. CAUSAL DIAGRAMS: APPLICATIONS IN ACCOUNTING

A typical paper in accounting research will include many variables to “control for” the potential confounding of causal effects. While many of these variables should be considered confounders, less attention is given to explaining why it is reasonable to assume that they are not mediators or colliders. Such a discussion is important because the inclusion of “controls” that are mediators or colliders will generally lead to bias.

One paper that does discuss this distinction is Larcker, Richardson, and Tuna [2007], who use a multiple regression (or logistic) model of the form¹¹

$$Y = \alpha + \sum_{r=1}^R \gamma_r Z_r + \sum_{s=1}^S \beta_s X_s + \epsilon. \quad (1)$$

Larcker, Richardson, and Tuna [2007, p.983] suggest that:

One important feature in the structure of Equation 1 is that the governance factors [X] are assumed to have no impact on the controls (and thus no indirect impact on the dependent variable). As a result, this structure may result in conservative estimates for the impact of governance on

⁹ Inclusion of Z blocks the “back-door” path from Y to X via Z .

¹⁰ The two arrows from X and Y “collide” in Z .

¹¹ We alter the mathematical notation of Larcker, Richardson, and Tuna [2007] to conform to the notation we use here.

the dependent variable. Another approach is to only include governance factors as independent variables, or:

$$Y = \alpha + \sum_{s=1}^S \beta_s X_s + \epsilon \quad (2)$$

The structure in Equation 2 would be appropriate if governance impacts the control variables and both the governance and control variables impact the dependent variable (i.e., the estimated regression coefficients for the governance variables will capture the total effect or the sum of the direct effect and the indirect effect through the controls).

But, there are some subtle issues here. If some elements of Z_r are mediators and others are confounders, then both equations will be subject to bias. Equation (2) will be biased due to omission of confounders, while equation (1) will be biased due to inclusion of mediating variables. Additionally, the claim that the estimates are “conservative” is only correct if the indirect effect via mediators is of the same sign as the direct (i.e., unmediated) effect. If this is not the case, then the relation between the magnitude (and even the sign) of the direct effect and the indirect effect is unclear.

Additionally, this discussion does not allow for the possibility of colliders. For example, governance plausibly affects leverage choices, while performance is also likely to affect leverage. If so, “controlling for” leverage might induce associations between governance and performance even absent in a true relation between these variables.¹² While the with-and-without-controls approach used by Larcker, Richardson, and Tuna [2007] has intuitive appeal, a more robust approach requires careful thinking about the plausible causal relations between the treatment variables, outcomes of interest, and candidate control variables.

3. Quasi-Experimental Methods in Accounting Research

While most studies in accounting use regression or matching methods to condition out confounding variables, a number of studies use quasi-experimental methods that rely on “as if” random assignment to identify causal effects (Dunning [2012]). Of the 91 papers in our 2014 survey seeking to draw a causal inference from observational data, we classify 14 as relying on quasi-experimental methods. Despite the low count, we believe that papers using these methods are considered stronger research contributions, and there seems an increasing trend toward the use of quasi-experimental methods. Additionally, a number of papers use methods such as DD or fixed-effect estimators, which are widely believed to approximate quasi-experimental methods. This section discusses and evaluates the usefulness of these methods for accounting research.

¹² Note that Larcker, Richardson, and Tuna [2007] do not in fact use leverage as a control when performance is a dependent variable.

3.1. NATURAL EXPERIMENTS

Natural experiments occur when observations are assigned by nature (or some other force outside the control of the researcher) to treatment and control groups in a way that is random or “as if” random (Dunning [2012]). Truly random assignment to treatment and control provides a sound basis for causal inference, enhancing the appeal of natural experiments for social science research. However, Dunning [p. 3, emphasis added] argues that this appeal “may provoke *conceptual stretching*, in which an attractive label is applied to research designs that only implausibly meet the definitional features of the method.”

Our survey of accounting research in 2014 identified five papers that exploited either a “natural experiment” or an “exogenous shock” to identify causal effects.¹³ An examination of these papers reveals how difficult it is to find a plausible natural experiment in observational data.

An important difficulty is that most “exogenous shocks” (e.g., Securities and Exchange Commission (SEC) regulatory changes or court rulings) do not randomly assign units to treatment and control groups and thus do not qualify as natural experiments. For example, an early version of Dodd-Frank contained a provision that would force companies to remove a staggered board structure.¹⁴ It is tempting to use this event to assess the valuation consequences of having a staggered board by looking at excess returns for firms with and without a staggered board around the announcement of this Dodd-Frank provision. Although potentially interesting, the Dodd-Frank “natural experiment” does not randomly assign firms to treatment and control groups. Instead, firms made an endogenous choice about whether to have a staggered board, and the regulation is potentially forcing firms to change that choice. But, firms might have a variety of margins through which they can respond to such a requirement, some of which may have valuation consequences of their own.¹⁵ Absent an account of these margins, an event study that includes a staggered board treatment variable does not isolate the (pure) effect of staggered boards on valuations.

Another important concern is that there could be a reason to believe that the natural experiment affected treatment assignments, and this impact is correlated with unobserved factors that might impact the outcome of interest. In general, even claims of random assignment to treatment do not suffice to deliver unbiased estimates of causal effects. An example of a drug trial can help underscore these points. Suppose we wish to understand whether a drug lowers blood pressure. Imagine patients in the trial are drawn from two hospitals. One hospital is randomly selected as the hospital

¹³ These are Lo [2014], Aier, Chen, and Pevzner [2014], Kirk and Vincent [2014], Houston et al. [2014], and Hail, Tahoun, and Wang [2014].

¹⁴ See Larcker, Ormazabal, and Taylor [2011].

¹⁵ For instance, if forced to remove a staggered board, some firms may put in another anti-takeover provision.

in which the drug will be administered. The other hospital's patients serve as controls. Suppose, in addition, that we know the patient populations in both hospitals are similar.

Most researchers would argue that we have all the ingredients for a successful treatment effect study. In particular, assignment to treatment is random. Now imagine that patients actually have to take the drug for it to have an effect. In this case, if there are unobserved reasons why some assigned to treatment opt out, modify the dosage, or stop taking medications for which there might be interactions, then being assigned to treatment is not the same as treatment. To take an extreme example, suppose the drug has a slight negative effect on blood pressure, everyone in fact takes the drug, but doctors in the hospital where patients are treated tell patients to stop taking their regular blood pressure medication. In this case, if regular blood pressure medications lower blood pressure more than the new drug, we might conclude that the new drug actually *raises* blood pressure! In sum, even showing that a treatment is randomly assigned does not guarantee that a regression will uncover the causal effect of interest.

Finally, it is important to carefully consider the choice of explanatory variables in studies that rely on natural experiments. In particular, researchers sometimes inadvertently use covariates that are affected by the treatment in their analysis. As noted by Imbens and Rubin [2015, p. 116], including such posttreatment variables as covariates can undermine the validity of causal inferences.¹⁶

Extending our survey beyond research published in 2014, we find papers with very plausible natural experiments. One such paper is Michels [2015], who exploits the difference in disclosure requirements for significant events that occur before financial statements are issued. Because the timing of these events (e.g., fires and natural disasters) relative to balance sheet dates is plausibly random, the assignment to the disclosure and recognition conditions is plausibly random. Nevertheless, even in this relatively straightforward setting, Michels [2015] recognizes the possibility of different materiality criteria for disclosed and recognized events, which could affect the relation between underlying events and observed disclosures. Michels' paper takes care to address this concern.¹⁷

Another plausible natural experiment is examined in Li and Zhang [2015, p. 80], who study a regulatory experiment in which the SEC "mandated temporary suspension of short-sale price tests for a set of randomly selected pilot stocks." Li and Zhang [2015, p. 79] conjecture "that managers respond to a positive exogenous shock to short selling pressure . . . by reducing the precision of bad news forecasts." But if the treatment affects

¹⁶ See the discussion of mediators above.

¹⁷ The setting of Michels [2015] plausibly involves a natural experiment. The endogenous nature of the disclosure and reporting responses by firms to these events, which is what is observable to the researcher, makes drawing causal inferences about the effect of recognition versus disclosure problematic.

the properties of these forecasts, and Li and Zhang [2015, p. 79] sought to condition on such properties, they would risk undermining the “natural experiment” aspect of their setting.

When true natural experiments can be found, they are an excellent setting for drawing causal inferences from observational data. Unfortunately, credible natural experiments are very rare. Certainly researchers should exploit these natural experiments when they occur (e.g., Li and Zhang [2015], Michels [2015]), but care also is needed when doing so.

3.2 INSTRUMENTAL VARIABLES

Angrist and Pischke [2008, p. 114] describe IVs as “the most powerful weapon in the arsenal” of econometric tools. Accounting researchers have long used IVs to address concerns about endogeneity (Larcker and Rusticus [2010], Lennox, Francis, and Wang [2012]) and continue to do so. Our survey of research published in 2014 identifies 10 papers using IVs.¹⁸ Much has been written on the challenges for researchers using IVs as the basis for causal inference (e.g., Roberts and Whited [2013]), and it is useful to use this background to evaluate the application of this approach in accounting research.

3.2.1. Evaluating IVs Requires Careful Theoretical Causal (Not Statistical) Reasoning. With respect to accounting research, Larcker and Rusticus [2010] lament that “some researchers consider the choice of IVs to be a purely statistical exercise with little real economic foundation” and call for “accounting researchers . . . to be much more rigorous in selecting and justifying their instrumental variables.” Angrist and Pischke [2008, p. 117] argue that “good instruments come from a combination of institutional knowledge and ideas about the process determining the variable of interest.” One study that illustrates this is Angrist [1990]. In that setting, the draft lottery is well understood as random and the process of mapping from the lottery to draft eligibility is well understood. Furthermore, there are good reasons to believe that the draft lottery does not affect anything else directly except for draft eligibility.¹⁹

Note that simply arguing that the only effect of an instrument on the outcome variable of interest is via the treatment of interest does not suffice to establish the exclusion restriction. Even if the claim that Z only affects Y via its effect on X is true, the researcher also needs to argue that variation in the instrument (Z) is “as if” random. For example, suppose that the only effect of Z on Y occurs via X , but Z is a function of a variable W that is also

¹⁸ These are Cannon [2014], Cohen et al. [2014], Kim, Mauldin, and Patro [2014], Vermeer, Edmonds, and Asthana [2014], Fox, Luna, and Schaur [2014], Guedhami, Pittman, and Saffar [2014], Houston et al. [2014], de Franco et al. [2014], Erkens, Subramanyam, and Zhang [2014], and Correia [2014].

¹⁹ Though some have questioned the exclusion restriction even in this case, arguing that the outcome of the draft lottery may have caused some, for example, to move to Canada (see Imbens and Rubin [2015]).

associated with Y . In this case, IV estimates of the effect of X on Y will be biased. Thus, a researcher should also account for the sources of variation in the chosen instrument and why these are not expected to be associated with variation in the outcome variable.²⁰

Unfortunately, there are few (if any) accounting variables that meet the requirement that they randomly assign observations to treatments, and do not affect the outcome of interest outside of effects on the treatment variable. Sometimes researchers turn to lagged values of endogenous variables or industry averages as instruments, but these too are subject to criticism.²¹

3.2.2. There Are No Simple (Statistical) Tests for the Validity of Instruments. Some accounting researchers appear to believe that statistical tests can resolve the question of whether their instrument is “valid.” Indeed, many studies choose to test the validity of their IVs using statistical tests (see Larcker and Rusticus [2010]). But such tests of instruments are of dubious value. Consider, for example, the following simulation of a setting where X does not cause y , but we nevertheless estimate the regression $y = X\beta + \epsilon$. That is, we estimate a regression model where $\beta = 0$. To make matters interesting, suppose $\rho(X, \epsilon) > 0$ (i.e., X is correlated with the error). Clearly, if we estimated the equation by OLS, we would conclude that there is a (positive) relationship between X and y . Suppose that, after being told that X is “endogenous,” we found three instruments: z_1 , z_2 , and z_3 . Unbeknown to us, the three instruments were determined as follows: $z_1 = X + \eta_1$, $z_2 = \eta_2$, and $z_3 = \eta_3$, with $\eta_1, \eta_2, \eta_3 \sim N(0, \sigma_\eta^2)$ and independent. That is, z_1 is X plus noise (e.g., industry averages or lagged values of X would seem to approximate z_1), while z_2 and z_3 are random noise (many variables could be candidates here).

Assuming that X and ϵ are bivariate-normally distributed with variance of 1 and $\rho(X, \epsilon) = 0.2$, and $\sigma_\eta = 0.03$, we performed 1,000 IV regression simulations with 1,000 firm-level observations in each case. Both OLS and IV coefficients are close, with the IV-estimated coefficient averaging 0.201. The IV coefficient estimates are statistically significant at the 5% level 100% of the time.²² Based on a test statistic of 30, which easily exceeds the thresholds suggested by Stock, Wright, and Yogo [2002], the null hypothesis of weak instruments is rejected 100% of the time. The Sargan [1958] test of overidentifying restrictions fails to reject a null hypothesis of valid instruments (at the 5% level) 95.7% of the time.

This example illustrates why it is that no statistical test allows the researcher to verify that their instruments satisfy the exclusion restriction.²³

²⁰ In the case of Angrist [1990], this was plausibly satisfied using a lottery for assignment of Z to subjects.

²¹ See Reiss and Wolak [2007] for a discussion regarding the implausibility of general claims that industry averages are valid instruments.

²² Note that this coefficient is close to $\rho(X, \epsilon) = 0.2$, which is to be expected, given how the data were generated.

²³ This is a corollary of the “causal reasoning is not statistical reasoning” point made above.

Obviously, causal inferences based on such IVs is completely inappropriate. Yet, this shows that it is quite possible for completely spurious instruments to deliver bad inferences, yet easily pass tests for weak instruments and tests of overidentifying restrictions.

3.2.3. Causal Diagrams Can Clarify Causal Reasoning. To illustrate the application of causal diagrams to the evaluation of IVs, we consider Armstrong, Gow, and Larcker [2013]. Armstrong, Gow, and Larcker study the effect of shareholder voting ($Shareholder\ support_t$) on future executive compensation ($Comp_{t+1}$). Because of the plausible existence of unobserved confounding variables that affect both future compensation and shareholder support, a simple regression of $Comp_{t+1}$ on $Shareholder\ support_t$, and controls would not allow Armstrong, Gow, and Larcker [2013] to obtain an unbiased or consistent estimate of the causal relation. Among other analyses, Armstrong, Gow, and Larcker [2013] use an IV to estimate the causal relation of interest. Armstrong, Gow, and Larcker [2013] claim that their instrument is valid. Their reasoning is represented graphically in figure 2. By conditioning on $Comp_{t-1}$ and using Institutional Shareholder Services (ISS) recommendations as an instrument, Armstrong, Gow, and Larcker [2013] argue that they can identify a consistent estimate of the causal effect of shareholder voting on $Comp_{t+1}$, even though there is an unobserved confounder, namely determinants of future compensation observed by shareholders, but not the researcher.²⁴

While the authors note that “validity of this instrument depends on ISS recommendations not having an influence on future compensation decisions conditional on shareholder support (i.e., firms listen to their shareholders, with ISS having only an indirect impact on corporate policies through its influence on shareholders’ voting decisions),” they are unable to test the assumption (Armstrong, Gow, and Larcker [2013, p. 912]). Unfortunately, this assumption seems inconsistent with the findings of Gow et al. [2013], who provide evidence that firms calibrate compensation plans (i.e., factors that directly affect $Comp_{t+1}$) to comply with ISS’s policies so as to get a favorable recommendation from ISS. As depicted in figure 2(B), this implies a path from $ISS\ recommendation_t$ to $Comp_{t+1}$ that does not pass through $Shareholder\ support_t$, suggesting that the instrument of Armstrong, Gow, and Larcker [2013, p. 912] is not valid.²⁵

3.2.4. IVs in Accounting Research: An Evaluation. A review of IV applications in our 2014 survey suggests that accounting researchers have paid

²⁴In figure 2, we depict the unobservability of this variable (to the researcher) by putting it in a dashed box. Note that we have omitted the controls included by Armstrong, Gow, and Larcker [2013] for simplicity, though a good causal analysis would consider these carefully.

²⁵Armstrong, Gow, and Larcker [2013] recognize the possibility that the instrument they use is not valid and conduct sensitivity analysis to examine the robustness of their result to violation of the exclusion restriction assumptions. This analysis suggests that their estimate is highly sensitive to violation of this assumption.

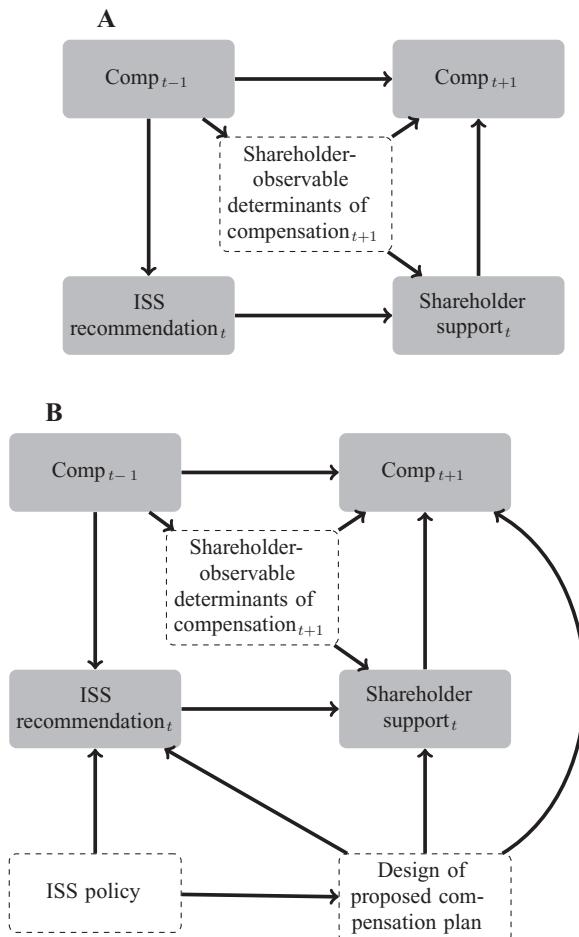


FIG. 2.—Identifying effects of shareholder support on compensation. (A) Causal diagram for Armstrong, Gow, and Larcker [2013] and (B) alternative causal diagram for Armstrong, Gow, and Larcker [2013].

little heed to the suggestions and warnings of Larcker and Rusticus [2010], Lennox, Francis, and Wang [2012], and Roberts and Whited [2013]. This is perhaps not surprising, as most studies do not have a theoretical model that can explain why a variable can naturally be excluded from the equation of interest but still matter. Thus, while instruments work in theory, in practice there remains a substantial burden of proof on researchers to justify the assumptions that justify IV estimators.

3.3. REGRESSION DISCONTINUITY DESIGNS

Recently, regression discontinuity (RD) designs have attracted the interest of accounting researchers, as a number of phenomena of interest to

accounting researchers involve discontinuities. For example, whether an executive compensation plan is approved is a discontinuous function of shareholder support (e.g., Armstrong, Gow, and Larcker [2013]) and whether a firm initially had to comply with provisions of the Sarbanes-Oxley Act was a discontinuous function of market float (Iliev [2010]).

In discussing the recent “flurry of research” using RD designs in other fields, Lee and Lemieux [2010, p. 282] point out that they “require seemingly mild assumptions compared to those needed for other nonexperimental approaches . . . and that causal inferences from RD designs are potentially more credible than those from typical ‘natural experiment’ strategies.” While RD designs make relatively mild assumptions, in practice these assumptions may be violated. In particular, manipulation of the running variable (or the variable that determines whether an observation is assigned to a treatment) may occur and researchers should carefully examine their data for this possibility (see, e.g., Listokin [2008], McCrary [2008]).

Another issue with RD designs is that the causal effect estimated is a local estimate (i.e., it relates to observations close to the discontinuity). This effect may be very different from the effect at points away from the discontinuity. For example, in designating a public float of \$75 million, the SEC may have reasoned that at that point the benefits of Sarbanes–Oxley were approximately equal to the fixed costs of complying with the law. If true, we would expect to see an estimate of approximately zero effect, even if there were substantial benefits of the law for shareholders of firms having a public float well above the threshold.

Another critical assumption is the bandwidth used in estimation (i.e., in effect how much weight is given to observations according to their distance from the cutoff). We encourage researchers using RD designs to employ methods that exist to estimate optimal bandwidths and the resulting estimates of causal effects (e.g., Imbens and Kalyanaraman [2012]).

Finally, one strength of RD designs is that the estimated relation is often effectively univariate and easily plotted. As suggested by Lee and Lemieux [2010], it is highly desirable for researchers to plot both underlying data and fitted regression functions around the discontinuity. This plot will enable readers to evaluate the strength of the results. If there is a substantive impact associated with the treatment, this should be obvious from a plot of the actual data and the associated fitted function.

3.4. OTHER METHODS

3.4.1. Difference-in-Differences and Fixed-Effect Estimators. Accounting researchers have come to view some statistical methods as requiring fewer assumptions and thus being less subject to problems when it comes to drawing causal inferences. Angrist and Pischke [2010, p. 12] include so-called “DD estimators” on their list of such quasi-experimental methods, along with “IV and RD methods.”²⁶ Enthusiasm for DD designs perhaps stems

²⁶ As Angrist and Pischke [2008, p. 228] argue that “DD is a version of fixed effects estimation,” we discuss these methods together.

from a belief that these are “quasi-experimental” methods in the same sense as the other two approaches cited by Angrist and Pischke [2010, p. 12]. But the essential feature that IVs and RD methods rely on is the “as if” random treatment assignment mechanism. If treatment assignment is driven by unobserved confounding variables, then DD and fixed-effect estimates of causal effects will be biased and inconsistent. As few settings in accounting satisfy random treatment assignment, there is a heavy burden on researchers using DD or fixed-effect estimators to explain why they believe these methods allow them to recover unbiased or consistent estimates of causal effects.

Proponents of DD methods argue that they rely on the relatively innocuous assumption of “parallel trends.” But it is far from clear that this assumption is actually a mild one. First, it is a highly parametric assumption: parallel trends might hold for levels of a variable, but that does not mean they would hold for log-transformations of the variable. Second, many variables of interest to accounting researchers are mean-reverting, which is inconsistent with parallel trends when treatment and control observations differ in pretreatment outcomes. Third, as DD studies typically rely on some kind of quasi-natural experiment, the existence of pretreatment differences raises questions about the claimed “as if” random assignment to treatment and control. For example, the frequently cited study of Kelly and Ljungqvist [2012] uses supposedly random shocks to brokerage coverage and a DD design. But the existence of a 0.039 difference in spreads between treatment and matched control firms suggests that the assignment was far from random.²⁷

The causal interpretation of regressions that use fixed effects to control for unobservable differences in observations also can be problematic, particularly when there are heterogeneities in treatment effects. If the true effect is positive for some units (e.g., firms) and negative for others, then, depending on the composition of the sample, the sign of the effect can be positive, negative, or indistinguishable from zero. Additionally, if units self-select into a binary treatment for the entire sample period, then a fixed-effect estimator will not use these observations in estimating the effect, even though these might plausibly be the observations with the greatest treatment effect.

Heterogeneity in effects is not the only problem that fixed-effect strategies cannot necessarily handle. In particular, when there are complex relations between unobservables and treatments, as is likely to be the case in many accounting research settings, it is unclear what a fixed-effect strategy would produce. If time-invariant heterogeneity is correlated with potential outcomes, then fixed-effect estimators can have greater bias than estimators that omit fixed effects.

²⁷ See Kelly and Ljungqvist [2012, p. 1388, table 2]. This pretreatment difference is material, given the estimated treatment effect of 0.020. Perhaps recognizing this issue, the subsequent paper by Balakrishnan et al. [2014] matches on pretreatment values.

In our view, accounting researchers need to be much more careful using and interpreting fixed-effect estimators. In particular, researchers need to clearly demonstrate how their fixed-effect estimates are related to the causal effect of interest, particularly when that effect could differ across observations.

3.4.2. Propensity-Score Matching. Another method that has become popular in accounting research is PSM. Regression methods can be viewed as making model-based adjustments to address confounding variables. Stuart and Rubin [2007, p.157] argue that:

[M]atching methods are preferable to these model-based adjustments for two key reasons. First, matching methods do not use the outcome values in the design of the study and thus preclude the selection of a particular design to yield a desired result. Second, when there are large differences in the covariate distributions between the groups, standard model-based adjustments rely heavily on extrapolation and model-based assumptions. Matching methods highlight these differences and also provide a way to limit reliance on the inherently untestable modeling assumptions and the consequential sensitivity to those assumptions.

For these reasons, PSM methods can prove useful when faced with observational data. However, PSM does *not* provide “the closest archival approximation to a true random experiment” and does *not* represent “the most appropriate and rigorous research design for testing the effects of an *ex ante* treatment” (Kirk and Vincent [2014, p. 1429]). Rosenbaum [2009, pp. 73–75] points out that matching is “a fairly mechanical task,” and when assignment to treatment is driven by unobservable variables, PSM-based estimates may be biased as much as regression estimates. We agree with Minutti-Meza [2014], who argues that “matching does not necessarily eliminate the endogeneity problem resulting from unobservable variables driving [treatment] and [outcomes].”

3.5 QUASI-EXPERIMENTAL METHODS: AN EVALUATION

We agree that the revolution in econometric methods for causal inference represents an opportunity for accounting researchers. However, the assumptions required for these methods to deliver credible estimates of causal effects are unlikely to be met in many applications that rely on observational data. In this regard, we echo the observation in Leuz and Wysocki [2016, p. 29] that “finding valid instruments to implement selection models and IV regressions is very difficult.”

Given the dominance of causal questions and observational data in accounting research, and the difficulty researchers will face in applying quasi-experimental methods in accounting research, our appraisal may seem disappointing. Yet, these methods can be used in certain settings. In what follows, we offer some alternative paths that accounting researchers might consider going forward.

4. Causal Mechanisms, Causal Inference, and Descriptive Studies

In the first part of this paper, we have argued that, while causal inference is the goal of most accounting research, it is extremely difficult to find settings and statistical methods that can produce credible estimates of causal effects. Does this mean accounting researchers must give up making causal statements? We believe the answer is no. There are viable paths forward. The objective of the second part of this paper is to discuss these paths.

The first path we discuss is an increased focus on causal mechanisms. Accounting research is not alone in its reliance on observational data with the goal of drawing causal inferences. It is, therefore, natural to look to other fields using observational data to identify causal mechanisms and ultimately to draw causal inferences. Epidemiology and medicine are two fields that are often singled out in this regard. In what follows, we briefly provide examples and highlight the features of the examples that enhanced the credibility of the inferences drawn. A key implication of this discussion is that accounting researchers need to identify clearly and rigorously the causal mechanism that is producing their results.

4.1 JOHN SNOW AND CHOLERA

A widely cited case of successful causal inference is John Snow's work on cholera. As there are many excellent accounts of Snow's work, we will focus on the barest details. As discussed in Freedman [2009, p. 339], "John Snow was a physician in Victorian London. In 1854, he demonstrated that cholera was an infectious disease, which could be prevented by cleaning up the water supply. The demonstration took advantage of a natural experiment. A large area of London was served by two water companies. The Southwark and Vauxhall company distributed contaminated water, and households served by it had a death rate 'between eight and nine times as great as in the houses supplied by the Lambeth company,' which supplied relatively pure water." But there was much more to Snow's work than the use of a convenient natural experiment. First, Snow's reasoning (much of which was surely done before "the arduous task of data collection" began) was about the mechanism through which cholera spread. Existing theory suggested "odors generated by decaying organic material." Snow reasoned qualitatively that such a mechanism was implausible. Instead, drawing on his medical knowledge and the facts at hand, Snow conjectured that "a living organism enters the body, as a contaminant of water or food, multiplies in the body, and creates the symptoms of the disease. Many copies of the organism are expelled with the dejecta, contaminate water or food, then infect other victims" (Freedman [2009, p. 342]).

With a hypothesis at hand, Snow then needed to collect data to prove it. His data collection involved a house-to-house survey in the area surrounding the Broad Street pump operated by Southwark and Vauxhall. As part of his data collection, Snow needed to account for anomalous cases (such as the brewery workers who drank beer, not water). It is important to note

that this qualitative reasoning and diligent data collection were critical elements in establishing (to a modern reader) the “as if” random nature of the treatment assignment mechanism provided by the Broad Street pump. Snow’s deliberate methods contrast with a shortcut approach, which would have been to argue that in his data he had a natural experiment.

Another important feature of this example is that widespread acceptance of Snow’s hypothesis did not occur until compelling evidence of the precise causal mechanism was provided. “However, widespread acceptance was achieved only when Robert Koch isolated the causal agent (*Vibrio cholerae*, a comma-shaped bacillus) during the Indian epidemic of 1883” (Freedman [2009, p. 342]). Only once persuasive evidence of a plausible mechanism was provided (i.e., direct observation of microorganisms now known to cause the disease) did Snow’s ideas become widely accepted.

We expect the same might be true in the accounting discipline if researchers carefully articulate the assumed causal mechanism for their observations. It is, of course, necessary for researchers to show that the proposed mechanism is actually consistent with behavior in the institutional setting being examined. As we discuss below, detailed descriptive studies of institutional phenomenon provide an important part of the information to evaluate the proposed mechanism.

4.2 SMOKING AND HEART DISEASE

A more recent illustration of plausible causal inference is discussed by Gillies [2011]. Gillies focuses on the paper by Doll and Peto [1976], which studies the mortality rates of male doctors between 1951 and 1971. The data of Doll and Peto [1976] showed “a striking correlation between smoking and lung cancer” (Gillies [2011, p. 111]). Gillies [2011] argues that “this correlation was accepted at the time by most researchers (if not quite all!) as establishing a causal link between smoking and lung cancer.” Indeed Doll and Peto [1976, p. 1535] themselves say explicitly that “the excess mortality from cancer of the lung in cigarette smokers is caused by cigarette smoking.” In contrast, while Doll and Peto [1976] had highly statistically significant evidence of an association between smoking and heart disease, they were cautious about drawing inferences of a direct causal explanation for the association. Doll and Peto [1976, p. 1528] point out that “to say that these conditions were related to smoking does not necessarily imply that smoking caused . . . them. The relation may have been secondary in that smoking was associated with some other factor, such as alcohol consumption or a feature of the personality, that caused the disease.”

Gillies [2011] then discusses extensive research into atherosclerosis between 1979 and 1989 and concludes that “by the end of the 1980s, it was established that the oxidation of LDL was an important step in the process which led to atherosclerotic plaques.” Later research provided “compelling evidence” that smoking causes oxidative modification of

biologic components in humans.²⁸ Gillies [2011, p. 120] points out that this evidence alone did not establish a confirmed mechanism linking smoking with heart disease, because the required oxidation needs to occur in the artery wall, not in the blood stream, and it fell to later research to establish this missing piece.²⁹ Thus, through a process involving multiple studies over two decades, a plausible set of causal mechanisms between smoking and atherosclerosis was established.

Gillies [2011] avers that the process by which a causal link between smoking and atherosclerosis was established illustrates the “Russo–Williamson thesis.” Russo and Williamson [2007, p. 159] suggest that “mechanisms allow us to generalize a causal relation: while an appropriate dependence in the sample data can warrant a causal claim ‘*C* causes *E* in the sample population,’ a plausible mechanism or theoretical connection is required to warrant the more general claim ‘*C* causes *E*.’ Conversely, mechanisms also impose negative constraints: if there is no plausible mechanism from *C* to *E*, then any correlation is likely to be spurious. Thus mechanisms can be used to differentiate between causal models that are underdetermined by probabilistic evidence alone.”

The Russo–Williamson thesis was arguably also at work in the case of Snow and cholera, where the establishment of a mechanism (i.e., *Vibrio cholerae*) was essential before the causal explanation offered by Snow was widely accepted. It also appears in the case of smoking and lung cancer, which was initially conjectured based on correlations, prior to a direct biological explanation being offered.³⁰

4.3 CAUSAL MECHANISMS IN ACCOUNTING RESEARCH

Our view is that accounting researchers can learn from fields such as epidemiology, medicine, and political science.³¹ These fields grapple with observational data and eventually draw inferences that are causal. While randomized controlled trials are a gold standard of sorts in epidemiology, in many cases it is unfeasible or unethical to use such trials. For example, in political science, it is not possible to randomly assign countries to treatment conditions such as democracy or socialism. Nevertheless, these fields have often been able to draw plausible causal inferences by establishing clear mechanisms, or causal pathways, from putative causes to putative effects.

²⁸ This evidence is much higher levels of a new measure (levels of *F*₂-isoprostanes in blood samples) of the relevant oxidation in the body due to smoking. This conclusion was greatly strengthened by the finding that levels of *F*₂-isoprostanes in the smokers “fell significantly after two weeks of abstinence from smoking” (Morrow et al. [1995, pp. 1201 and 1202]).

²⁹ “Smoking produced oxidative stress. This increased the adhesion of leukocytes to the . . . artery, which in turn accelerated the formation of atherosclerotic plaques” (Gillies [2011, p. 123]).

³⁰ The persuasive force of Snow’s natural experiment, coming decades before the work by Neyman [1923] and Fisher [1935], might be considered greater today.

³¹ In this regard, we echo the suggestion by Leuz and Wysocki [2016] that it “might be useful for regulators, policy makers and academics to study the experience in medicine.”

One paper that has a fairly compelling identification strategy is Brown, Stice and White [2015], which examines “the influence of mobile communication on local information flow and local investor activity using the enforcement of state-wide distracted driving restrictions.” The authors find that “these restrictions . . . inhibit local information flow and . . . the market activity of stocks headquartered in enforcement states.” Miller and Skinner [2015, p. 229] suggest that “given the authors’ setting and research design, it is difficult to imagine a story under which the types of reverse causality or correlated omitted variables explanations that we normally worry about in disclosure research are at play.” However, notwithstanding the apparent robustness of the research design, the results would be much more compelling if there were more detailed evidence regarding the precise causal mechanism through which the estimated effect occurs and the authors appear to go to lengths to provide such an account.³² For example, evidence of trading activity by local investors while driving prior to, but not after, the implementation of distracted driving restrictions would add considerable support to conclusions in Brown, Stice and White [2015].³³

As another example, many published papers have suggested that managers adopt conditional conservatism as a reporting strategy to obtain benefits such as reduced debt costs. However, as Beyer et al. [2010, p. 317] point out, an *ex ante* commitment to such a reporting strategy “requires a mechanism that allows managers to credibly commit to withholding good news or to commit to an accounting information system that implements a higher degree of verification for gains than for losses,” yet research has only recently begun to focus on the mechanisms through which such commitments are made (e.g., Erkens, Subramanyam, and Zhang [2014]).

It is very clear that we need a much better understanding of the precise causal mechanisms for important accounting research questions. A clear discussion of these mechanisms will enable reviewers and readers to see what is being assumed and assess the reasonableness of the theoretical causal mechanisms.

³² Brown, Stice and White [2015, pp. 277 and 278] “argue that constraints on mobile communication while driving could impede or delay the collection and diffusion of local stock information across local individuals. Anecdotal evidence suggests that some individuals use car commutes as opportune times to gather and disseminate stock information via mobile devices. For instance, some commuters use mobile devices to collect and pass on stock information either electronically or by word of mouth to other individuals within their social network. Drivers also use mobile devices to wirelessly check stock positions and prices in real-time, stream the latest financial news, or listen to earnings calls.”

³³ Note that the authors disclaim reliance on trading while driving: “our conjectures do not depend on the presumption that local investors are driving when they execute stock trades . . . [as] we expect such behavior to be uncommon.” However, even if not *necessary*, given the small effect size documented in the paper (approximately 1% decrease in volume), a small amount of such activity could be *sufficient* to provide a convincing account in support of their results.

4.4 DESCRIPTIVE STUDIES

Accounting is an applied discipline and it would seem that most empirical research studies should be solidly grounded in the details of how institutions operate. These descriptions can form a basis for identifying and justifying causal mechanisms for explaining empirical results. Unfortunately, there are very few studies published in top accounting journals that focus on providing detailed descriptions of institutions in accounting research settings. Part of this likely reflects the perception that research that pursues causal questions (i.e., tests of theories) is more highly prized, and thus more likely to be published in top accounting journals.³⁴ We believe that accounting research can benefit substantially from more in-depth descriptive research. As we discuss below, this type of research is essential to improve our understanding of causal mechanisms and develop structural models.³⁵

One reason to value descriptive research is that it can uncover realistic structures and mechanisms that would be exceedingly difficult to arrive at from basic economic theory or the simple intuition of the researcher. In the compensation area, the early research by Lewellyn [1968] and the more recent work by Frydman and Saks [2010] are also essentially descriptive studies that caused researchers to explore why certain patterns of remuneration arrangements are used, revised, or eliminated over time. These types of data motivate researchers to frame research studies that have the potential to uncover the causal mechanisms that produce these institutional observations.

A good example in the accounting literature is the study by Healy [1985]. Using proxy statement disclosures and conversations with actual executives and consultants, Healy [1985] studies the bonus contracts of 94 large U.S. companies and identifies a common structure of these bonus plans, including the existence of caps and floors. The paper also suggests hypotheses worth investigating regarding the effects of these plan features on accounting decisions. It seems highly unlikely that a model derived from fundamental economic theory would arrive at these plan features found in his data.

Another example is work by Smith and Warner [1979], Kalay [1982], and many others who look at debt covenant provisions. Institutional knowledge

³⁴ At one point, the *Journal of Accounting Research* published papers in a section entitled “Capsules and Comments.” The editor at the time (Nicholas Dopuch) would seem to place a paper into this section if it “did not fit” as a main article, but examined new institutional data or ideas. Such a journal section might have provided a credible signal of a willingness to publish descriptive studies of institutionally interesting settings.

³⁵ There are many “classic” descriptive studies that have had a major impact on subsequent theoretical and empirical research in organizational behavior and strategy (e.g., Cyert, Simon, and Trow [1956], Mintzberg [1973], Bower [1986]). Cyert, Simon, and Trow [1956] argue that “a realistic description and theory of the decision-making process are of central importance to business administration and organization theory. Moreover, it is extremely doubtful whether . . . economics does in fact provide a realistic account of decision-making in large organizations operating in a complex world.”

about debt covenants has generated hypotheses about managerial wealth and accounting manipulation. Moreover, descriptive statistics regarding covenants also provided Dichev and Skinner [2002] with the data to show that leverage is not a valid proxy for “closeness to covenant.” This is an important finding because the empirical literature to this point simply assumed that leverage was a reliable and valid proxy for potential covenant violations. An in-depth examination of actual debt covenants and an understanding of how covenant violations are dealt with by financial institutions would have substantially improved much of the research on how debt covenants influence firm behavior (i.e., so-called “positive theory” research).

In the corporate governance area, the descriptive data on board of director interlocks in Brandeis [1913], U.S. Federal Trade Commission [1951], and U.S. Congress Senate Committee on Governmental Affairs and Ribicoff [1978] provided novel descriptive insights into the structure of boards of directors. These and other similar studies had an important impact on starting the large literature on how boards of directors function. Similarly, the initial collection of equity ownership by executives, directors, and large shareholders by the Securities and Exchange Commission [1936] enabled researchers to understand the extent to which ownership is separated from control, and examine the implications of the classic Berle and Means [1932] hypotheses regarding economic activity.

Descriptive data on antitakeover provisions collected by the Investor Responsibility Research Center (IRRC) have provided the basis for a considerable amount of research on the market for corporate control. Gompers, Ishii, and Metrick [2003], Bebchuk, Cohen, and Ferrell [2009], and many others use these data to form and test a multitude of research questions related to corporate governance. Perhaps more importantly, Daines and Klausner [2001] provided an institutionally grounded examination of how these specific antitakeover provisions actually work from a legal perspective (which contrasts with conjectures made by researchers in other disciplines). The Daines and Klausner [2001] analysis provides a good example of how descriptive data combined with institutional and legal knowledge can provide appropriate insights into the workings of corporate governance.

The descriptive disclosure data compiled by the Association for Investment Management and Research (AIMR) have had a similar impact on financial accounting research. These ratings reflect the assessments of analysts specializing in specific industries as to the informativeness of disclosures made by firms. The ratings data have provided a variety of useful information about differences in disclosure practices across firms, industries, and time. We suspect that these statistics were instrumental in motivating Lang and Lundholm [1993, p. 6], Healy, Hutton, and Palepu [1999], and many others. They provided new insights into whether firm disclosure is associated with performance, consensus among investors, stock liquidity, and other important outcome variables. In related work, Groysberg, Healy, and Maber [2011] provide an informative analysis of how analysts are

compensated using descriptive proprietary data and statistical analyses to uncover the fundamental features of the reward system.

Recently published research suggests an increased recognition of the value of descriptive research. Soltes [2014] examines the interactions between sell-side analysts and company management in one firm that granted proprietary access to its data to “offer insights into which analysts privately meet with management, when analysts privately interact with management, and why these interactions occur.” By comparing private interaction to observed interaction between analysts and managers on conference calls, and highlighting that private interaction with management is an important communication channel for analysts, Soltes [2014] suggests a plausible mechanism through which information transfers actually occur.

That private communication with management is an important source of information is confirmed by Brown et al. [2015]. Brown et al. survey and interview financial analysts to understand how they think about a variety of issues. Their findings suggest that analysts’ views on earnings quality differ from those most researchers explore. For instance, analysts do not use the “red flags” used by academics to identify manipulation. Analysts also generally are not attempting to uncover manipulation and use forecasts to figure out a stock price target. These insights should shape research seeking to develop hypotheses and models of accounting information and analyst behavior. Despite the dearth of descriptive research in top accounting journals, we believe that our discipline can benefit substantially from this style of research. An interesting question is what makes a descriptive study an important contribution that should be published in a top journal. An obvious required attribute is that the descriptive study examines an interesting institutional question where researchers care about understanding the phenomenon producing the observations. Stated differently, would anyone change their research agenda or their (causal) interpretations of prior work if provided with these descriptive results?

The descriptive research needs to be neutral and unbiased in terms of data collection and interpretations. If expert opinions are used, can we be assured that the opinions are not biased because of their business dealings? Data collected using surveys or interviews by consulting firms may provide great descriptive data, but researchers need to be convinced that the data are not confounded by selection bias or other sampling concerns.

The research should also provide deep insight into the causal mechanisms underlying observed institutional data. There may well be alternative mechanisms suggested by the research, and these alternatives may be a function of nuances and contextual variables for the setting. Provided the researcher is clear that their aim is description and not the last word on causality, the presence of several alternative explanations should not detract from the insight of the descriptive work.

Obviously, the evaluation of descriptive research is somewhat subjective, but the evaluation of more traditional accounting research is similarly subjective. As a discipline, we do not have much recent experience assessing

descriptive research, and we are unfamiliar with recent advances in descriptive methods, such as nonparametric regression. However, given the possibility that descriptive research can help us begin to think about causal mechanisms, it should be encouraged and accepted in the top accounting journals.

5. Structural Modeling

5.1 STRUCTURAL MODELING: AN OVERVIEW

In sections 2 and 3, we suggested that researchers minimally consider using diagrams to communicate the basis for their causal inferences, and in section 4, we suggested that researchers be more precise in describing how their data permit causal inferences. This section explores a formal approach to developing a causal model, namely, the “structural” approach. Structural models are empirical models that are derived from theoretical models of behavior. The term structural model originated with economists and statisticians working at the Cowles Foundation in the 1940s and 1950s. The earliest structural models used economic models of consumer and producer behavior to derive demand and supply equations. By adding an equilibrium condition, such as quantities demanded equal quantities supplied, economists obtained a set of mathematical equations that could be used to understand movements in observed prices and quantities. A question then arose as to whether economists could reverse-engineer this modeling process and use observed prices and quantities to recover the underlying demand and supply relations. The models made it clear that the empiricist could only recover estimates of the unobserved demand and supply equations if certain exogenous (IV) variables were available.

The impact of these early models on empirical work in economics encouraged other social scientists to begin using theoretical models to interpret data. Structural models have found widest application in situations where causality is an issue, such as the determinants of educational choices, voting, contraception, addiction, and financing decisions. Other applications of structural models are discussed in Reiss and Wolak [2007] and Reiss [2011].

A structural empirical model comprises a theoretical model of the phenomenon of interest and a stochastic model that links the theoretical model to the observed data. The theoretical model minimally describes who makes decisions, the objectives of decision-makers, and constraints on their behavior. In developing and analyzing the theoretical model, the researcher decides what conditions (variables) matter and what is endogenous and exogenous. While the theoretical model typically draws on economic principles, it could also be derived from behavioral theories in other fields, such as psychology and sociology.³⁶

³⁶ Some researchers refer to any mathematical model fit to data as a structural model. For instance, one might assume that the number of restatements in an industry follows a

Structural models offer a number of benefits for empirical researchers. First, structural modeling is a process that forces a researcher to make explicit assumptions about what determines behavior and outcomes (i.e., the causal mechanism). Second, structural models make it clear what data are needed to identify unobserved parameters and random variables, such as coefficients of risk aversion. Third, structural models provide a foundation for estimation and inference. Finally, structural models facilitate counterfactual analyses, such as what might happen under conditions not observed in the data. To illustrate these benefits, as well as some of their limitations, we next explore an accounting application.

5.2 STRUCTURAL MODELS IN ACCOUNTING: AN ILLUSTRATION

This section develops a model of managerial incentives to misstate accounting information. This topic has been the focus of many papers in recent years (see the review in Armstrong, Jagolinzer, and Larcker [2010]). The key question in this literature is whether certain kinds of managerial incentives increase the tendency for managers to misstate (or attempt to misstate) financial information. A number of papers hypothesize that tying managers' compensation to the information that they provide will increase their desire to misstate that information. However, some researchers suggest that, by aligning the long-term interests of shareholders and managers, certain kinds of incentives could actually *reduce* misstatements (Burns and Kedia [2006]).

Efendi, Srivastava, and Swanson [2007] illustrate a fairly typical descriptive empirical paper in this literature. Efendi, Srivastava, and Swanson [2007, p. 687] estimate a logistic regression with an indicator for restatements as the dependent variable and measures of CEO incentives as independent variables of interest, along with controls for firm size, financial structure, and corporate governance proxies.³⁷

A key assumption implicit in much of this literature is that restatements are a good proxy for actual *misstatements* (e.g., Efendi, Srivastava, and Swanson [2007], Armstrong, Jagolinzer, and Larcker [2010]). This assumption is made because, in practice, accounting researchers only observe misstatements that are detected and corrected by external monitors after the financial statements were issued. Examples of these external monitors include whistleblowers, regulators, media, and others (e.g., Dyck, Morse, and Zingales [2010]). For simplicity, we refer to the actions of these external monitors collectively as "subsequent investigations." If subsequent

Poisson process and then fit the parameters of the Poisson model using industry-level data on restatements. We do not view such models as structural because they lack specific behavioral or institutional components that permit a causal inference. We would classify this approach as descriptive or statistical modeling.

³⁷ Efendi, Srivastava, and Swanson [2007] also employ a case-control design that involves matching firms with restatements with firms without. We do not focus on that aspect of their research design in our discussion here.

investigations are perfect and detect all misstatements, then there is a one-to-one correspondence between misstatements and restatements.³⁸ Realistically, these subsequent investigations are not perfect, meaning that we need to recognize the difference between misstatements and restatements when estimating the effect of managerial incentives on misstatements.

In the following analysis, we consider two alternative models of the causal mechanism linking managerial incentives to accounting restatements. Each model explicitly considers the incentives of the manager and the role of the external auditor. The two models, however, lead to different conclusions about how CEO incentives affect restatements. These differences permit us to illustrate the value of having a theoretical model that can interpret competing empirical estimates, as well as the difficulty of interpreting estimates in the absence of such models.

5.2.1. Model 1: A Nonstrategic Auditor Model. We assume that firm misstatements are deliberate and are made by a single agent, whom we refer to as the “CEO.” The CEO is assumed to be rational in the sense that he or she trades off private expected benefits and costs of misstatements when deciding whether to misstate. Specifically, suppose that the CEO receives a benefit of B^* from the *successful* manipulation of earnings (i.e., a misstatement that is not detected either by the firm’s auditors before a report is released or by subsequent investigations).

Besides the CEO, we assume that the firm’s auditors independently detect and correct attempted misstatements at a constant rate p_A and that the (conditional) probability of subsequent investigations catching a misstatement is p_I . Given these assumptions, the probability of a misstatement getting past the firm’s auditor and subsequent investigations is $(1 - p_A) \times (1 - p_I)$. The CEO’s expected benefit from a successful misstatement is then

$$B^* = (1 - p_I) \times (1 - p_A) \times B,$$

where B is a gross benefit to the manager from a misstatement.

Assume the CEO must exert a fixed cost of effort C_M in order to misstate performance. Combining this cost with the manager’s expected benefits from of misstatement gives

$$y_M^* = \begin{cases} \text{Misstate} & \text{if } (1 - p_I) \times (1 - p_A) \times B - C_M \geq 0 \\ \text{Don't misstate,} & \text{otherwise.} \end{cases} \quad (3)$$

This (structural) inequality describes the unobserved misstatement process. In general, researchers will not observe the structural parameters of interest: B , C_M , p_A , or p_I .

To complete the structural model and recover these parameters, the researcher must add assumptions that relate the parameters to the data

³⁸ There will still be a difference between *attempted* misstatements and actual misstatements due to the external auditor correcting some attempted misstatements.

available. Suppose we only observe a (zero-one) indicator variable y for restatements. These restatements are the result of three decisions:

- 1) The manager misstates (or not).
- 2) The firm auditor detects and corrects an attempted misstatement (or not).
- 3) A subsequent investigation detects a misstatement and a restatement occurs (or not).

Mathematically, this sequence can be modeled as

$$y = I(\text{Restate}) = I(y_M^* \geq 0) \times (1 - I(y_A^* \geq 0)) \times I(y_I^* \geq 0), \quad (4)$$

where $I(\cdot)$ is a zero-one indicator function equaling 1 when the condition in parentheses is true. The unobserved variables y_M^* , y_A^* , and y_I^* reflect the criteria that underlie the CEO's, firm's auditor's, and subsequent investigators' decisions. Note that equation (4) uses $(1 - I(y_A^* \geq 0))$, an indicator for the firm's auditor missing the misstatement.

Equation (4) somewhat resembles a traditional binary discrete choice model. The easiest way to see this is to take expectations (from the researcher's standpoint). Assuming the decision variables are independent,

$$\begin{aligned} E(y) &= E [I(y_M^* \geq 0) \times (1 - I(y_A^* \geq 0)) \times I(y_I^* \geq 0)] \\ &= \Pr(\text{Misstate}) \times \Pr(\text{Auditor Misses}) \times \Pr(\text{Investigation Finds}) \\ &= \beta^* \times (1 - p_A) \times p_I = \Pr(\text{Restate}), \end{aligned} \quad (5)$$

where β^* is the (researcher's) forecasted probability that a misstatement occurs, or, from equation (3),

$$\beta^* = \Pr ((1 - p_A)(1 - p_I)B - C_M \geq 0). \quad (6)$$

At this point, the theory has delivered a structure for relating the *unobserved* probability of a misstatement, β^* , to the potentially estimable probability of a restatement. Now, we face a familiar structural modeling problem, which is that the model does not anticipate all the reasons why, in practice, these probabilities might vary across firm accounting statements. For example, the theory so far does not point to reasons why CEOs might differ in their benefits and costs of misstatements. To move theoretical relations closer to the data, researchers typically allow parts of the model to depend on differentiating variables. Often the specifications of these dependencies are ad hoc. Empiricists are willing to do this, however, because they believe that it is important to account for practical aspects of the application that the theory does not recognize.

To illustrate this approach, and following suggestions of what might matter from the accounting research literature, suppose the CEO's unobserved costs and benefits vary as follows:

$$\begin{aligned} B &= b_0 + b_1 \text{EQUITY} + X_B\beta \\ C_M &= m_0 + m_1 \text{SALARY} + X_C\gamma + \xi, \end{aligned} \quad (7)$$

where EQUITY is the fraction of a CEO's total pay that is stock-based compensation, the X_B are other observable factors that impact the manager's benefits from misstatements, SALARY is the CEO's annual base salary, and the X_C are observable factors impacting the CEO's perceived costs of misstatements.³⁹ The EQUITY variable is intended to capture the idea that the more a CEO is rewarded for performance, the greater will be his or her incentive to misstate results so as to increase (perceived) performance. Thus, we would expect the unknown coefficient b_1 to be positive if providing more equity incentives increases the tendency of the CEO to misstate earnings, but expect $b_1 < 0$ if it reduces that tendency. Similarly, we include the variable SALARY as a driver of the cost of making misstatements, with the idea that a CEO caught misstating might lose his or her job, including salary (and other benefits). Thus, we would expect the unknown coefficient m_1 also to be positive. For now, we leave the other X variables unnamed.

We have no strong theoretical reason for the assumption of linearity. Its motivation is practical, as it facilitates estimation of the model unknowns (as we will shortly see).⁴⁰

With these assumptions, the probability of a restatement becomes

$$\Pr(\text{Restate}) = \theta_0 \Pr(\theta_1 + \theta_2 \text{EQUITY} + \theta_3 \text{SALARY} \geq \xi). \quad (8)$$

The new θ parameters are functions of the underlying incentive parameters as follows: $\theta_0 = (1 - p_A) \times p_I$, $\theta_1 = (1 - p_A)(1 - p_I)b_0 - m_0$, $\theta_2 = (1 - p_A)(1 - p_I)b_1$, and $\theta_3 = -m_1$. Apart from the scalar multiple θ_0 , which can be absorbed into the probability statement (and thus is not identified), this probability model has the form of a familiar binary choice model (e.g., a probit or logit). Thus, the value of the structure imposed so far is that it can motivate the application of a familiar statistical model as in Efendi, Srivastava, and Swanson [2007], as well as explain how the estimated coefficients are potentially connected to quantities that impact the probability of a misstatement.

5.2.2. Estimating the Nonstrategic Auditor Model. To illustrate how to estimate this structural model, we simulated a data set containing 10,000 firm-year observations on whether or not financial results were restated.⁴¹

For verisimilitude, we simulated variables that have been used to model restatements. RESTATE is a zero-one indicator variable for whether a firm

³⁹ For expositional purposes, we assume away X_B and X_C in our analysis.

⁴⁰ Another key variable in the above model is the unobserved cost ξ . While it makes sense to say that the researcher cannot measure all misstatement costs, why not also allow for unobserved benefits as well? The answer here is that adding an unobserved benefit would not really add to the model, as it is the net difference that the model is trying to capture. The sense in which it could matter is if we thought we observed the probabilities p_A and p_I . In this case, we might be able to distinguish between the cost and benefit unobservables based on their variances.

⁴¹ The parameter values used to generate the data are $a_0 = 0.5$, $a_1 = 3.5$, $a_2 = 3.5$, $m_0 = 7$, $m_1 = 1.5$, $b_0 = 20$, $b_1 = 10$, $p_0 = 0.75$, $v_0 = 0.05$, $p_I = 0.45$, and $r_0 = 60$. For those interested, the data are available at <http://web.stanford.edu/~preiss/Data.page.html>.

TABLE 1
Descriptive Statistics

Variable	Sample Mean (SE)
RESTATE	0.099 (0.30)
SALARY	1.06 (0.27)
EQUITY	0.45 (0.26)
BIG4	0.75 (0.43)
FINDIRECT	0.09 (0.08)
SEG	1.49 (0.50)
INT	0.31 (0.46)

RESTATE is a zero-one indicator for whether a sample firm made a restatement in a particular year. SALARY is the CEO's annual base salary (in millions of \$). EQUITY is the fraction of a CEO's total pay that is equity-based compensation. BIG4 is a zero-one indicator for whether the firm uses a Big 4 auditor. FINDIRECT is the fraction of the board of directors with a professional finance or accounting background. INT is a zero-one indicator for whether the firm derives most of its revenue outside the United States. SEG is the firm's number of two-digit SIC business segments.

restated (RESTATE = 1) their financial results in a given year. The variable BIG4 also is a zero-one indicator for whether the firm's auditor is one of the four largest U.S. accounting firms. It is included in the specifications because Big 4 auditing firms might have more accounting expertise and this expertise might make them more likely to catch misstatements. Similarly, the corporate governance literature suggests that board oversight from directors with accounting or finance backgrounds reduces the likelihood of misstatements. We proxy this possibility with FINDIREC, the percentage of directors who have professional accounting or finance backgrounds. Finally, the variables INT and SEG are included to capture the complexity and costs of audits. Specifically, INT is a zero-one indicator for whether the firm does a majority of its business outside the United States. We assume that international companies have higher auditing costs. Similarly, SEG is a count of the firm's business segments. We assume that more segments likely will increase the costs of auditing.

Table 1 reports descriptive statistics for our sample and table 2 reports the results of logit regressions in which the dependent variable is the restatement indicator variable. These specifications parallel prior descriptive statistical models that correlate restatements with other variables that might impact misstatements. The table contains both a simple specification containing an intercept along with the two CEO compensation variables, and a more intricate specification involving the other variables in the data set. For each specification, we report the estimated coefficients of the logit and the corresponding marginal effects evaluated at the sample means of the exogenous variables.

TABLE 2
Logit Regression Results

Coefficient	Specification 1		Specification 2	
	Coefficient (SE)	Marginal Effect (SE)	Coefficient (SE)	Marginal Effect (SE)
Intercept	-2.278 (0.141)		-3.498 (0.198)	
SALARY	0.280 (0.120)	0.025 (0.011)	0.326 (0.121)	0.028 (0.010)
EQUITY	-0.504 (0.130)	-0.045 (0.011)	-0.503 (0.131)	-0.043 (0.011)
BIG4			0.135 (0.080)	0.011 (0.006)
FINDIRECT			-0.239 (0.408)	-0.020 (0.034)
INT			0.548 (0.069)	0.051 (0.007)
SEG			0.578 (0.069)	0.049 (0.006)

This table presents results from logistic regressions of RESTATE, a zero-one indicator for whether the firm made a restatement in a particular year, on a proxy for managerial incentives and controls. The controls are as follows: SALARY is the CEO's annual base salary (in millions of \$), EQUITY is the fraction of a CEO's total pay that is equity-based compensation, BIG4 is a zero-one indicator for whether the firm uses a Big 4 auditor, FINDIRECT is the fraction of the board of directors with a professional finance or accounting background, INT is a zero-one indicator for whether the firm derives most of its revenue outside the United States, and SEG is the firm's number of business segments.

The results for the pay coefficients in both specifications run counter to those the previous accounting literature might predict and counter to those predicted by the structural model that assumes the benefit coefficient on equity pay, b_1 , is greater than zero. Specifically, more base pay is associated with more restatements, while more equity-based compensation is associated with fewer restatements.

Besides the intercepts and the EQUITY and SALARY coefficients, the only other coefficients that are statistically significant are those on INT and SEG. While we can say (descriptively) that INT and SEG are associated with higher restatement rates, unless we take a position on how they enter X_C or X_B , it is difficult to interpret whether these signs make sense.

The question we now address is what to make of the fact that the coefficients on EQUITY seem inconsistent with our informal arguments and with the prediction from our structural model that assumes $b_1 > 0$. One possible interpretation of this finding is that our beliefs about the effects of incentives on misstatements were wrong. Another possibility is that the measures we employ and the functional forms assumed are incorrect, which leads to spurious results. Yet another possibility is that our theory of misstatements is incorrect. It is this last possibility that we consider now.

5.2.3. Model 2: A Strategic Auditor Model. A key weakness of the previous model is that it ignores the incentives of the external auditor. According to PCAOB guidance in Auditing Standard No. 12, assessment of the risk of

material misstatement should take into account “incentive compensation arrangements.” Similarly, Auditing Standard No. 8 suggests that audit effort should increase if risk is higher. To make the model richer in a manner consistent with these institutional details, we assume that auditors trade off the costs of audit effort against the reputational losses they might incur should they miss a managerial misstatement that is subsequently detected.⁴²

In the previous model, the firm’s auditor impacted the manager’s misstatement benefits through p_A (which is assumed to be constant). Suppose that p_A is in fact a choice variable for the firm’s auditor. To make matters simple, suppose that the auditor detects manipulation with probability p_{AH} if they exert high effort and, otherwise, they detect manipulation with the lower probability p_{AL} . Let the cost of high effort be a fixed cost $C_A > 0$. Without loss of generality, suppose the cost of low effort is zero. When deciding whether to audit with high or low effort, the auditor perceives a cost to its reputation, C_R , because of not detecting a misstatement that is caught by subsequent investigations. This structure implies that the total cost of high effort to the auditor is $C_A + (1 - p_{AH}) \times p_I \times C_R$ or the cost of high effort plus the expected cost of missing a misstatement that is subsequently caught with probability p_I . The total expected cost of low effort is similarly equal to $(1 - p_{AL}) \times p_I \times C_R$.

To complete this new model, we need to make an (equilibrium) assumption about how the CEO and firm auditor interact. Following the literature, we assume that the two simultaneously and independently make decisions, and their strategies form a Nash equilibrium. That is, we assume the players’ strategies are such that they optimize their objectives taking the actions of the other players as fixed. This means that, in a Nash equilibrium, the players are taking actions that they cannot unilaterally improve upon.

In this type of auditing game, the Nash equilibrium has the CEO and the auditor playing mixed (randomized) strategies. That is, the auditor will independently exert high effort with probability α^* and the CEO independently misstates with probability β^* . These probabilities are such that each party has no incentive to change strategies. That is,

- 1) the CEO is indifferent between misstating and not misstating, or

$$(1 - p_A^*) (1 - p_I) B - C_M = 0, \quad (9)$$

where $p_A^* = \alpha^* p_{AH} + (1 - \alpha^*) p_{AL}$ is the equilibrium probability a misstatement is detected; and

- 2) the auditor is indifferent between exerting high and low effort, or

$$\beta^* (1 - p_{AH}) p_I C_R + C_A = \beta^* (1 - p_{AL}) p_I C_R.$$

⁴² Here we have in mind the findings by Dyck, Morse, and Zingales [2010], who show that many egregious forms of misstatements are detected subsequently by employees, directors, regulators, and the media.

Solving these two equations for the equilibrium probabilities α^* and β^* yields

$$\begin{aligned}\alpha^* &= \frac{(1 - p_{AL})(1 - p_I)B - C_M}{(1 - p_I)(p_{AH} - p_{AL})B}, \\ \beta^* &= \frac{C_A}{(p_{AH} - p_{AL})p_I C_R}.\end{aligned}\quad (10)$$

From these equations, we can calculate the equilibrium probability of a restatement⁴³

$$\begin{aligned}\Pr(\text{Restate}) &= \Pr(\text{Misstate}) \times \Pr(\text{Auditor Misses}) \times \Pr(\text{Investigation Finds}) \\ &= \beta^* \times (1 - p_A^*) \times p_I.\end{aligned}\quad (11)$$

This equation illustrates how the probability of a restatement is related to the unobserved frequency of misstatements. In particular, if we knew the frequency with which auditors and subsequent investigations caught misstatements, we could easily link the two. Otherwise, we would have to estimate these probabilities (or make assumptions about them).

Substituting the equilibrium strategies (10) into (11) yields

$$\Pr(\text{Restate}) = \frac{C_A C_M (1 - p_{AL})}{(p_{AH} - p_{AL})(1 - p_I) C_R B}.\quad (12)$$

Now we are in a position to use the theory to help interpret the conflicting logistic regression results in table 3.

Equation (12) shows that the presence of a strategic external auditor changes how the CEO's incentives impact the probability of a restatement.⁴⁴ Partial derivatives of equation (12) show that the restatement probability is:

- Decreasing in the benefit B that the CEO enjoys from misstatement;
- Increasing in the personal cost of manipulation C_M incurred by the CEO;
- Decreasing in the reputational cost C_R incurred by the external auditor;
- Increasing in the cost of high effort C_A incurred by the external auditor.

Thus, in contrast to the model with a nonstrategic auditor, increasing the benefit that managers enjoy from misstatement, or decreasing the misstatement cost, leads to fewer restatements being observed by researchers. These two effects might explain the negative sign on EQUITY and the positive sign on SALARY observed in the previous logit results.

⁴³ As part of the solution, we require α^* and β^* to be probabilities between 0 and 1. This is true, provided C_R and B satisfy the inequalities $C_R > \frac{C_A}{(p_{AH} - p_{AL})p_I}$ and $B > \frac{C_M}{1 - p_I}$.

⁴⁴ Note that the probability statement in equation (12) differs from the statement in equation (8). The probability statement in equation (12) reflects the randomness of the strategies, whereas in equation (8) it reflects variables the researcher does not observe.

T A B L E 3
Logit Generalized Method of Moments (GMM) Estimates for the Strategic Auditor Model

	Estimated Coefficient (SE)	Estimated Coefficient (SE)	Estimated Coefficient (SE)	True Coefficient
$\theta_0 = \frac{(1-v_0)a_0m_0}{(1-p_I)(p_0-v_0)r_0b_0}$	-0.016 (0.045)	-0.009 (0.044)	0.008 (0.002)	0.007
$\theta_1 = \frac{(1-v_0)a_1m_0}{(1-p_I)(p_0-v_0)r_0b_0}$	0.049 (0.034)	0.055 (0.035)	0.053 (0.007)	0.050
$\theta_2 = \frac{(1-v_0)a_2m_0}{(1-p_I)(p_0-v_0)r_0b_0}$	0.060 (0.030)	0.055 (0.030)	0.051 (0.006)	0.050
$\theta_3 = \frac{(1-v_0)a_0m_1}{(1-p_I)(p_0-v_0)r_0b_0}$	0.025 (0.042)	0.018 (0.042)	0.002 (0.001)	0.002
$\theta_4 = \frac{(1-v_0)a_1m_1}{(1-p_I)(p_0-v_0)r_0b_0}$	0.014 (0.032)	0.010 (0.032)	0.012 (0.003)	0.011
$\theta_5 = \frac{(1-v_0)a_2m_1}{(1-p_I)(p_0-v_0)r_0b_0}$	0.003 (0.028)	0.008 (0.028)	0.012 (0.001)	0.011
$\theta_6 = \frac{b_1}{b_0}$	0.570 (0.184)	0.566 (0.184)	0.576 (0.261)	0.500

This table presents GMM estimates for the strategic auditor model of section 5. The first (exactly identified) specification (1) uses the instruments Constant, SALARY, EQUITY, INT, SEG, SALARY \times SEG, and SALARY \times INT. Specification (2) is overidentified. It uses the same instruments as specification (1) plus EQUITY \times SEG, EQUITY \times INT, SALARY², and EQUITY². Specification (3) uses the instruments in specification (1) and imposes the constraints $\theta_0/\theta_3 = \theta_1/\theta_4 = \theta_2/\theta_5$, $\theta_0/\theta_1 = \theta_3/\theta_4$, and $\theta_1/\theta_3 = \theta_1/\theta_2$. The "True" values are those corresponding to the parameters used to generate the data.

Robust standard errors are reported for specifications (1) and (2). The standard errors for specification (3) are based on 500 Monte Carlo replications. The parameter values used to generate the data are $a_0 = 0.5$, $a_1 = 3.5$, $a_2 = 3.5$, $m_0 = 7$, $m_1 = 1.5$, $b_0 = 20$, $b_1 = 10$, $p_0 = 0.75$, $v_0 = 0.05$, $p_I = 0.45$, and $r_0 = 60$.

To have a better sense of how one might connect the strategic auditor theory to the logistic models in table 2, suppose, similar to the motivation for equation (7), that

$$\begin{aligned}
 B &= b_0 + b_1 \text{EQUITY}, \\
 C_M &= m_0 + m_1 \text{SALARY}, \\
 C_A &= a_0 + a_1 \text{INT} + a_2 \text{SEG}, \\
 C_R &= r_0, \quad p_{AH} = p_0, \quad \text{and} \quad p_{AL} = v_0,
 \end{aligned} \tag{13}$$

where a_0 , a_1 , a_2 , r_0 , b_0 , p_0 , and v_0 are constant parameters. Inserting these expressions into the expected restatement rate (12) gives

$$\begin{aligned}
 \Pr(\text{Restate}) &= \frac{C_A C_M (1 - p_{AL})}{(p_{AH} - p_{AL})(1 - p_I) C_R B} \\
 &= \frac{(1 - v_0)(a_0 + a_1 \text{INT} + a_2 \text{SEG})(m_0 + m_1 \text{SALARY})}{(p_0 - v_0)r_0(b_0 + b_1 \text{EQUITY})} \tag{14}
 \end{aligned}$$

$$= \frac{\theta_0 + \theta_1 \text{INT} + \theta_2 \text{SEG} + \theta_3 \text{SALARY} + \theta_4 \text{INT} \times \text{SALARY} + \theta_5 \text{SEG} \times \text{SALARY}}{1 + \theta_6 \text{EQUITY}}. \tag{15}$$

Note that the θ s absorb unknown quantities such as r_0 and p_0 , and the denominator intercept is normalized to 1. This last restriction is required to identify the ratio of the two linear functions.

Although this model does not have a logit form, it is potentially estimable using a nonlinear estimation method such as the generalized method of moments (GMM).⁴⁵ GMM attempts to match sample moments to what the structural model implies these moments should be. For example, an obvious sample moment would be the average restatement rate in the sample. The corresponding theoretical moment would be the probability expression in equation (15). In our estimations, we use sample moments of the form

$$\mathcal{M}_j = \sum_{i=1}^{10,000} X'_{ji} [\text{RESTATE}_i - \text{Pr}(\text{Restate}_i)], \quad (16)$$

where $\text{Pr}(\text{Restate})$ comes from equation (15).⁴⁶ Practically, we need at least as many moments as we have θ parameters to estimate (there are seven θ s in the model). The X_j used in the moments include all explanatory variables, plus some interactions (see table 3 for a list). Again, to illustrate how we estimate the θ parameters of equation (15), one of the X s is a dummy variable for whether the firm is an international company. The corresponding moment equation in (16) seeks to match international companies' average restatement rate to the model's prediction for that rate.

Table 3 reports the results of estimating the new (strategic auditor) structural model. The results show that, in this particular case, even without sample information on the unobserved probabilities p_A and p_I , we can recover estimates of the model parameters up to normalizations. For instance, the coefficient ratio θ_4/θ_1 estimates the ratio of cost parameters m_1/m_0 . The parameter m_1 is the cost coefficient on SALARY and $m_0 > 0$ is a fixed cost of manipulation. The sign of θ_4/θ_1 thus reveals the sign of m_1 . From the theory, we expect the sign to be positive, and this is what we find in the estimation results.⁴⁷

Similarly, θ_6 equals the (scaled) misstatement benefit coefficient on the EQUITY variable. Recall that the descriptive logit regression coefficients in table 2 suggest that EQUITY has a negative effect on misstatements. In contrast, we now find the expected positive relation because we explicitly model the difference between misstatements and restatements in our structural estimation.

Table 3 contains three different sets of estimates. Column (1) is for an exactly identified model, where there are as many instruments as

⁴⁵ Other estimation strategies are possible, but we do not consider them here.

⁴⁶ To ensure that the model parameters imply restatement probabilities between 0 and 1, we add a penalty function to the GMM objective function. This penalty increases with the number of estimated probabilities below 0 or above 1. For most replications, this penalty is immaterial to the results obtained.

⁴⁷ The same is not true for θ_3/θ_0 .

parameters to estimate. Column (2) presents estimates for an overidentified model in which there are more moments than parameters to estimate. Column (3) presents constrained GMM estimates, where the constraints are motivated by the fact that pairs of ratios of the θ coefficients are equal (e.g., $\theta_3/\theta_0 = \theta_4/\theta_1$). The three sets of estimates are similar, with the constrained estimates yielding more precision.

5.2.4. Implications of Structural Modeling Analyses. From the discussion above, it seems that there are (at least) two alternative explanations (or hypotheses) for the results we find. One hypothesis is that the process generating the data is best modeled with a nonstrategic auditor and the effect of EQUITY on incentives to misstate is either negative (or perhaps zero). The support for this hypothesis comes from table 2, which is an appropriate regression analysis for the model with a nonstrategic auditor, where a negative (and weakly statistically significant) coefficient on EQUITY is found. However, a second, and in our view a considerably more plausible, hypothesis is that the process generating the data is best modeled with a strategic auditor and that the effect of EQUITY on incentives to misstate is positive (or perhaps zero). The support for this hypothesis comes from table 3, which is predicated on the model with a strategic auditor, and where a positive coefficient on b_1 (the parameter linking EQUITY to benefits from misstatement) is found.

The point of this discussion is not to resolve the debate regarding the effect of incentives on misstatements. Rather, the goal is to illustrate the necessity of having an underlying structural model of the process by which the data we observe were generated. The importance of such models was illustrated in sections 2 and 4, where we used causal diagrams as a kind of (nonparametric) causal model. Here we have shown that more can be inferred from a formal model tied to behavioral assumptions.

Not only does a structural model enable us to derive sharper predictions regarding the relations between variables for various parameterizations, but it also provides a basis for actually estimating those relations. In particular, the comparative statistics of the model shed light on the difference between restatements and misstatements, and what assumptions (e.g., a strategic or a nonstrategic auditor) and data were needed to draw inferences about misstatements from restatements. Additionally, we were able to recover some of the primitive parameters impacting incentives for managers to misstate results, as well as perform counterfactual analyses. Finally, although structural modeling does not allow us to completely resolve questions of causality, if the model is based on reasonable assumptions and has a close fit to the data, we arguably have better insight into the likely causal relations underlying the phenomenon being examined.

5.2.5. Counterfactuals. While the coefficient magnitudes only allow us to infer relative coefficient magnitudes and signs, we nevertheless can use the model to perform counterfactual calculations. A counterfactual calculation determines what is the consequence of changing some parameters or

variables while holding others fixed. There are many different counterfactuals that could be considered. For illustrative purposes, we evaluate what would happen to misstatements and restatements if we do away with equity-based compensation and nothing else changes in the model. The value of having an equilibrium model to analyze this change is that we explicitly allow the auditing process to adjust to the removal of CEO incentives to misstate. From the equilibrium strategies in equation (10), we observe that removing equity-based pay ($b_1 = 0$ or EQUITY=0) does not change the equilibrium frequency of misstatements (i.e., β^*), but does change the frequency of high-effort auditing (α^*). Inserting the estimated parameters into equation (14), we find

$$\frac{\text{Pr}(\text{Restate} \mid \text{No Equity})}{\text{Pr}(\text{Restate} \mid \text{Equity})} = \frac{\beta^* \times (1 - p_A^{**}) \times p_I}{\beta^* \times (1 - p_A^*) \times p_I} = \frac{(1 - p_A^{**})}{(1 - p_A^*)} = 1.24.$$

This result illustrates that the restatement rate would increase by 24% (or from 9.99% to 12.36%) if equity-based incentives were removed or did not impact the benefits of misstatements. The fact that the restatement rate goes up may at first seem somewhat odd, given that the benefits to the CEOs have fallen. The model, however, shows that the increase comes about because the auditors exert *less* effort in detecting misstatements, thereby catching fewer, leaving more for subsequent investigations to detect.

5.3 LIMITATIONS OF STRUCTURAL MODELS

When we discuss structural models with colleagues, we sometimes encounter two negative reactions. One objection appears to be that these models are both too complicated and too unrealistic to yield much insight into practical accounting research questions. The other is related to a preconception that structural modelers believe that all empirical work needs to be structural.

To be clear, we do not believe that all accounting researchers need estimate structural models. Indeed, in many situations, descriptive models can be more informative. Further, no structural modeling exercise should go forward, unless the researcher is convinced that the benefits of a structural model would outweigh the substantial costs entailed in developing and estimating a structural model.

We have mentioned some of the benefits of structural models, particularly when it comes to having a clear basis for making causal statements. What are the costs to developing and estimating structural models? First, structural models can be technically demanding to develop. Additionally, while constructing a theoretical model that can be derived from the data, the empirical researcher is typically forced to make simplifications that a pure theorist might never make and other empiricists would criticize as unrealistic. Further, as we saw in the restatement model, the structural modeler often is in the uncomfortable position of adding features *ex post*, such as covariates, to the model that make it more in line with the realities of the application.

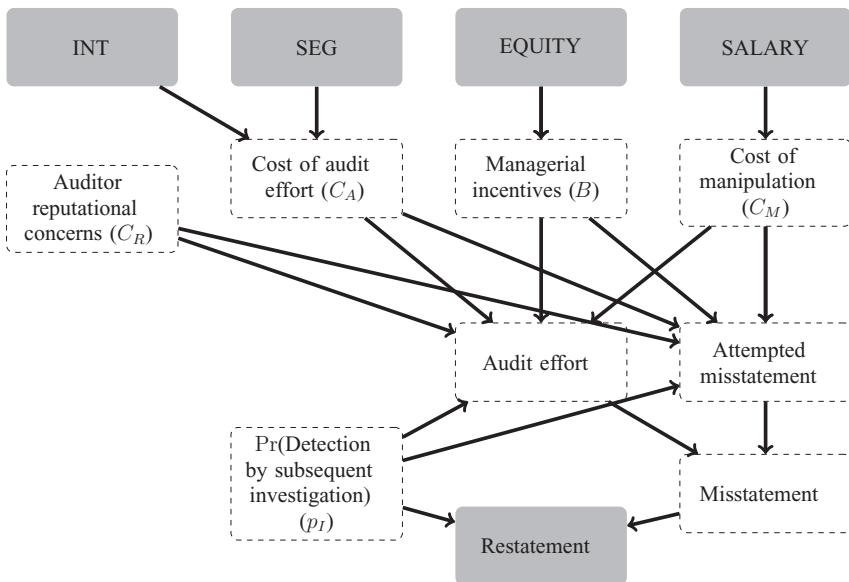


FIG. 3.—Causal diagram for the strategic auditor model.

While these costs are important to recognize, it is also important to realize that, without structural models, there likely will continue to be a substantial divide between theoretical and empirical research in accounting. With a few exceptions, theoretical accounting researchers do not explain how to map the specifics of their models to data. In many cases, extant theory is not sufficient to motivate the hypotheses tested by empirical researchers. Consistent with the existence of this gap, few empirical research papers in accounting rely on formal theoretical models to motivate their hypotheses. Often, when empirical researchers *do* rely on theoretical papers to motivate hypotheses, the predictions claimed to be derived from those papers have little obvious connection with the actual content of those papers. Instead, almost all empirical research papers in accounting use more informal, verbal approaches to hypothesis development.

Second, although structural models can, in principle, make it clear what a researcher is assuming about causality, it is incumbent on the structural modeler to make the model's causal relations clear. One way to do this is to provide a causal diagram. To illustrate, figure 3 provides a causal diagram representing the strategic auditor restatement model.⁴⁸ As can be seen, we are assuming that p_I is independent of EQUITY. But, it is quite plausible that these investigations are conducted (in part) by a regulator who is as

⁴⁸ While the mixed strategy of our model has β not being a function of B or C_M and α not being a function of C_A or C_R , we have retained these links as being plausible in a more general model.

strategic as the auditor, thus giving rise to a link between EQUITY and p_I . We also assumed that EQUITY is exogenous, whereas it is plausibly related to the complexity of the business, which may also affect the cost of auditing. These links could be added to the structural model, albeit at some cost. Evaluation of their accuracy then could be made using traditional in-sample and out-of-sample goodness-of-fit tests.

Third, just because a researcher can write down a theoretical model and estimate it does not make the empirical model “right.” Clearly, there is a risk of incorrect causal inferences being drawn from estimation of a structural model based on faulty assumptions. But, structural models are capable of recovering behavioral information, provided the model is correct, which can be seen in table 3 by comparing the “true” coefficients in the last column (i.e., those used to generate the data) with those in the other columns. Because we also used the same data to estimate the logit regression in table 2, we are able to establish that if we used the wrong (nonstrategic) auditor structural model, we would have been led astray in our inferences about the effect of EQUITY.

In practice, we do not have this kind of insight into the correct process that generated the data. Hence, the researcher will need to weigh models based upon how well the model’s assumptions match the practical and institutional realities of the phenomena being studied. Despite these challenges, we believe that there is significant value in making the theory underlying empirical research transparent and rigorous.⁴⁹

6. Concluding Remarks

In this paper, we examined the various approaches used by accounting researchers to draw causal inferences from analyses of observational (or nonexperimental) data. The vast majority of empirical papers using such data seek to draw causal inferences, notwithstanding the well-known difficulties in doing so. While some papers seek to use quasi-experimental methods to develop unbiased estimates of causal effects, the assumptions required to deliver such estimates are not often credible. We believe that clearer communication of research questions and design choices would help researchers avoid some of the conceptual traps that affect accounting research. One tool that may help in this regard are causal diagrams because these require the research to be very transparent about the causal mechanism that is being assumed for the research question.

⁴⁹ The use of structural models in accounting research has been fairly limited to date. Recent examples include Gerakos and Kovrijnykh [2013], Gerakos and Syverson [2015], Zakharyukina [2015], and Bertomeu, Ma, and Marinovic [2015]. These four papers model an institutionally rich problem, estimate the derived model, provide estimates for important structural parameters, and also give interesting counterfactuals based on their theoretical models. We view these papers as useful initial steps in applying structural approaches to accounting research questions.

We also argued that accounting research could benefit from a more complete understanding of causal pathways. In particular, we believe that structural models based on rigorous theory will see greater use in the coming years. Finally, we see great value to in-depth descriptive studies that inform causal issues and deepen our knowledge of the behavior and institutions we seek to model. Although our suggestions do not completely resolve controversies surrounding causal inferences drawn from observational data, we believe that they offer a viable and exciting path forward.

APPENDIX

Causal Diagrams: Formalities

In this appendix, we provide a more formal treatment of some of the ideas on causal diagrams discussed in the text (see Pearl [2009b] for more detailed coverage).

A.1 DEFINITIONS AND A RESULT

We first introduce some basic definitions and a key result.

DEFINITION 1. (*d*-separation, block, collider). *A path p is said to be d-separated (or blocked) by a set of nodes Z if and only if*

- 1) *p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node m is in Z, or*
- 2) *p contains an inverted fork (or collider) $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and such that no descendant of m is in Z.*

DEFINITION 2. (Back-door criterion). *A set of variables Z satisfies the back-door criterion relative to an ordered pair of variables (X, Y) in a directed acyclic graph (DAG) G if*

- *no node in Z is a descendant of X, and*
- *Z blocks every path between X and Y that contains an arrow into X.*⁵⁰

Given this criterion, Pearl [2009b, p. 79] proves the following result.

THEOREM 1. (Back-door adjustment). *If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula*

$$P(y|x) = \sum_z P(y|x, z)P(z),$$

⁵⁰The “arrow into X” is the portion of the definition that explains the “back-door” terminology.

where $P(y|x)$ stands for the probability that $Y = y$, given that X is set to level $X = x$ by external intervention.⁵¹

A.2 APPLICATION OF BACK-DOOR CRITERION TO FIGURE 1

Applying the back-door criterion to figure 1(A) is straightforward and intuitive. The set of variables $\{Z\}$ or simply Z satisfies the criterion, as Z is not a descendant of X and Z blocks the back-door path $X \leftarrow Z \rightarrow Y$. So, by conditioning on Z , we can estimate the causal effect of X on Y . This situation is a generalization of a linear model in which $Y = X\beta + Z\gamma + \epsilon_Y$ and ϵ_Y is independent of X and Z , but X and Z are correlated. In this case, it is well known that omission of Z would result in a biased estimate of β , the causal effect of X on Y , but, by including Z in the regression, we get an unbiased estimate of β . In this situation, Z is a *confounder*.

In figure 1(B), we see that Z , which is a *mediator* of the effect of X on Y , does not satisfy the back-door criterion, because Z is a descendant of X . However, \emptyset (i.e., the empty set) does satisfy the back-door criterion. Clearly, \emptyset contains no descendant of X . Furthermore, the only path other than $X \rightarrow Y$ that exists is $X \rightarrow Z \rightarrow Y$, which does not have a back door into X . Note that the back-door criterion not only implies that we need not condition on Z to obtain an unbiased estimate of the causal effect of X on Y , but we should not condition on Z to get such an estimate.

Finally, in figure 1(C), we have Z acting as what Pearl [2009a, p. 17] refers to as a “collider” variable.⁵² Again, we see that Z does not satisfy the back-door criterion, because Z is a descendant of X . However, \emptyset again satisfies the back-door criterion. First, it contains no descendant of X . Second, the only path other than $X \rightarrow Y$ that exists is $X \rightarrow Z \leftarrow Y$, which does not have a back door into X . Again, the back-door criterion not only implies that we need not condition on Z , but we should not condition on Z to get an unbiased estimate of the causal effect of X on Y .

A.3 CAUSAL DIAGRAMS AND INSTRUMENTAL VARIABLES

We now discuss how *correct* causal diagrams can be used to identify valid (or invalid) instruments.

DEFINITION 3. (Instrument). Let G denote a causal graph in which X has an effect on Y . Let $G_{\bar{X}}$ denote the causal graph created by deleting all arrows emanating from X . A variable Z is an instrument relative to the total effect of X on Y if there exists a set of nodes S , unaffected by X , such that

- 1) S d-separates Z from Y in $G_{\bar{X}}$, and
- 2) S does not d-separate Z from X in G .

⁵¹ How the quantities $P(y|x)$ map into estimates of causal effects is not critical to the current discussion; it suffices to note that, in a given setting, it can be calculated if the needed variables are observable.

⁵² The two arrows from X and Y “collide” in Z .

Applying this definition to figure 2, we can evaluate the instrument used in Armstrong, Gow, and Larcker [2013]. There we have $S = Comp_{t-1}$, $X = Shareholder\ support_t$, $Y = Comp_{t+1}$, and $Z = ISS\ recommendation_t$. We use U to denote the observed variables depicted in the dashed box of figure 2(A). If $G_{\bar{X}}$ is created by deleting the single arrow emanating from $Shareholder\ support_t$, we can see that there are two back-door paths running from Y to Z : $Z \leftarrow S \rightarrow U \rightarrow Y$ and $Z \leftarrow S \rightarrow Y$. However, both these paths are blocked by S and the first requirement is satisfied. The second requirement is clearly satisfied as Z is directly linked to X .⁵³ Note that this analysis can be expressed intuitively as requiring that the ISS recommendation only affects $Comp_{t+1}$ through its effect on $Shareholder\ support_t$, and the ISS recommendation has an effect on $Shareholder\ support_t$.

But this analysis presumes that the causal diagram figure 2(A) is correct. Armstrong, Gow, and Larcker [2013, p. 912] note that the “validity of this instrument depends on ISS recommendations not having an influence on future compensation decisions conditional on shareholder support (i.e., firms listen to their shareholders, with ISS having only an indirect impact on corporate policies through its influence on shareholders’ voting decisions).” This assumption is represented in figure 2(A) by the *absence* of an arrow from $ISS\ recommendation_t$ to $Comp_{t+1}$.

Unfortunately, this assumption seems inconsistent with the findings of Gow et al. [2013], who provide evidence that firms are carefully calibrating compensation plans (i.e., factors that directly affect $Comp_{t+1}$) to comply with the requirements of ISS’s policies, implying a path from $ISS\ recommendation_t$ to $Comp_{t+1}$ that does not pass through $Shareholder\ support_t$. This path is represented in figure 2(B) and the plausible existence of this path suggests that the instrument of Armstrong, Gow, and Larcker [2013, p. 912] is not credibly valid for the causal effect they seek to estimate.

REFERENCES

- AIER, J. K.; L. CHEN; AND M. PEVZNER. “Debtholders’ Demand for Conservatism: Evidence from Changes in Directors’ Fi(Fiduciary Duties.” *Journal of Accounting Research* 52 (2014): 993–1027.
- ANGRIST, J. D. “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.” *American Economic Review* 80 (1990): 313–36.
- ANGRIST, J. D.; G. W. IMBENS; AND D. B. RUBIN. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91 (1996): 444–55.
- ANGRIST, J. D., AND J.-S. PISCHKE. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press, 2008.
- ANGRIST, J. D., AND J.-S. PISCHKE. “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics.” *Journal of Economic Perspectives* 24 (2010): 3–30.
- ARMSTRONG, C. S.; I. D. GOW; AND D. F. LARCKER. “The Efficacy of Shareholder Voting: Evidence from Equity Compensation Plans.” *Journal of Accounting Research* 51 (2013): 909–50.

⁵³This is a necessary condition, but assumptions about functional form are also critical in using an instrument to estimate a causal effect. However, this is not essential to our argument here.

- ARMSTRONG, C. S.; A. D. JAGOLINER; AND D. F. LARCKER. "Chief Executive Officer Equity Incentives and Accounting Irregularities." *Journal of Accounting Research* 48 (2010): 225–71.
- BALAKRISHNAN, K.; M. B. BILLINGS; B. KELLY; AND A. LJUNGQVIST. "Shaping Liquidity: On the Causal Effects of Voluntary Disclosure." *The Journal of Finance* 69 (2014): 2237–78.
- BEBCHUK, L.; A. COHEN; AND A. FERRELL. "What Matters in Corporate Governance?" *Review of Financial Studies* 22 (2009): 783–827.
- BERK, R. A. "Review of *Observational Studies* by Paul Rosenbaum." *Journal of Educational and Behavioral Statistics* 24 (1999): 95–100.
- BERLE, A. A., AND G. C. MEANS. *The Modern Corporation and Property*. New York: Macmillan, 1932.
- BERTOMEU, J.; P. MA; AND I. MARINOVIC. "How Often Do Managers Withhold Information: A Structural Estimation." Unpublished working paper, 2015.
- BEYER, A.; D. A. COHEN; T. Z. LYS; AND B. R. WALTHER. "The Financial Reporting Environment: Review of the Recent Literature." *Journal of Accounting and Economics* 50 (2010): 296–343.
- BLOOMFIELD, R. J.; M. W. NELSON; AND E. F. SOLTES. "Gathering Data for Archival, Field, Survey and Experimental Accounting Research." *Journal of Accounting Research* (2016): this issue.
- BOWER, J. L. *Managing the Resource Allocation Process*. Boston, MA: Harvard Business School Press, 1986.
- BRANDEIS, L. "Breaking the Money Trusts." *Harper's Weekly*, December 6, 1913.
- BROWN, L. D.; A. C. CALL; M. B. CLEMENT; AND N. Y. SHARP. "Inside the 'Black Box' of Sell-Side Financial Analysts." *Journal of Accounting Research* 53 (2015): 1–47.
- BROWN, N. C.; H. STICE; AND R. M. WHITE. "Mobile Communication and Local Information Flow: Evidence from Distracted Driving Laws." *Journal of Accounting Research* 53 (2015): 275–329.
- BURNS, N., AND S. KEDIA. "The Impact of Performance-Based Compensation on Misreporting." *Journal of Financial Economics* 79 (2006): 35–67.
- CANNON, J. N. "Determinants of 'Sticky Costs': An Analysis of Cost Behavior Using United States Air Transportation Industry Data." *The Accounting Review* 89 (2014): 1645–72.
- CLOR-PROELL, S. M., AND L. A. MAINES. "The Impact of Recognition Versus Disclosure on Financial Information: A Preparer's Perspective." *Journal of Accounting Research* 52 (2014): 671–701.
- COHEN, J. R.; U. HOITASH; G. KRISHNAMOORTHY; AND A. M. WRIGHT. "The Effect of Audit Committee Industry Expertise on Monitoring the Financial Reporting Process." *The Accounting Review* 89 (2014): 243–73.
- CORREIA, M. M. "Political Connections and SEC Enforcement." *Journal of Accounting and Economics* 57 (2014): 241–62.
- CREADY, W. H.; A. KUMAS; AND M. SUBASI. "Are Trade Size-Based Inferences About Traders Reliable? Evidence from Institutional Earnings-Related Trading." 52 (2014): 877–909.
- CYERT, R. M.; H. A. SIMON; AND D. B. TROW. "Observation of a Business Decision." *The Journal of Business* 29 (1956): 237–48.
- CZERNY, K.; J. J. SCHMIDT; AND A. M. THOMPSON. "Does Auditor Explanatory Language in Unqualified Audit Reports Indicate Increased Financial Misstatement Risk?" *The Accounting Review* 89 (2014): 2115–49.
- DAINES, R., AND M. KLAUSNER. "Do IPO Charters Maximize Firm Value? Antitakeover Protection in IPOs." *Journal of Law, Economics, and Organization* 17 (2001): 83–120.
- DE FRANCO, G.; F. P. VASVARI; D. VYAS; AND R. WITTENBERG-MOERMAN. "Debt Analysts' Views of Debt-Equity Conflicts of Interest." *The Accounting Review* 89 (2014): 571–604.
- DICHEV, I. D., AND D. J. SKINNER. "Large-Sample Evidence on the Debt Covenant Hypothesis." *Journal of Accounting Research* 40 (2002): 1091–123.
- DOLL, R., AND R. PETO. "Mortality in Relation to Smoking: 20 Years' Observations on Male British Doctors." *British Medical Journal* 2 (1976): 1525–36.
- DUNNING, T. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge, UK: Cambridge University Press, 2012.
- DYCK, A.; A. MORSE; AND L. ZINGALES. "Who Blows the Whistle on Corporate Fraud?" *The Journal of Finance* 65 (2010): 2213–53.

- EFENDI, J.; A. SRIVASTAVA; AND E. P. SWANSON. "Why Do Corporate Managers Misstate Financial Statements? The Role of Option Compensation and Other Factors." *Journal of Financial Economics* 85 (2007): 667–708.
- ERKENS, D. H.; K. R. SUBRAMANYAM; AND J. ZHANG. "Affiliated Banker on Board and Conservative Accounting." *The Accounting Review* 89 (2014): 1703–28.
- FISHER, R. A. *The Design of Experiments*. Oliver and Boyd, Edinburgh, UK 1935.
- FLOYD, E., AND J. A. LIST. "Using Field Experiments in Accounting and Finance." *Journal of Accounting Research* (2016): this issue.
- FOX, W. F.; L. LUNA; AND G. SCHAUR. "Destination Taxation and Evasion: Evidence from U.S. Inter-State Commodity Flows." *Journal of Accounting and Economics* 57 (2014): 43–57.
- FREEDMAN, D. A. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge, UK: Cambridge University Press, 2009.
- FRYDMAN, C., AND R. E. SAKS. "Executive Compensation: A New View from a Long-Term Perspective, 1936–2005." *Review of Financial Studies* 23 (2010): 2099–138.
- GERAKOS, J., AND A. KOVRIJNYKH. "Performance Shocks and Misreporting." *Journal of Accounting and Economics* 56 (2013): 57–72.
- GERAKOS, J., AND C. SYVERSON. "Competition in the Audit Market: Policy Implications." *Journal of Accounting Research* (2015): Forthcoming.
- GILLIES, D. "The Russo–Williamson Thesis and the Question of Whether Smoking Causes Heart Disease," in *Causality in the Sciences*, edited by P. M. Illari, F. Russo, and J. Williamson. Oxford University Press, Oxford, UK, 2011: 110–25.
- GLYMOUR, M. M., AND S. GREENLAND. "Causal Diagrams," in *Modern Epidemiology*, Third edition, Chapter 12, edited by K. J. Rothman, S. Greenland, and T. L. Lash. Philadelphia, PA: Lippincott Williams & Wilkins, 2008: 183–209.
- GOLDBERGER, A. S. "Structural Equation Methods in the Social Sciences." *Econometrica* 40 (1972): 979–1001.
- GOMPERS, P.; J. ISHII; AND A. METRICK. "Corporate Governance and Equity Prices." *The Quarterly Journal of Economics* 118 (2003): 107–56.
- GOW, I. D.; D. F. LARCKER; A. MCCALL; AND B. TAYAN. *Sneak Preview: How ISS Dictates Equity Plan Design*. Working paper, Stanford Graduate School of Business, 2013.
- GROYSBERG, B.; P. M. HEALY; AND D. A. MABER. "What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?" *Journal of Accounting Research* 49 (2011): 969–1000.
- GUEDHAMI, O.; J. A. PITTMAN; AND W. SAFFAR. "Auditor Choice in Politically Connected Firms." *Journal of Accounting Research* 52 (2014): 107–62.
- HAAVELMO, T. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11 (1943): 1–12.
- HAAVELMO, T. "The Probability Approach in Econometrics." *Econometrica* 12 (1944).
- HAIL, L.; A. TAHOUN; AND C. WANG. "Dividend Payouts and Information Shocks." *Journal of Accounting Research* 52 (2014): 403–56.
- HEALY, P. M. "The Effect of Bonus Schemes on Accounting Decisions." 7 (1985): 85–107.
- HEALY, P. M.; A. P. HUTTON; AND K. PALEPU. "Stock Performance and Intermediation Changes Surrounding Sustained Increases in Disclosure." *Contemporary Accounting Research* 16 (1999): 485–520.
- HECKMAN, J., AND R. PINTO. "Causal Analysis After Haavelmo." *Econometric Theory* 31 (2015): 115–51.
- HOLLAND, P. W. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (1986): 945–60.
- HOUSTON, J. F.; L. JIANG; C. LIN; AND Y. MA. "Political Connections and the Cost of Bank Loans." *Journal of Accounting Research* 52 (2014): 193–243.
- ILIEV, P. "The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices." *The Journal of Finance* 65 (2010): 1163–96.
- IMBENS, G. W., AND K. KALYANARAMAN. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (2012): 933–59.
- IMBENS, G. W., AND D. B. RUBIN. *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge, UK: Cambridge University Press, UK, 2015.

- KALAY, A. "Stockholder-Bondholder Conflict and Dividend Constraints." *Journal of Financial Economics* 10 (1982): 211–33.
- KELLY, B., AND A. LJUNGQVIST. "Testing Asymmetric-Information Asset Pricing Models." *Review of Financial Studies* 25 (2012): 1366–413.
- KIM, K.; E. G. MAULDIN; AND S. PATRO. "Outside Directors and Board Advising and Monitoring Performance." *Journal of Accounting and Economics* 57 (2014): 110–31.
- KIRK, M. P., AND J. D. VINCENT. "Professional Investor Relations Within the Firm." *The Accounting Review* 89 (2014): 1421–52.
- LANG, M., AND R. LUNDHOLM. "Cross-Sectional Determinants of Analyst Ratings of Corporate Disclosures." *Journal of Accounting Research* 31 (1993): 246–71.
- LANG, M. H., AND R. J. LUNDHOLM. "Corporate Disclosure Policy and Analyst Behavior." *The Accounting Review* 71 (1996): 467–92.
- LARCKER, D. F.; G. ORMAZBAL; AND D. J. TAYLOR. "The Market Reaction to Corporate Governance Regulation." *Journal of Financial Economics* 101 (2011): 431–48.
- LARCKER, D. F.; S. A. RICHARDSON; AND I. TUNA. "Corporate Governance, Accounting Outcomes, and Organizational Performance." *The Accounting Review* 82 (2007): 963–1008.
- LARCKER, D. F., AND T. O. RUSTICUS. "On the Use of Instrumental Variables in Accounting Research." *Journal of Accounting and Economics* 49 (2010): 186–205.
- LEE, D. S., AND T. LEMIEUX. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2010): 281–355.
- LENNOX, C. S.; J. R. FRANCIS; AND Z. WANG. "Selection Models in Accounting Research." *The Accounting Review* 87 (2012): 589–616.
- LEUZ, C., AND P. WYSOCKI. "The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research." *Journal of Accounting Research* (2016): this issue.
- LEWELLYN, W. *Executive Compensation in Large Industrial Corporations*. New York: Columbia University Press, 1968.
- LI, Y., AND L. ZHANG. "Short Selling Pressure, Stock Price Behavior, and Management Forecast Precision: Evidence from a Natural Experiment." *Journal of Accounting Research* 53 (2015): 79–117.
- LISTOKIN, Y. "Management Always Wins the Close Ones." *American Law and Economics Review* 10 (2008): 159–84.
- LO, A. K. "Do Declines in Bank Health Affect Borrowers' Voluntary Disclosures? Evidence from International Propagation of Banking Shocks." *Journal of Accounting Research* 52 (2014): 541–81.
- MCCRARY, J. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2008): 698–714.
- MICHELS, J. "Disclosure Versus Recognition: Inferences from Subsequent Events." Working paper, University of Pennsylvania, 2015.
- MILLER, G. S., AND D. J. SKINNER. "The Evolving Disclosure Landscape: How Changes in Technology, the Media, and Capital Markets Are Affecting Disclosure." *Journal of Accounting Research* 53 (2015): 221–39.
- MINTZBERG, H. *The Nature of Managerial Work*. New York: Harper & Row, 1973.
- MINUTTI-MEZA, M. "Issues in Examining the Effect of Auditor Litigation on Audit Fees." *Journal of Accounting Research* 52 (2014): 341–56.
- MORROW, J. D.; B. FREI; A. W. LONGMIRE; J. M. GAZIANO; S. M. LYNCH; Y. SHYR; W. E. STRAUSS; J. A. OATES; AND L. J. ROBERTS. "Increase in Circulating Products of Lipid Peroxidation F_2 -Isoprostanes in Smokers: Smoking as a Cause of Oxidative Damage." *New England Journal of Medicine* 332 (1995): 1198–203.
- NEYMAN, J. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science* 5 (1923): 465–72.
- PEARL, J. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3 (2009a): 96–146.
- PEARL, J. *Causality: Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press, 2009b.

- REISS, P. C. "Descriptive, Structural, and Experimental Empirical Methods in Marketing Research." *Marketing Science* 30 (2011): 950–64.
- REISS, P. C., AND F. A. WOLAK. "Structural Econometric Modeling: Rationales and Examples from Industrial Organization," in *Handbook of Econometrics*, Amsterdam, Volume 6, Chapter 64, edited by J. J. Heckman and E. E. Leamer. Elsevier, 2007: 4277–415.
- ROBERTS, M. R., AND T. M. WHITED. "Endogeneity in Empirical Corporate Finance," in *Handbook of the Economics of Finance*, edited by G. M. Constantinides, M. Harris and R. M. Stulz, Chapter 7. Elsevier, Amsterdam, 2013: 493–572.
- ROSENBAUM, P. R. *Design of Observational Studies*. New York: Springer Science & Business Media, 2009.
- RUBIN, D. B. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (1974): 688–701.
- RUBIN, D. B. "Assignment to Treatment Group on the Basis of a Covariate." *Journal of Educational and Behavioral Statistics* 2 (1977): 1–26.
- RUSSO, F., AND J. WILLIAMSON. "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science* 21 (2007): 157–70.
- SARGAN, J. D. "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica* 26 (1958): 393–415.
- SECURITIES AND EXCHANGE COMMISSION(SEC). *Official Summary of Holdings of Officers, Directors, and Principal Stockholders*. Washington, DC: US Government Printing Office, 1936.
- SMITH, C. W., AND J. B. WARNER. "On Financial Contracting: An Analysis of Bond Covenants." *Journal of Financial Economics* 7 (1979): 117–61.
- SOLTES, E. "Private Interaction Between Firm Management and Sell-Side Analysts." *Journal of Accounting Research* 52 (2014): 245–72.
- STOCK, J. H.; J. H. WRIGHT; AND M. YOGO. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics* 20 (2002): 518–29.
- U.S. CONGRESS SENATE COMMITTEE ON GOVERNMENTAL AFFAIRS, AND A. RIBICOFF. *Interlocking Directorates Among the Major U.S. Corporations—Staff Study*, 95th Congress, 2nd Session, 1978. Washington, DC: U.S. Government Printing Office, 1978.
- U.S. FEDERAL TRADE COMMISSION. *Report on Interlocking Directorates*. Technical report. Washington DC: U.S. Government Printing Office, 1951.
- VERMEER, T. E.; C. EDMONDS; AND S. ASTHANA. "Organizational Form and Accounting Choice: Are Nonprofit or For-Profit Managers More Aggressive?" *The Accounting Review* 89 (2014): 1867–93.
- WRIGHT, S. "Correlation and Causation." *Journal of Agricultural Research* 20 (1921): 257–85.
- ZAKOLYUKINA, A. "Measuring Intentional GAAP Violations: A Structural Approach." Working paper, University of Chicago, 2015.