# Causal Inference in Accounting Research[1]

**Ian D. Gow**
Harvard Business School
email: igow@hbs.edu


**David F. Larcker**
Stanford Graduate School of Business
Rock Center for Corporate Governance
email: dlarcker@stanford.edu


**Peter C. Reiss**
Stanford Graduate School of Business
email: preiss@stanford.edu

April 27, 2015

# CAUSAL INFERENCE IN ACCOUNTING RESEARCH

IAN D. GOW, DAVID F. LARCKER, AND PETER C. REISS

ABSTRACT. This paper examines the approaches accounting researchers use to draw causal inferences using observational (or non-experimental) data. The vast majority of accounting research papers draw causal inferences notwithstanding the well-known difficulties in doing so with observational data. While a minority of papers seek to use quasi-experimental methods to draw inferences, there are concerns about how these methods are typically applied. We believe that accounting research would benefit from: a greater focus on the study of causal mechanisms (or causal pathways); increased emphasis on structural modeling of the phenomena of interest; and, more in-depth descriptive research. We argue these changes are possible and offer a practical path forward for rigorous accounting research.

## 1. INTRODUCTION

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

Most empirical research in accounting relies on observational (or non-experimental) data. This paper evaluates the different approaches accounting researchers use to draw causal inferences from observational data. 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 from the analysis of 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? That accounting researchers focus on causal inference 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 using such data seek to draw such inferences. Making causal inferences requires strong assumptions about the causal

relations between variables studied. 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. In Section 2, we discuss 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 the typical regression analyses that rely on observational data. Nonetheless, the difficulty of identifying, measuring, and controlling for all possible confounding variables leads many to be skeptical of the use of regression analyses of observational data for causal inference.

Recently, some social scientists have held out hope that better research designs and statistical methods can increase the credibility of causal 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 five studies claiming to study natural experiments (or "exogenous shocks") and ten studies using instrumental variables. Thus, quasi-experimental methods are used to some degree in accounting research, and we believe their use will increase in future research efforts.[1]

In Section 3, we examine and evaluate 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 the data relied upon. 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

---

[1]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.

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 *causal* 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, 5, and 6) identifies approaches that can provide a plausible framework for guiding future accounting research:

- There should be an increased emphasis on the study of causal mechanisms, i.e., 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.

- 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 readily acknowledge that, while structural modeling does not solve endogeneity concerns, it makes the assumed causal structure explicit and gives the researchers a rigorous way to assess what would happen if some features of the model change (i.e., to provide counterfactuals). We believe that causal diagrams can be a useful tool to convey the key elements of a structural model and can also act as a middle-level stand-in when structural modeling of a phenomenon is in its early stages or is incomplete.[2]

---

[2]"Middle-level" here refers to the placement of causal diagrams between relatively informal verbal reasoning and the rigors of a structural model.

- 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. Hopefully, these descriptive studies will provide insights that theorists can use to build models that empiricists can actually "take to data."

The remainder of the paper is structured as follows. Section 2 provides an overview of the issues observational data pose for drawing causal inferences in accounting research; it suggests frameworks for identifying and analyzing these issues. Section 3 evaluates the use of quasi-experimental methods in accounting research. Section 4 discusses mechanism-based causal inference. Section 5 illustrates how structural modeling approaches might be used by accounting researchers, with some emphasis of the strengths and weaknesses of this approach. In Section 6 we argue for richer descriptive research that can shed light on causal issues. Concluding remarks are provided in Section 7.

## 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 four categories: "Theoretical" (7); "Experimental" (12); "Field" (3); or "Archival Data" (103). For our discussion below, we collect the field and archival data papers into a single "Observational" category.

For each non-theoretical 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. Cohen, Hoitash, Krishnamoorthy, and Wright, 2014; Clor-Proell and Maines, 2014). In other cases, the abstracts reveal that authors have causal inferences as a goal. For example, de Franco, Vasvari, Vyas, and Wittenberg-Moerman (2014) asks "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, 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, inference.[3]

Of the 106 original papers using observational data, we coded 91 as seeking to draw causal inferences.[4] Of the remaining empirical papers, we coded 7 papers as having a goal of "description" (including two of the three field papers). For example, Soltes (2014b) 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 5 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 3 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 to the measurement of variables used by accounting researchers.

---

[3]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., that $Z$ causes $X$ and $Z$ causes $Y$. However, we do not find this kind of reasoning to be common in our survey.

[4]While we exclude research papers using experimental methods, all of these papers also seek to draw causal inferences.

In summary, we find that most original research papers use observational data and that 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-difference estimates, and propensity-score matching. 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.[5]

2.1.1. *Difference-in-difference 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 difference-in-difference (DD) estimators on their list of such quasi-experimental methods, along with "instrumental variables and regression discontinuity methods."[6] Enthusiasm for DD designs perhaps stems 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 instrumental variables and regression discontinuity 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 estimates of causal effects.

---

[5]There are settings where difference-in-difference and fixed effect estimators may deliver causal estimates. For example, if assignment to treatment is random, then it is possible for a difference-in-difference estimate using pre- and post-treatment 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.
[6]As Angrist and Pischke (2008, p. 228) argue that "DD is a version of fixed effects estimation," we discuss these methods together.

2.1.2. *Propensity score matching.* Another method that has become popular in accounting research is propensity score matching (PSM). Regression methods can be viewed as making model-based adjustments to address confounding variables. Stuart and Rubin (2007) 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]."

2.2. **Causal inference: A brief overview.** In recent decades, the definition and logic of causality has been revisited by researchers in such diverse fields as epidemiology, sociology, statistics, and computer science. Work by Rubin (1974, 1977) 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 tied to theoretical economic models. As discussed by Heckman and Pinto (2015), Haavelmo (1943, 1944) 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 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 that 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).

2.3. **Causal diagrams: A primer.** 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 (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, instrumental variable 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.

---

[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).

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.[8]

The criterion developed by Pearl (2009b) implies that very different conditioning strategies are needed for each of the causal diagrams (see Appendix A for a more formal treatment). 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, $Z$, that satisfies what Pearl (2009b) labels the "back-door criterion" (Pearl, 2009b, p.79).[9] 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.[10] 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.

We now discuss what each of the three graphs in Figure 1 suggest about how one might model the causal effect of $X$ on $Y$. Figure 1a is straightforward. It shows that we need to condition on $Z$ in order to estimate the causal effect of $X$ on $Y$. Note the notion

---

[8]That arrows have a direction accounts for the "D" in DAG (Directed Acyclic Graph). The acyclic ("A") component means that there must be no cycles in the graph. Cycles cause obvious problems in causal reasoning. An example would be $X \to Y \to Z \to X$. In this graph there is no ultimate cause. Graphs make a distinction between observed and unobserved random variables. In some cases, an unobserved joint determinant of two random variables will not be represented explicitly, but replaced by a dashed, "undirected" edge between those two random variables.

[9]Intuitively, the back-door criterion requires that $Z$ blocks (and does not open) "back-door" paths. A back-door path can be thought of as a way for $X$ to be associated with $Y$ due to associations with other variables rather than causal links from $X$ to $Y$. See Appendix A for a more formal discussion.

[10]Including variables in a linear regression framework only "controls for" only under strict assumptions, such as linearity in the relations between $X$, $Y$, and $Z$.

of "condition on" again is more general than just including $Z$ in a parametric (linear) model.[11] The need to condition on $Z$ leads to $Z$ being called a *confounder*.

Figure 1b is a bit different. Here $Z$ is a *mediator* of the effect of $X$ on $Y$. No conditioning is required in this setting to obtain an unbiased estimate of the effect of $X$ on $Y$. But, 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 that we should not condition on $Z$ to get such an estimate.

Finally in Figure 1c, we have $Z$ acting as what is referred to as a "collider" variable (Glymour and Greenland, 2008; Pearl, 2009a).[12] The back-door criterion not only implies that we need not condition on $Z$, but that we *should* not condition on $Z$ to get an unbiased estimate of the causal 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.[13]

2.3.1. *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 they is reasonable to assume that they are not mediators or colliders. Such a discussion is important because the use of mediators and colliders may 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 form:[14]

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

---

[11]Inclusion of $Z$ blocks the back-door path from $Y$ to $X$ via $Z$.

[12]The two arrows from $X$ and $Y$ "collide" in $Z$.

[13]Many intuitive examples of collider bias involve selection or stratification. Admission to a college could be a function of combined test scores and interview performance exceeding a threshold, i.e., $T + I \geq C$. 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.

[14]We alter the mathematical notation of Larcker et al. (2007) to conform with notation we use here.

Larcker et al. (2007) 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 the dependent variable. Another approach is to only include governance factors as independent variables, or:

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

> 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 a true relation between these variables.[15] While the with-and-without-controls approach used by Larcker et al. (2007) has intuitive appeal, a more robust approach would involve careful thinking about the plausible causal relations between the treatment variables, the outcomes of interest, and the candidate control variables.

---

[15]Note that Larcker et al. (2007) do not in fact use leverage as a control when performance is a dependent variable.

## 3. QUASI-EXPERIMENTAL METHODS IN ACCOUNTING RESEARCH

While most studies in accounting use methods of conditioning on confounding variables in some form of regression or matching model, 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 a clear trend toward the use of quasi-experimental methods. In this section, we discuss and evaluate the use of these methods in accounting research.

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 (as if) 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 (2012, 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.[16] An examination of these papers reveals how difficult it is to find a plausible natural experiment in observational data.

The most important concern is that that most "exogenous shocks" (e.g., SEC regulatory changes or court rulings) generally do not randomly assign firms into treatment and control groups. For example, an early version of Dodd-Frank contained a provision that would force companies to remove a staggered board structure.[17] It is tempting to use

---

[16]These are Lo (2014); Aier, Chen, and Pevzner (2014); Kirk and Vincent (2014); Houston, Jiang, Lin, and Ma (2014) and Hail, Tahoun, and Wang (2014).

[17]See Larcker, Ormazabal, and Taylor (2011).

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, this "natural experiment" does not randomly assign firms to treatment and control groups regarding a staggered board. That is, firms made an endogenous choice about staggered boards and the regulation is potentially forcing firms to change their choice. But firms might have a variety of margins through which they might respond to such a requirement, some of which may have valuation consequences of their own. 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 be a strong reason to believe that the natural experiment impacted assignment to treatment and this impact is uncorrelated 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 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 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 post-treatment variables as covariates can undermine the validity of causal inferences.

Extending our survey beyond research published in 2014, we find papers with very credible 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 his events (e.g., fires and natural disasters) relative to balance sheet date is plausibly random, the assignment to the disclosure and recognition conditions is also 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 been underlying events and disclosures, and takes care to address this concern.

Another credible 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 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.

If true natural experiments can be found, they are an excellent design 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. Michels, 2015; Li and Zhang, 2015), but care is needed in doing so.

3.2. **Instrumental variables.** Angrist and Pischke (2008, p.114) describe instrumental variables (IV) as "the most powerful weapon in the arsenal" of tools in econometrics. Accounting researchers have long used instrumental variables 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 instrumental variables.[18] Much has been written on the challenges for researchers in using instrumental variables (IV) 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 instrumental variables 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]

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

---

[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).

values of endogenous variables or industry averages as instruments, but these too are subject to criticism.[20]

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 instrumental variables using statistical tests (see Larcker and Rusticus, 2010). But such tests of instruments may be of dubious value. Consider, for example, the following simulation exercise. Consider a world where $X$ does not cause $y$, but we nevertheless estimate the regression $y = X\beta + \epsilon$. 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 $\epsilon$ are bivariate-normally distributed with variance of $1$ and $\rho(X, \epsilon) = 0.2$, and that $\sigma_\eta = 0.03$, we performed 1,000 IV regression simulations with 1,000 firm-level observations in each case. Both the 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.[21] 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.

---

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

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

This example illustrates why it is that no statistical test allows the researcher to verify that their instruments satisfy the exclusion restriction.[22] Obviously, causal inferences based on such instrumental variables 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 instrumental variables, we consider Armstrong, Gow, and Larcker (2013). Armstrong et al. (2013) 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 et al. (2013) to obtain an unbiased estimate of the causal relation. Among other analyses, Armstrong et al. (2013) use an instrumental variable to estimate the causal relation of interest. Armstrong et al. (2013) claim that their instrument is valid. Their reasoning is represented graphically in Figure 2. By conditioning on $Comp_{t-1}$ and using ISS recommendations as an instrument, Armstrong et al. (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.[23]

While the authors note this possibility: "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

---

[22]This is a corollary of the "causal reasoning is not statistical reasoning" point made above.

[23]In 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 et al. (2013) for simplicity, though a good causal analysis would consider these carefully.

indirect impact on corporate policies through its influence on shareholders' voting decisions)", they are unable to test the assumption (Armstrong et al., 2013, p. 912). Unfortunately, this assumption seems inconsistent with the findings of Gow, Larcker, McCall, and Tayan (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 2b, 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 et al. (2013, p. 912) is not valid.[24]

3.2.4. *IV in accounting research: An evaluation.* A review of instrumental variable applications in our 2014 survey suggests that accounting researchers have paid little heed to the suggestions and warnings of Larcker and Rusticus (2010); Lennox et al. (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 is a substantial burden of proof on researchers to justify appropriateness of making the stringent assumptions that IV estimators require.

3.3. **Regression discontinuity designs.** Recently, 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 et al., 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 regression discontinuity (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

---

[24]Armstrong et al. (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.

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).

3.4. **Quasi-experimental methods: An evaluation.** We agree that the revolution in econometric methods for causal inference represents an opportunity for accounting researchers. However, we suggest caution. 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. This suggests that researchers might temper their conclusions in light of the stringency of these assumptions. Additionally, researchers should discuss how their conclusions are likely to generalize beyond the (special) environment in which they were found.

## 4. Mechanisms and causal inference

> In complex fields like the social sciences and epidemiology, there are only few (if any) real life situations where we can make enough compelling assumptions that would lead to identification of causal effects.
>
> (Judea Pearl, cited in Freedman 2004, p. 287)

In the first half of the paper, we have argued that, while causal inference is the goal of most accounting research, it is difficult to find datasets and statistical methods that can produce reliable 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. These paths do not rely on researchers finding clever and appropriate identification strategies to answer questions of interest. The objective of the second part of this paper is to discuss these paths forward. The first path we discuss is an increased focus on causal mechanisms.

4.1. **Causal mechanisms: Some examples.** Accounting research is not alone in its reliance on observational data. It is therefore possible to look to other fields to find cases in which they have used observational data 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.

4.1.1. *John Snow and cholera.* A widely cited case of causal inference involves 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 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 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 accounting research.

4.1.2. *Smoking and heart disease.* A more recent illustration of plausible causal inference is discussed by Gillies (2011). Gillies (2011) 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 themselves say explicitly (p. 1535) that the excess mortality from cancer of the lung in cigarette smokers is caused by cigarette smoking." In contrast, while Doll and Peto (1976) also 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.[25] 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.[26] Thus, through a process involving multiple studies over two decades, a plausible set of causal mechanisms between smoking and atherosclerosis was established.

---

[25]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. 1201–2).

[26]"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).

4.1.3. *Implications of cases on mechanism.* 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.[27]

Our view is that accounting researchers can learn from fields such as epidemiology, medicine, and political science. 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. And 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.

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

---

[27]The persuasive force of Snow's natural experiment, coming decades before the work of Neyman (1923) and Fisher (1935), might be considered greater today.

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, it seems that the results would be even more compelling if there were more detailed evidence regarding a causal mechanism through which the estimated effect occurs and the authors appear to go to lengths to provide such an account.[28] For example, evidence of trading activity by local investors while driving prior to, but not after, the implementation of distracted driving restrictions would seem to be quite persuasive even incremental to the compelling identification strategy provided.[29]

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, Cohen, Lys, and Walther (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 et al., 2014).

It seems very clear that we need a much better understanding of the precise causal mechanisms for important accounting research questions. A clear discussion of these

---

[28]Brown et al. (2015, pp. 277-278) "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."

[29]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.

mechanisms will enable reviewers and readers to see what is being assumed and assess the reasonableness of the theoretical causal paths.[30]

## 5. STRUCTURAL MODELING

5.1. **Structural modeling: An overview.** In Sections 2 and 3, we suggested that researchers should consider using diagrammatic models to communicate the basis for their causal inferences and in Section 4 we suggested that researchers need to be much more precise in presenting their causal mechanism. This section explores a more 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 the idea that observed prices and quantities were equilibrium objects (i.e., that quantities demanded and quantities supplied must be equal at the equilibrium price), economists obtained a mathematical model that could be used to understand movements in observed prices and quantities. A question then arose as to whether economists could use observed prices and quantities to recover their underlying determinants. The models made it clear that the empiricist could only recover estimates of the unobserved demand and supply equations if certain exogenous (instrumental) 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).

---

[30]As we discuss in Section 6, this type of knowledge enables the the identification of gaps in research based on strictly archival data and somewhat unstructured verbal theorizing. Soltes (2014a) provides an insightful discussion of the pitfalls associated with an exclusive reliance on archival data.

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.

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. 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 costs, 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 Armstrong, Jagolinzer, and Larcker, 2010). The key question in this literature is whether certain kinds of incentives cause an increase in 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) illustrates a fairly typical approach in this literature. Efendi et al. (2007, p. 687) estimate a logistic regression with an indicator for restatements as the dependent variable, and measures of CEO incentive variables as the

independent variables of interest, as well as controls such as firm size, financial structure measures, and proxies for dimensions of corporate governance.[31]

A key assumption implicit in much of this literature is that restatements are a good proxy for *mis*statements (e.g., Efendi et al., 2007; Armstrong et al., 2010). This assumption is made because in practice we 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 (Dyck, Morse, and Zingales, 2010). For simplicity, we refer to the actions of these external monitors collectively as "subsequent investigations." If subsequent investigations are perfect and detect all misstatements not detected by the firm's auditor, then there is a one-to-one correspondence of misstatements and restatements.[32] Realistically, these subsequent investigations are not perfect, meaning that we need to recognize the difference between the two 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. We show that different assumptions made regarding the causal relationships between observable and unobservables lead to very different empirical implications. In doing so, we illustrate the value of a structural model in understanding these relationships, and why structural modeling is a useful approach for conducting accounting research with observational data.

5.2.1. *A simple model with a non-strategic auditor.* In our models, we assume that misstatements are deliberate and made by a single agent, whom we refer to as the 'CEO.' Although this is clearly a strong assumption, it allows us to deliver clear predictions about

---

[31]Efendi et al. (2007) also employ a case-control design, which involves matching firms with restatements with firms without. We do not focus on that aspect of their research design in our discussion here.
[32]There will still be a difference between *attempted* misstatements and actual misstatements due to the external auditor correcting some attempted misstatements.

the unobserved rate of misstatements.[33] The CEO is assumed to be rational in the sense that he or she trades off the expected benefits and costs of misstatements when deciding whether to misstate. 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).

We assume the firm's auditors independently catch 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.

To misstate performance, the CEO must exert costly effort, which is a fixed $C_M$. Combining this cost with the manager's expected benefits from of misstatement gives

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

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

To complete the structural model, the researcher must relate these objects to the observed data. Because we observe a (zero-one) indicator variable for restatements $y$ and not the actual misstatement behavior, we need to link the two. In this model, restatements are the result of three stochastic processes:

(1) The manager misstates (or not).

(2) The firm auditor detects and corrects an attempted misstatement (or not) .

---

[33]A strong (visible) assumption can be seen as a weakness of the model, but also can be seen as an advantage. For example, if the assumption is considered inappropriate, it is very clear how to revise the model.

(3) A subsequent investigation detects a misstatement and a restatement occurs (or

not).

Mathematically, this sequence can be modeled as

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

where $I(\cdot)$ is a zero-one indicator function equaling one when the condition in paren-

theses is true. The unobserved variables $y_A^*$ and $y_I^*$ reflect the likelihood that the firm's

auditor and subsequent investigations, respectively, will detect a misstatement. Notice

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)

$$
\begin{aligned}
E(y) &= E\left[ I(y_M^* \geq 0) \times (1 - I(y_A^* \geq 0)) \times I(y_I^* \geq 0) \right] \\[2mm]
&= \text{Pr(Misstate)} \times \text{Pr(Auditor Misses)} \times \text{Pr(Investigation Finds)} \tag{5} \\[2mm]
&= \beta^* \times (1 - p_A) \times p_I = \text{Pr(Restate)}
\end{aligned}
$$

From equation (3), $\beta^*$ is the (researcher's) forecasted probability that a misstatement oc-

curs, or

$$\beta^* = \text{Pr}\left( (1 - p_A)(1 - p_I)B - C_M \geq 0 \right) \tag{6}$$

At this point, the theory has delivered a structure for relating the *unobserved* proba-

bility of a misstatement, $\beta^*$, to the potentially estimable probability of a restatement. Now

we face a familiar structural modeling problem, which is that the model does not antici-

pate all the reasons why in practice these probabilities might vary across firm accounting

statements. For example, the theory so far does not point to different reasons why CEOs

might differ in their benefits and costs of misstatements. To move theoretical relations

closer to the data, researchers typically introduce observable reasons into them. Often

there is an ad hoc element to these additions. Empiricists are willing to do this, however,

because they believe that it is important to account for practical specifics that the theory does not recognize.

To illustrate this approach, here we assume that CEO's unobserved costs and benefits do vary systematically with observables. In addition, because these observables do not perfectly represent the observed and unobserved benefits, it is important to allow for unobservable differences in the costs and benefits of misstatements. One specification that does this assumes that

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

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.[34] The EQUITY variable is intended to capture the idea that, the more a CEO is rewarded for performance, the greater 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. 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 shall shortly see).[35]

---

[34]For expositional purposes, we assume away $X_B$ and $X_C$ in our analysis.

[35]Another key variable in the above model is the unobserved cost $\xi$. 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.

Further, the probability of a restatement becomes

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

where $\theta_0 = (1-p_A) \times p_I$, $\theta_1 = (1-p_A)(1-p_I)b_0 - m_0$, $\theta_2 = (1-p_A)(1-p_I)b_1$, and $\theta_3 = -m_1$.

Apart from the scalar multiple $\theta_0$, which can be absorbed into the probability statement (and thus is not identified), this probability model has the form of a familiar binary choice (e.g., a probit or logit model). Thus, the value of the structure imposed so far is that it can motivate the application of a familiar statistical model (as in Efendi et al., 2007), as well as explain how the estimated coefficients are potentially connected to quantities that impact the probability of a misstatement.

5.2.2. *Estimation of model of a non-strategic auditor.* To illustrate the application of this structural model to data, we assembled a dataset containing 5,000 firm-year observations on whether or not financial results were restated in a given year. Definitions of the variables in our data set are provided in Table 1.[36]

The data include variables that have previously been used to model restatements. The variable BIG4 is included because it is believed that Big 4 auditing firms have more expertise and are therefore more likely to catch misstatements. Similarly, the corporate governance literature suggests that board oversight from directors with accounting or finance backgrounds reduces the likelihood that CEOs will make misstatements. Finally, the variables INT and SEG are included to capture the complexity and costs of audits. Specifically, international companies and companies with more business segments are thought to raise the costs of auditing. Similar to prior accounting research, these additional variables are (somewhat arbitrarily) included to "control" for the conjecture that if the audit is more costly, less auditing will be done, and there will be more manipulations and restatements.

Table 1 reports descriptive statistics for the sample. CEOs on average receive about one million dollars in base pay and their incentive-related pay averages 26% of their total

---

[36]We discuss the source of the data in more detail below.

pay. About three-quarters of the sample has a Big 4 accounting firm as its auditor. The fraction of directors with financial expertise is less than ten percent. The average firm has about 4.4 business segments and is primarily based in the United States.

Table 2 reports the results of logit regressions in which the dependent variable is the restatement indicator variable. The table contains both a simple specification containing an intercept along with the two CEO pay variables, and a more intricate specification involving the other variables we have in our data. For each specification we report the estimated coefficients of the logit and the corresponding marginal effects evaluated at the sample averages of the exogenous variables.

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 for $b_1 > 0$. Specifically, more base pay is associated with more restatements, while more equity-based compensation is associated with fewer restatements.

Besides the intercepts and EQUITY 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 assuming that $b_1 > 0$.

5.2.3. *A simple model with a strategic auditor.* A key weakness of the nonstrategic auditor model analyzed above 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.[37]

In the previous model, the firm's auditor impacted the manager's misstatement benefits through $p_A$. 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, $C_R$, due to 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 that 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 $\alpha^*$ and the CEO independently misstates with probability $\beta^*$. These probabilities are such that each party to the game has no incentive to change their randomized strategy. That is:

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

$$(1 - p_A^*)(1 - p_I)B - C_M = 0 \tag{9}$$

---

[37]Here we have in mind the findings of Dyck et al. (2010) who show that many egregious forms of misstatements are detected subsequently by employees, directors, regulators, and the media.

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

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

Solving these two equations for the equilibrium probabilities $\alpha^*$ and $\beta^*$ yields:

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

From these equations, we can calculate the equilibrium probability of a restatement[38]

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

This equation tells us how the observed (or measurable) 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

$$\Pr(\text{Restate}) = \frac{C_A C_M (1 - p_{IL})}{(p_{IH} - p_{IL})(1 - p_I) C_R B}. \tag{12}$$

We now are in a position to use the theory to help interpret the conflicting logistic regression results in Table 3.

---

[38] As part of the solution, we require $\alpha^*$ and $\beta^*$ to be probabilities between zero and one. This is true provided $C_R$ and $B$ satisfy the inequalities $C_R > \frac{C_A}{(p_H - p_L) p_I}$ and $B > \frac{C_M}{1 - p_I}$.

Equation (12) shows that the presence of a strategic external auditor changes how the CEO's incentives impact the probability of a restatement.[39] 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 non-strategic 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.

To have a better sense of how one might connect the strategic auditor theory to the logistic models in Table 2, suppose, similar to ways we motivated (7), that

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

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

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

---

[39]Notice that the probability statement in equation (12) differs from that 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.

$$= \frac{\theta_0 + \theta_1 \text{INT} + \theta_2 \text{SEG} + \theta_3 \text{SALARY} + \theta_4 \text{INT} \times \text{SALARY} + \theta_5 \text{SEG} \times \text{SALARY}}{1 + \theta_6 \text{EQUITY}} \tag{14}$$

Notice that the $\theta$'s absorb unknown quantities such as $r_0$ and $p_0$, and that the denominator intercept is normalized to one. 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 generalized method of moments (GMM). This method attempts to match so-called sample moments to what the structural model implies the 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 probabilty expression in equation (14). Because we need at least as many moments as we have $\theta$ parameters to estimate (there are seven $\theta$'s in the model), we use seven sample moments, each of the form:

$$\mathcal{M}_j = \sum_{i=1}^{5,000} X'_{ji} \left[ \text{RESTATE}_A - \text{Pr(Restate)}_A \right].$$

where Pr(Restate) comes from equation (14).[40] The $X_j$ used in the moments include all explanatory variables. Thus, because $X$ includes a dummy variable for whether the firm is an international company, the corresponding moment equation seeks to match the sample 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 on the sample of 5,000 firms. 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 a normalization.[41] For instance, the coefficient ratio $\theta_3/\theta_0$ estimates the ratio of cost parameters $m_1/m_0$. Since $m_1$ is the cost coefficient on SALARY

---

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

[41]A simple way to see this might be to observe that there are seven $\theta$ coefficients and eight underlying structural parameters.

and $m_0 > 0$ for costs to make sense, the sign of $\theta_3/\theta_0$ 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.

Similarly, $\theta_6$ equals the (scaled) misstatement benefit coefficient on the EQUITY variable. Recall that the descriptive logit regression coefficients in Table 2 suggest EQUITY has a negative affect 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. However, it is important to note that $\theta_6$ is only marginally significant.

The one sign that does not make sense given the other coefficient estimates is the negative sign on $\theta_5$, but this coefficient is insignificantly different from zero. Further, even with a sample size of 5,000, restatements are relatively rare, thus making it difficult for the model to predict them with much accuracy.

While the coefficient magnitudes do not allow us to estimate the underlying benefits and costs to managers from misstatements, we can illustrate the value of the model by performing a counterfactual calculation. There are many different counterfactuals that could be considered. For illustrative purposes, we can ask 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 see that removing equity-based pay does not change the equilibrium frequency of misstatements, but does change the frequency of high effort auditing. From (5), the model and the data we find

$$\frac{\text{Pr(Restate} \mid \text{No Equity)}}{\text{Pr(Restate} \mid \text{Equity)}} = \frac{\beta^* \times (1 - p_A^{**}) \times p_I}{\beta^* \times (1 - p_A^*) \times p_I} = \frac{(1 - p_A^*)}{(1 - p_A^*)} = 1.10.$$

This result tells us that the restatement rate would increase by 10% (from 10.24% to 11.25%) if equity-based incentives were withdrawn. 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.

5.2.4. *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 non-strategic auditor and that 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 non-strategic auditor, where a negative (and weakly statistically significant) coefficient on EQUITY is found. However a second, and in our view 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 3, where we used causal diagrams as a kind of (non-parametric) causal model. Here we have shown 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 non-strategic 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.

5.3. **Limitations of structural models.** There are, of course, costs to developing and estimating structural models. First, structural models can be technically demanding to develop. Additionally, when constructing a theoretical model that can be taken the data, the empirical researcher will typically be forced to make simplifications that a pure theorist might never make and that other empiricists criticize as unrealistic. Unfortunately, there is a substantial divide between theoretical and empirical research in accounting. With 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, structural models do not avoid the need to make causal assumptions. A causal diagram representing the model developed above is provided in Figure 3.[42] As can be seen, we are assuming that $p_I$ is independent of EQUITY. But is quite plausible that these investigations are conducted (in part) by a regulator who is as strategic as the auditor in our model, thus giving rise to a link between EQUITY and $p_I$. We also assume 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

---

[42]While the mixed-strategy of our model has $\beta$ not being a function of $B$ or $C_M$ and $\alpha$ not being a function of $C_A$ or $C_R$, we have retained these links as being plausible in a more general model.

structural model, albeit at some cost. Evaluation of their accuracy then could be made via in- or 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. In our analyses, the data we used were generated by simulation using the model with the strategic auditor, so in a sense we might feel confidence that we have the right model. But, in practice, we do not have this kind of insight into the process generating the data we observe and we need to make assumptions.[43] Of course, there is no guarantee that after all the effort that went into developing the model, the estimates will make sense or that the model will otherwise be validated. Despite these challenges, we believe that there is significant value in making the theory underlying empirical research transparent and rigorous.[44]

## 6. 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. Unfortunately, there are very few studies published in top accounting journals that focus on providing deep description of institutions relevant to 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.[45] We believe that accounting research could benefit substantially from more

[43]We should be careful not to ascribe greater value to the latter set of estimates, as they are based on the kind of *causal* knowledge (i.e., that the data were simulated using a particular) model that do not in general come from *statistical* descriptions of the data.

[44]The use of structural models in accounting research has been fairly limited to date. Recent examples include Gerakos and Kovrijnykh (2013), Zakolyukina (2015), and Bertomeu, Ma, and Marinovic (2015). These three 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.

[45]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

in-depth descriptive research. As we discuss below, this type of research is essential for those who seek to develop structural models or improve our understanding of causal mechanisms.[46]

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. For example, using proxy statements and conversations with actual executives and consultants, Healy (1985) studies the bonus contracts of 94 large US companies and identifies a common structure of these bonus plans, including the existence of caps and floors (Healy, 1985, p. 89). 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 actually used by firms. These institutional features can be used to identify precise mechanisms and also as elements of structural models in which other features might be motivated more directly by economic theory.

Recent published research suggests an increased recognition of the value of descriptive research. Soltes (2014b) examines the interactions between sell-side analysts and company management in one firm that granted him proprietary access 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 (2014b) provides a plausible mechanism through which information transfers hypothesized in more traditional empirical papers actually occur.

---

have provided a credible signal of a willingness to publish descriptive studies of institutionally interesting settings.

[46]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; Bower, 1986; Mintzberg, 1973). Cyert et al. (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."

That private communication with management is an important source of informa-
tion is confirmed by Brown, Call, Clement, and Sharp (2015). Brown et al. (2015) sur-
vey 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
researchers focus on. For instance, analysts do not use the "red flags" used by academics
to identify manipulation; and analysts generally are not attempting to uncover manip-
ulation and use forecasts, not as ends in themselves, but to figure out the stock price
target. These insights should shape research seeking to develop hypotheses and models
of accounting information and analyst behavior.

Other fields provide interesting examples likely to be of interest to accounting re-
searchers. For example, Ahern (2014) examines 183 illegal insider networks using pri-
mary source documents from the SEC, DOJ, and various public records. It provides rich
insights into investor networks and it suggests questions for future work. For example,
network relationships can be divided into familial (3%), business-related (35%), friend-
ships (35%), or "not clear" (21%). Insiders are more likely to be an accountant or lawyer,
less likely to be a Democrat, and more likely to have a "criminal record." These results
provide new institutional insights that have the potential to identify causal mechanisms
regarding information transfer and disclosure. In economics, Bloom and Van Reenen
(2007) provide a descriptive study of 732 medium-sized firms to assess whether manage-
ment practices, such as lean manufacturing and use of incentives, are related to produc-
tivity. These descriptive results have lead to the development of theoretical models of
innovating and productivity along with increasingly sophisticated empirical studies.

## 7. CONCLUDING REMARKS

In this paper, we examined the approaches used by accounting researchers to draw
causal inferences from analyses of observational (or non-experimental) data. The vast

majority of empirical papers using such data seek to draw causal inferences, notwith-standing the well-known difficulties with doing so. While some papers seek to use quasi-experimental methods to develop unbiased estimates of causal effects, we find that 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.

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 they offer a viable and exciting path forward.

## References

Ahern, K. R. (2014). Information networks: Evidence from illegal insider trading tips. *SSRN Electronic Journal*.

Aier, J. K., L. Chen, and M. Pevzner (2014, November). Debtholders' demand for conservatism: Evidence from changes in directors' fiduciary duties. *Journal of Accounting Research 52*(5), 993–1027.

Angrist, J. D. (1990, June). Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records. *American Economic Review 80*(3), 313–336.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996, June). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444.

Angrist, J. D. and J.-S. Pischke (2008, December). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010, May). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Armstrong, C. S., I. D. Gow, and D. F. Larcker (2013, December). The efficacy of shareholder voting: Evidence from equity compensation plans. *Journal of Accounting Research 51*(5), 909–950.

Armstrong, C. S., A. D. Jagolinzer, and D. F. Larcker (2010, May). Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research 48*(2), 225–271.

Berk, R. A. (1999). Review of *Observational Studies* by Paul Rosenbaum. *Journal of Educational and Behavioral Statistics 24*(1), 95–100.

Bertomeu, J., P. Ma, and I. Marinovic (2015, March). How often do managers withhold information: A structural estimation. Working paper.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010, December). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics 50*(2-3), 296–343.

Bloom, N. and J. Van Reenen (2007, November). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics 122*(4), 1351–1408.

Bower, J. L. (1986). *Managing the Resource Allocation Process*. Harvard Business School Press.

Brown, L. D., A. C. Call, M. B. Clement, and N. Y. Sharp (2015, February). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research 53*(1), 1–47.

Brown, N. C., H. Stice, and R. M. White (2015, May). Mobile communication and local information flow: Evidence from distracted driving laws. *Journal of Accounting Research 53*(2), 275–329.

Burns, N. and S. Kedia (2006, January). The impact of performance-based compensation on misreporting. *Journal of Financial Economics 79*(1), 35–67.

Cannon, J. N. (2014, April). Determinants of "sticky costs": An analysis of cost behavior using United States air transportation industry data. *The Accounting Review 89*(5), 1645–1672.

Clor-Proell, S. M. and L. A. Maines (2014, June). The impact of recognition versus disclosure on financial information: A preparer's perspective. *Journal of Accounting Research 52*(3), 671–701.

Cohen, J. R., U. Hoitash, G. Krishnamoorthy, and A. M. Wright (2014, January). The effect of audit committee industry expertise on monitoring the financial reporting process. *The Accounting Review 89*(1), 243–273.

Correia, M. M. (2014, April). Political connections and SEC enforcement. *57*(2–3), 241–262.

Cready, W. H., A. Kumas, and M. Subasi (2014, September). Are trade size-based inferences about traders reliable? Evidence from institutional earnings-related trading. *52*(4), 877–909.

Cyert, R. M., H. A. Simon, and D. B. Trow (1956, October). Observation of a business decision. *The Journal of Business 29*(4), 237–248.

Czerney, K., J. J. Schmidt, and A. M. Thompson (2014, November). Does auditor explanatory language in unqualified audit reports indicate increased financial misstatement risk? *The Accounting Review 89*(6), 2115–2149.

de Franco, G., F. P. Vasvari, D. Vyas, and R. Wittenberg-Moerman (2014, March). Debt analysts' views of debt-equity conflicts of interest. *The Accounting Review 89*(2), 571–604.

Doll, R. and R. Peto (1976). Mortality in relation to smoking: 20 years' observations on male british doctors. *British Medical Journal 2*, 1525–1536.

Dunning, T. (2012, September). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press.

Dyck, A., A. Morse, and L. Zingales (2010, November). Who blows the whistle on corporate fraud? *The Journal of Finance 65*(6), 2213–2253.

Efendi, J., A. Srivastava, and E. P. Swanson (2007, September). Why do corporate managers misstate financial statements? the role of option compensation and other factors. *Journal of Financial Economics 85*(3), 667–708.

Erkens, D. H., K. R. Subramanyam, and J. Zhang (2014, April). Affiliated banker on board and conservative accounting. *The Accounting Review 89*(5), 1703–1728.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd.

Fox, W. F., L. Luna, and G. Schaur (2014, February). Destination taxation and evasion: Evidence from U.S. inter-state commodity flows. *57*(1), 43–57.

Freedman, D. A. (2004, August). Graphical models for causation, and the identification problem. *Evaluation Review 28*(4), 267–293.

Freedman, D. A. (2009, November). *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge University Press.

Gerakos, J. and A. Kovrijnykh (2013, July). Performance shocks and misreporting. *Journal of Accounting and Economics 56*(1), 57–72.

Gillies, D. (2011). The Russo-Williamson thesis and the question of whether smoking causes heart disease. In P. M. Illari, F. Russo, and J. Williamson (Eds.), *Causality in the Sciences*, pp. 110–125. Oxford University Press.

Glymour, M. M. and S. Greenland (2008). Causal diagrams. In K. J. Rothman, S. Greenland, and T. L. Lash (Eds.), *Modern Epidemiology* (3 ed.)., Chapter 12. Lippincot, Williams and Wilkins.

Goldberger, A. S. (1972, November). Structural equation methods in the social sciences. *Econometrica 40*(6), 979–1001.

Gow, I. D., D. F. Larcker, A. McCall, and B. Tayan (2013). Sneak preview: How ISS dictates equity plan design. *Stanford Closer Look Series*.

Guedhami, O., J. A. Pittman, and W. Saffar (2014, March). Auditor choice in politically connected firms. *Journal of Accounting Research 52*(1), 107–162.

Haavelmo, T. (1943, January). The statistical implications of a system of simultaneous equations. *Econometrica 11*(1), 1–12.

Haavelmo, T. (1944, July). The probability approach in econometrics. *Econometrica 12*.

Hail, L., A. Tahoun, and C. Wang (2014, May). Dividend payouts and information shocks. *Journal of Accounting Research 52*(2), 403–456.

Healy, P. M. (1985, April). The effect of bonus schemes on accounting decisions. *7*(1-3), 85–107.

Heckman, J. and R. Pinto (2015, February). Causal analysis after Haavelmo. *Econometric Theory 31*(01), 115–151.

Holland, P. W. (1986, December). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Houston, J. F., L. Jiang, C. Lin, and Y. Ma (2014, March). Political connections and the cost of bank loans. *Journal of Accounting Research 52*(1), 193–243.

Iliev, P. (2010, June). The effect of SOX Section 404: Costs, earnings quality, and stock prices. *The Journal of Finance 65*(3), 1163–1196.

Imbens, G. W. and K. Kalyanaraman (2012, January). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. Cambridge University Press.

Kim, K., E. G. Mauldin, and S. Patro (2014, April). Outside directors and board advising and monitoring performance. *Journal of Accounting and Economics 57*(2–3), 110–131.

Kirk, M. P. and J. D. Vincent (2014, July). Professional investor relations within the firm. *The Accounting Review 89*(4), 1421–1452.

Larcker, D. F., G. Ormazabal, and D. J. Taylor (2011, August). The market reaction to corporate governance regulation. *Journal of Financial Economics 101*(2), 431–448.

Larcker, D. F., S. A. Richardson, and I. Tuna (2007). Corporate governance, accounting outcomes, and organizational performance. *The Accounting Review 82*(4), 963–1008.

Larcker, D. F. and T. O. Rusticus (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics 49*(3), 186–205.

Lee, D. S. and T. Lemieux (2010, June). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Lennox, C. S., J. R. Francis, and Z. Wang (2012, March). Selection models in accounting research. *The Accounting Review 87*(2), 589–616.

Li, Y. and L. Zhang (2015, January). Short Selling Pressure, Stock Price Behavior, and Management Forecast Precision: Evidence from a Natural Experiment. *Journal of Accounting Research 53*(1), 79–117.

Listokin, Y. (2008, September). Management always wins the close ones. *American Law and Economics Review 10*(2), 159–184.

Lo, A. K. (2014, May). Do declines in bank health affect borrowers' voluntary disclosures? Evidence from international propagation of banking shocks. *Journal of Accounting Research 52*(2), 541–581.

McCrary, J. (2008, February). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Michels, J. (2015, February). Disclosure versus recognition: Inferences from subsequent events. Working paper, University of Pennsylvania.

Miller, G. S. and D. J. Skinner (2015, May). The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure. *Journal of Accounting Research 53*(2), 221–239.

Mintzberg, H. (1973). *The Nature of Managerial Work*. Harper & Row.

Minutti-Meza, M. (2014, May). Issues in examining the effect of auditor litigation on audit fees. *Journal of Accounting Research 52*(2), 341–356.

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 (1995, May). 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*(18), 1198–1203.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5*(4), 465–472.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics Surveys 3*(0), 96–146.

Pearl, J. (2009b, September). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Reiss, P. C. (2011, November). Descriptive, structural, and experimental empirical methods in marketing research. *Marketing Science 30*(6), 950–964.

Reiss, P. C. and F. A. Wolak (2007). Structural econometric modeling: Rationales and examples from industrial organization. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 64, pp. 4277–4415. Elsevier.

Roberts, M. R. and T. M. Whited (2013). Endogeneity in empirical corporate finance. In *Handbook of the Economics of Finance*, Chapter 7, pp. 493–572. Elsevier.

Rosenbaum, P. R. (2009, October). *Design of Observational Studies*. Springer Science & Business Media.

Rubin, D. B. (1974, October). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology 66*(5), 688–701.

Rubin, D. B. (1977, March). Assignment to treatment group on the basis of a covariate. *Journal of Educational and Behavioral Statistics 2*(1), 1–26.

Russo, F. and J. Williamson (2007, July). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science 21*(2), 157–170.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*(3), 393–415.

Soltes, E. (2014a, May). Incorporating field data into archival research. *Journal of Accounting Research 52*(2), 521–540.

Soltes, E. (2014b, March). Private interaction between firm management and sell-side analysts. *Journal of Accounting Research 52*(1), 245–272.

Stock, J. H., J. H. Wright, and M. Yogo (2002, October). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics 20*(4), 518–529.

Vermeer, T. E., C. Edmonds, and S. Asthana (2014, April). Organizational form and accounting choice: Are nonprofit or for-profit managers more aggressive? *The Accounting Review 89*(5), 1867–1893.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20*, 257–285.

Zakolyukina, A. (2015, April). Measuring intentional GAAP violations: A structural approach. Working paper, University of Chicago.

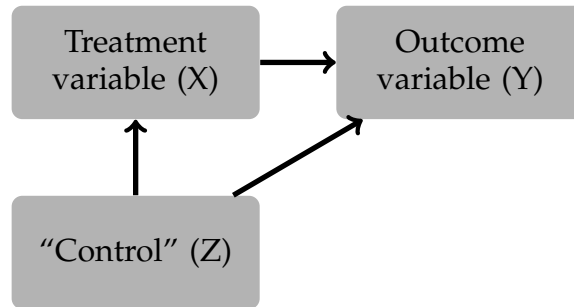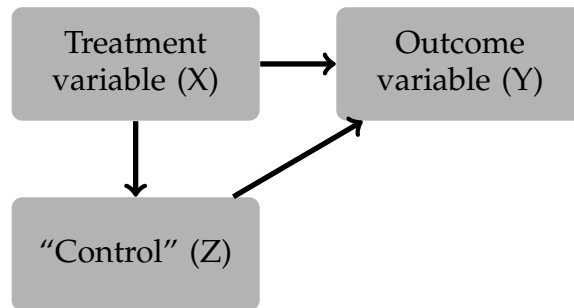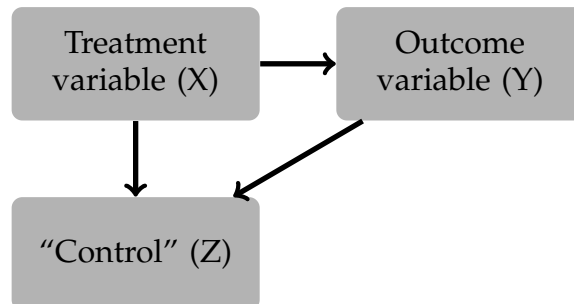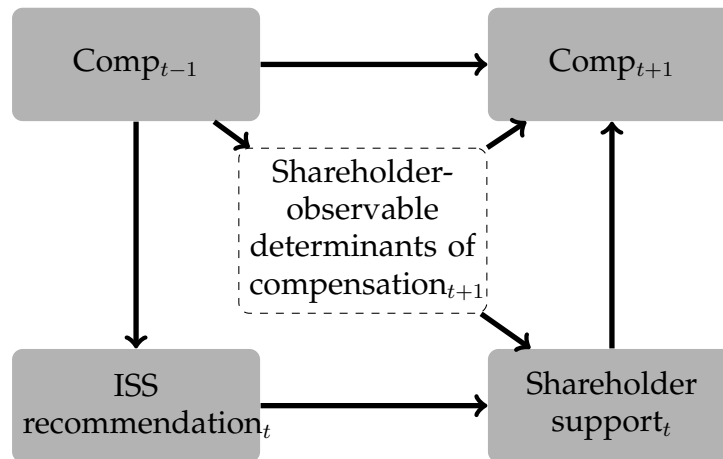FIGURE 1.  Three basic causal diagrams

(A) $Z$ is a confounder



(B) $Z$ is mediator



(C) $Z$ is a collider

FIGURE 2. Identifying effects of shareholder support on compensation

(A) Causal diagram for Armstrong et al. (2013)



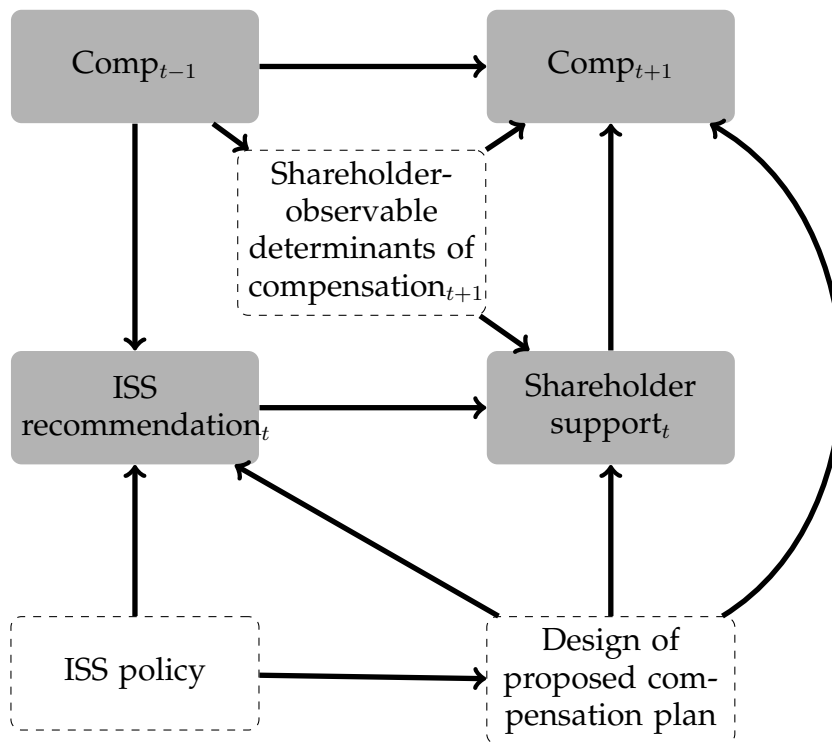(B) Alternative causal diagram for Armstrong et al. (2013)

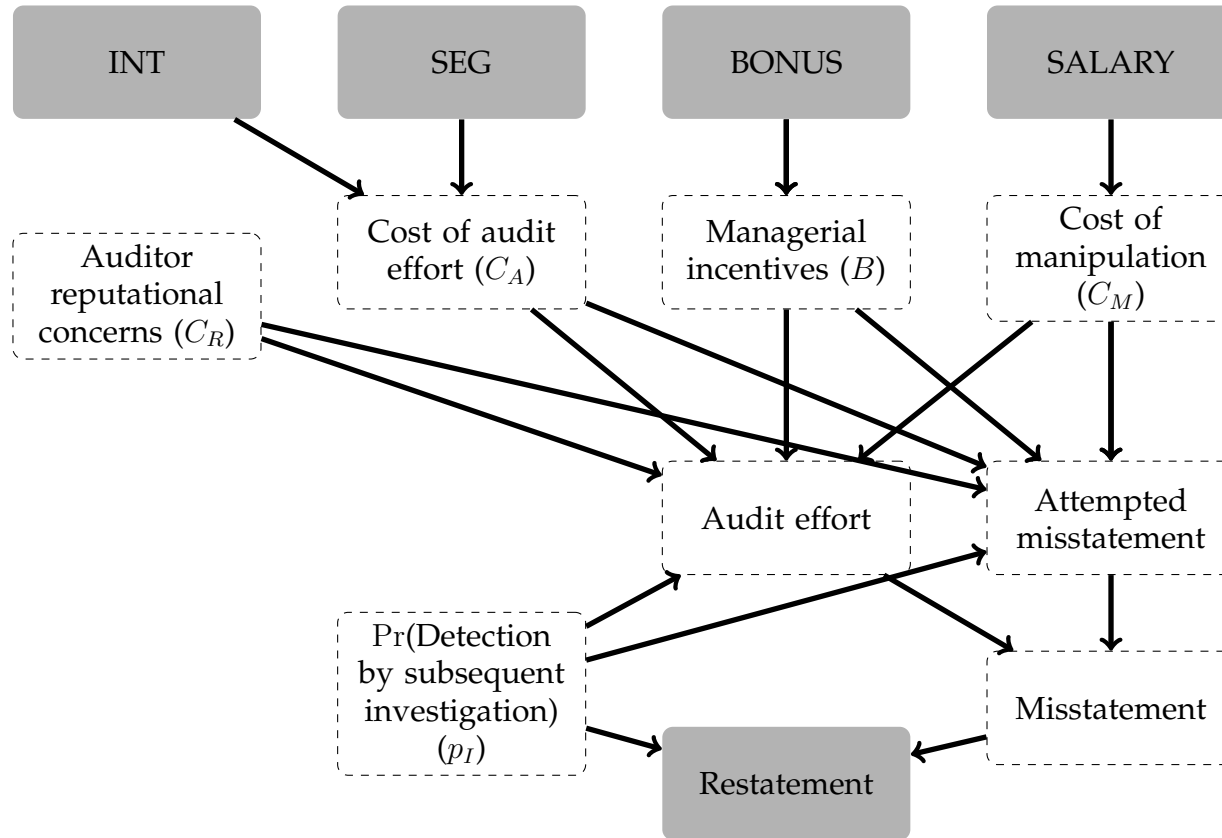FIGURE 3. Causal diagram for strategic auditor model

TABLE 1. Descriptive Statistics

**RESTATE** is a zero-one indicator for whether the firm made a restatement in a particular year. **SALARY** is the CEO's annual base salary. **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 finance background. **INT** is percentage of non-US revenue for the firm. **SEG** is number of the firm's business segments.

| Variable | Sample Mean (Std Dev) |
|---|---|
| **RESTATE** | 0.102 0.303 |
| **SALARY** | 0.95 0.15 |
| **EQUITY** | 0.26 0.29 |
| **BIG4** | 0.76 0.43 |
| **FINDIRECT** | 0.08 0.08 |
| **SEG** | 4.41 3.02 |
| **INT** | 0.30 0.46 |

TABLE 2.  Logit Regression Results

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 (**EQUITY**) and controls.  The controls are **SALARY**, the CEO's annual base salary., **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 finance background; **INT** is percentage of non-US revenue for the firm. **SEG** is number of the firm's business segments.

| Coefficient on | Specification 1 | | Specification 2 | |
| --- | --- | --- | --- | --- |
| | Coefficient (Std Dev) | Marginal Effect (Std Dev) | Coefficient (Std Dev) | Marginal Effect (Std Dev) |
| **Intercept** | -2.210* 0.308 | | -2.786* 0.349 | |
| **SALARY** | 0.118 0.317 | 0.010 0.027 | 0.010 0.327 | 0.001 0.027 |
| **EQUITY** | -0.644* 0.185 | -0.055* 0.016 | -0.673* 0.201 | -0.056* 0.017 |
| **BIG4** | | | 0.104 0.130 | 0.009 0.011 |
| **FINDIRECT** | | | -0.058 0.740 | -0.005 0.062 |
| **INT** | | | 0.630* 0.120 | 0.052* 0.010 |
| **SEG** | | | 0.086* 0.021 | 0.007* 0.002 |

TABLE 3. Logit GMM Estimates for the Strategic Auditor Model

This table presents results for GMM estimates of the strategic auditor model of Section 5.

| | Estimated Coefficient | Bootstrap Std Error |
|---|---|---|
| $\theta_0 = \dfrac{(1-v_0)a_0 m_0}{(p_0-v_0)r_0 b_0}$ | 0.0342 | 0.058 |
| $\theta_1 = \dfrac{(1-v_0)a_1 m_0}{(p_0-v_0)r_0 b_0}$ | 0.0022 | 0.068 |
| $\theta_2 = \dfrac{(1-v_0)a_2 m_0}{(p_0-v_0)r_0 b_0}$ | 0.0145 | 0.009 |
| $\theta_3 = \dfrac{(1-v_0)a_0 m_1}{(p_0-v_0)r_0 b_0}$ | 0.0108 | 0.051 |
| $\theta_4 = \dfrac{(1-v_0)a_1 m_1}{(p_0-v_0)r_0 b_0}$ | 0.1015 | 0.058 |
| $\theta_5 = \dfrac{(1-v_0)a_2 m_1}{(p_0-v_0)r_0 b_0}$ | -0.0064 | 0.008 |
| $\theta_6 = \dfrac{b_1}{b_0}$ | 0.4557 | 0.291 |

Standard errors are the estimated standard deviation of 500 replications.

APPENDIX A. 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 $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle node $m$ is in $Z$, or
   (2) $p$ contains an inverted fork (or *collider*) $i \to m \leftarrow j$ 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 a an ordered pair of variables $(X, Y)$ in a 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 $X$.[47]

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*

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

*where $P(y|x)$ stands for the probability that $Y = y$, given that $X$ is set to level $X = x$ by external intervention.*[48]

A.2. **Application of back-door criterion to Figure 1.** Applying the back-door criterion to Figure 1a 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 $X \leftarrow Z \to Y$. So by conditioning on $Z$, we can estimate the causal effect of $X$ on $Y$. This situation is a generalization of linear model in which $Y = X\beta + Z\gamma + \epsilon_Y$ and $\epsilon_Y$ 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 $\beta$, the causal effect of $X$ on $Y$, but by including $Z$ in the regression, we get an unbiased estimate of $\beta$. In this situation, $Z$ is a *confounder*.

Turning to Figure 1b, 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 $X \to Y$ that exists is $X \to Z \to Y$, 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 that we should not condition of $Z$ to get such an estimate.

Finally in Figure 1c, we have $Z$ acting as what Pearl (2009a, p. 17) refers to as a "collider" variable.[49] Again, we see that $Z$ does not satisfy the back-door criterion, because $Z$

---

[47]The "arrow into $X$" is the portion of the definition that is explains the "back-door" terminology.

[48]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.

[49]The two arrows from $X$ and $Y$ "collide" in $Z$.

is a descendant of $X$. However, $\emptyset$ again satisfies the back-door criterion. First, contains no descendant of $X$. Second, the only path other than $X \to Y$ that exists is $X \to Z \leftarrow Y$, 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 that we should not condition of $Z$ 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 $G_{\overline{X}}$ 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 $G_{\overline{X}}$
   (2) $S$ does not $d$-separate $Z$ from $X$ in $G$

Applying this definition to Figure 2, we can evaluating the instrument used in Armstrong et al. (2013). There we have 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 2a. If create $G_{\overline{X}}$ by deleting the single arrow emanating from *Shareholder support$_t$*, we can see that there are two back-door paths running from $Y$ to $Z$: $Z \leftarrow S \to U \to Y$ and $Z \leftarrow S \to Y$. However, both of these paths are blocked by $S$ and the first requirement is satisfied. The second requirement is clearly satisfied as $Z$ is directly linked to $X$.[50] Note that analysis can be expressed intuitively as requiring that the ISS recommendation only affects $Comp_{t+1}$ through its effect on *Shareholder support$_t$*, and that $Comp_{t+1}$ has an effect on *Shareholder support$_t$*.

But this analysis presumes that the causal diagram Figure 2a is correct. Armstrong et al. (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 represented in Figure 2a 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 2b and the plausible existence of this path suggests that the instrument of Armstrong et al. (2013, p. 912) is not credibly valid for the causal effect they seek to estimate.

---

[50]This 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.