From Casual to Causal Inference in Accounting Research: 
The Need for Theoretical Foundations

Jeremy Bertomeu  
Baruch College  
City University of New York  
jeremy.bertomeu@baruch.cuny.edu

Anne Beyer  
Graduate School of Business  
Stanford University  
abeyer@stanford.edu

Daniel J. Taylor  
The Wharton School  
University of Pennsylvania  
dtayl@wharton.upenn.edu

November 15, 2015

Abstract
On December 5th and 6th 2014, the Stanford Graduate School of Business hosted the Causality in the Social Sciences Conference. The conference brought together several distinguished speakers from philosophy, economics, finance, accounting and marketing with the bold mission of debating scientific methods that support causal statements. We highlight key themes from the conference as relevant for accounting researchers. First, we emphasize the role of formal economic theory in informing empirical research that seeks to draw causal inferences, and offer a skeptical perspective on attempts to draw causal inferences in the absence of well-defined constructs and assumptions. Next, we highlight some of the conceptual limitations of quasi-natural experimental methods that were discussed at the conference, and discuss the role of structural estimation. Finally, we illustrate many of the points from the conference by estimating a novel, theoretically-grounded measure of disclosure costs.

We thank an anonymous referee, Chris Armstrong, Qi Chen, Paul Fischer, Joseph Gerakos, Ian Gow, Wayne Guay, David Larcker, Christian Leuz, Ivan Marinovic, Jeremy Michels, Valeri Nikolaev, Ro Verrecchia, and Ivo Welch for helpful conversations and comments. We are grateful to the conference organizers for giving us the opportunity to write this piece.
1. Introduction

On December 5th and 6th 2014, the Stanford Graduate School of Business hosted the Causality in the Social Sciences Conference. The conference brought together distinguished speakers from philosophy, economics, finance, accounting and marketing with the bold mission of debating scientific methods that support causal statements in the social sciences. The conference was structured around a keynote by philosopher Nancy Cartwright and presentations by five notable economists: Joshua Angrist, Guido Imbens, Charles Manski, Peter Reiss and John Rust. Three panel discussions offered complementary views about causality in the areas of finance, accounting and marketing. Each panel was composed of journal editors and leading researchers, and focused on the state-of-the-art and opportunities for future research. The conference approached several issues of interest to accounting researchers, such as the role of theoretical foundations, quasi-natural experimental methods, and structural estimation.

The conference organizers asked us to report on and synthesize the proceedings from three unique perspectives: audience members, accounting researchers, and a unique blend of individuals with empirical and theoretical research interests. We were asked to illustrate some of the themes of the conference using examples drawn from the accounting literature and to include a brief application. While initially challenging, we found it intellectually rewarding to step outside the silos of “empirical researcher” and “theoretical researcher” and discuss issues of common interest. Perhaps this will be the conference’s legacy: the conference showed methodological and cross-disciplinary academic fertilization at its best, and has the potential to disseminate new perspectives about accounting topics.

Accounting researchers may look at the vast and divergent views being espoused in the proceedings of the conference published in this issue of Foundation and Trends, and wonder about
the takeaways as it relates to accounting. Our synthesis aims to highlight what we think were a few of the key takeaways, and expand on each using several examples and settings familiar to accounting researchers. In this respect, we caveat that our synthesis necessarily reflects our own views as accounting researchers. We view this report as contributing to the debate in accounting surrounding causal inferences, methodology, and the role of formal theory in informing empirical work—certainly not as the final word on that debate. Our synthesis is intended to complement a growing body of literature exploring issues related to causality within accounting, among others Gow, Larcker, and Reiss (2014) and Chen and Schipper (this issue). Given the shared topic—causality in accounting research—many of the issues raised at the conference and included in our synthesis also appear in prior work.

Two key themes emerged at the conference, and these themes pervade our discussion. First, regardless of method, causal inferences rely on untestable core assumptions (Cartwright, this issue). Identification of causal channels does not come from statistical techniques but from assumptions. For example, if the assumptions behind instrumental variable regression (IV) are satisfied (e.g., the instruments satisfy the exclusion restriction), then IV can be used to estimate a causal effect. While institutional knowledge might tell us whether certain assumptions are more plausible than others, certain untestable assumptions will always be necessary.

Second, there was considerable skepticism about statistical techniques commonly referred to as “quasi-natural experimental methods” (Rust, this issue), and whether strong, causal inferences typically associated with the use of such methods are reasonable. We found this

---

1 See also the Special Issue on Causality in the October 2014 volume of Accounting, Organizations, and Society (e.g., Balakrishnan and Penno, 2014; Gassen, 2014; Luft and Shield, 2014; Lukka, 2014; Van der Stede, 2014).

2 For example, the exclusion restriction of instrumental variables, the parallel trends assumption of difference-in-differences, and the continuity assumption of regression discontinuity are all inherently untestable (e.g., Larcker and Rusticus, 2010; Roberts and Whited, 2013; Gow, Larcker, and Reiss, 2014).
particularly surprising given the increasing emphasis on these methods within the accounting literature. Concerns about these methods focused on generalizability and the ability of these methods to identify underlying causal mechanisms in the absence of formal theory.³

At this point, let us simply note two practical implications of these themes for originating research that pushes the frontiers of the accounting literature. First, making assumptions should not be taken as a scientific compromise, or something to hide; instead, assumptions should be presented and opened for discussion. The message is simple: assumptions should be explicit, transparent, and deliver new insights. Without a clear understanding of the assumptions of empirical tests and measures, causal inferences will remain elusive.

Second, regardless of empirical method, a link between formal economic theory and empirical work is essential in helping researchers identify causal effects of interest. For example, when researchers measure systematic risk, they employ estimates of Beta from the Capital Asset Pricing Model—an empirical measure of systematic risk derived from formal theory; when researchers measure information asymmetry, they employ estimates of Kyle’s λ or the probability of informed trade (PIN)—empirical measures of information asymmetry between market participants derived from formal theory. It strikes us that none of these measures is particularly intuitive in the absence of the corresponding formal economic theory and assumptions. Formal theory provides the assumptions that guide how we interpret the relations in the data—for example, how we interpret the covariance between a firm’s returns and the market return. In this regard, formal theory makes transparent the assumptions that underlie each of the above measures.

Relative to the other fields represented at the conference, accounting research tends not to be very clear about what assumptions it is making. For example, researchers in accounting

³ These issues are not new to the economics literature (e.g., Heckman, 2000, 2005; Heckman and Vytlacil, 2007).
generally rely on intuition and less formal verbal descriptions to motivate empirical measures of theoretical constructs rather than mathematical descriptions derived from formal theory. The advantage of this approach is clear: it does not constrain empirical work to topics studied by theorists. Indeed, without additional assumptions driven by the particulars of the data, few formal theories are institutionally rich enough to suggest specific measures of the construct being studied. The disadvantage of this approach is also clear: great confusion—about the validity of empirical proxies and what theoretical constructs they are intended to capture.

As a result, the accounting literature simultaneously features a great variety of empirical proxies and fundamental disagreements about what these proxies capture. Few of the numerous proxies of accounting quality, conservatism, or proprietary costs are derived from formal theory. Consequently, the theoretical construct of interest is only defined at an intuitive level, and the assumptions that underlie various proxies are not transparent. For example, what assumptions are necessary for the asymmetric timeliness coefficient of Basu (1997) to measure conservatism? While the proxy has great intuitive appeal, it is only more than 15 years later—and only after its widespread use—that we are starting to get a sense for the assumptions that underlie this measure (Ball, Kothari, and Nikolaev, 2013). Understanding these assumptions, and their validity, is critical to reliably estimating causal effects. More generally, it seems difficult to credibly identify the causal effect of an intervention (e.g., regulation) on a theoretical construct, if that construct is only defined at an intuitive level, or if we do not understand the assumptions that underlie our empirical measures of the construct.

This is now being given the opportunity to change, as more researchers within accounting are interested in probing into the foundations of common measures of accounting quality,
conservatism, and proprietary costs—among other examples—each with the potential effect of a dramatic rethink in the measures and relations that have been exhausted in prior research.

We structure the remainder of the paper as follows. In Section 2, we discuss the role of formal economic theory in informing empirical research that seeks to make causal statements, and offer a skeptical perspective on attempts to draw causal inferences in the absence of well-defined constructs and assumptions. In Section 3, we highlight some of the conceptual limitations of quasi-natural experimental methods that were discussed at the conference; limitations that appear to be underappreciated within the accounting literature. In Section 4, we discuss the role of structural estimation as an emerging method within accounting research, and provide a simple application that illustrates many of the themes discussed at the conference. Specifically, we estimate a novel, theoretically-grounded measure of voluntary disclosure costs based on Verrecchia (1983).

2. Causal inferences: The role of theory

A sound, well-defined theory is necessary to draw causal inferences from observational data. All speakers and participants agreed that causal inferences require a theory—theory provides a framework for making predictions and interpreting estimated correlations. In the absence of a theory, correlations do not carry any economic meaning. Correlations may have statistical meaning, in the sense that two variables might co-move, but interpreting that co-movement requires a theory. Indeed, Heckman (2005) suggests that the first two tasks that confront empirical researchers seeking to draw causal inferences are: (1) use theory to describe a hypothetical world, and (2) identify the causal channel in that hypothetical world (see also Heckman and Vytlacil, 2007). To illustrate this idea further with an example, we consider an econometric method developed in Heckman (1979).
Given the need for a theory, it should not be surprising that there was widespread agreement among conference participants that researchers must understand: (a) the assumptions of the underlying theory that motivates the analysis, (b) the assumptions that underlie empirical proxies for the theoretical constructs of interest, and (c) the assumptions that underlie the specific statistical method being used to estimate correlations in the data. Thus, the debate appeared to be less about the need for theory and the need to understand assumptions, and more about the usefulness of different types of theory for interpreting estimated correlations.

Kahn and Whited (this issue) offers a distinction between two types of theory—“mathematical” and “verbal.” We expand on this classification and differentiate between formal economic theory, formal statistical theory, and verbal theory, where we use the terms formal and mathematical interchangeably. We adopt these labels as they were used by conference participants, and as they are expositionally convenient. All three types of theory are present across the related fields of economics, finance, accounting and marketing, and we have employed all three types ourselves. Undoubtedly, there is a role for all three types of theory in scientific inquiry. The remainder of this section discusses the extent to which each type of theory can inform empirical work seeking to draw causal inferences.

2.1 Formal economic theory

We define formal economic theory as a mathematical representation of the choice problem faced by individuals or organizations. Formal economic theory is a mathematical characterization of a system of preferences (or objectives) and institutions (or contexts) that explain individual decision making. Formal economic theory is micro-founded, meaning that it is based on the assumption that an individual, with a clearly stated set of objectives and constraints, makes the
best feasible choice to meet these objectives. In this regard, such theory is well-suited to make
causal statements on how an individual will respond to a change in the environment. As Heckman
(2000) describes:

“Formal economic models are logically consistent systems within which hypothetical
‘thought experiments’ can be conducted to examine the effects of changes in parameters
and constraints on outcomes. Within a model, the effects on outcomes of variation in
constraints facing agents in a market setting are well defined. Comparative statics
exercises formalize Marshall’s notion of a ceteris paribus change which is what economists
mean by a causal effect.”

One of the defining characteristics of formal economic theory is that it uses mathematics
to describe a hypothetical world and identify the causal channels in that world. Of course, formal
economic theory also requires a verbal narrative that explains the nature of assumptions in terms
of economic behavior, but the assumptions themselves, and their implications, are derived using
formal logic and expressed using more precise mathematical descriptions. In the terminology of
Heckman (2005), formal economic theory uses mathematics to describe a hypothetical world and
identify the causal channels in that world. Mathematical formalism ensures the absence of
ambiguity in logic and assumptions, and makes transparent any inconsistencies.

There are important advantages and disadvantages to relying on formal economic theory
to motivate empirical predictions and describe causal channels. In regards to disadvantages, few
formal economic theories are sufficiently rich to fully characterize economic phenomena. Many
economic theories relevant to accounting, such as agency or asset pricing theory, explain only a
tiny portion of the variation in observables and exhibit many anomalies; in fact, most theories—
even seminal ones that have guided decades of research—tend to be statistically rejected. Thus,
there are usually significant factors outside the context of an economic model that appear in the
data.
Second, while well-suited for identifying causal channels in a hypothetical world described by the model’s assumptions, formal economic theories tend to focus on unobservable characteristics of a decision. As a result, formal economic theories can sometimes be difficult to falsify because a rejection of the theory can be interpreted as an empirical measurement problem. For example, Roll (1977) famously notes that the theoretical market portfolio implied by the CAPM is, in practice, not observable or tradable. The theoretical market portfolio of the CAPM comprises the aggregate wealth of the economy and includes both private companies and human capital. Since accurate measures of this portfolio are difficult to come by, empirical tests that reject the CAPM cannot distinguish between a rejection of the empirical proxy or a rejection of the underlying theory (see also, Levy and Roll, 2010).

Third, economic theories are often heavy in assumptions, since restrictions must be made on behaviors and technologies to facilitate solving an equilibrium. In some cases, these assumptions are known to be violated in the data (e.g., individuals do not exhibit constant absolute risk aversion). However, even when assumptions are known to be violated, formal economic theory can still be useful in guiding empirical work. For example, despite knowledge that that the underlying assumptions are violated, economic models such as the Capital Asset Pricing Model or Modigliani-Miller have guided decades of research and provided a useful framework for thinking about issues related to diversification, leverage, and asset pricing. This suggests a usefulness for formal theory even when underlying assumptions are violated. Indeed, it is the falsification of these theories that has helped organize and direct subsequent research.

A theory that is light on assumptions will support few causal statements: in the limit, with no assumption, there is no theory or causal claim. Vice-versa, the stronger the assumptions, the stronger the causal statements one can make, but at the cost of assumptions to be taken at faith. An
economic theory is most successful when it can make extensive predictions with few assumptions or unobservable parameters. Researchers thus face a trade-off when articulating an economic theory. This returns us to a theme of the conference. Making assumptions should not be taken as a scientific compromise, or something to hide—all types of theory, whether verbal or formal, require assumptions. Assumptions should be presented and opened for discussion. Without an understanding of the assumptions, causal inferences will remain elusive.

We now turn to the strengths of formal economic theory, especially as it relates to informing empirical work that seeks to identify causal channels and draw causal inferences. First, in the context of an economic model, a causal channel is often labeled as a comparative static. A comparative static is, in literal terms, the nature of a causal relation. For this reason, social sciences rely heavily on formal theory.

Even though it may not connect to any data, a formal model is the cleanest setting in which strong causal claims may be made (e.g., “in the context of the assumed model, X causes Y, which causes Z”). In a formal model, we know precisely what economic forces are (and are not) present. A formal model provides the opportunity to distinguish the effects of multiple economic forces—forces that may not be separately observable in the data. For example, investigating whether accounting conservatism could arise in the absence of the stewardship role of accounting—the role of accounting in mitigating agency problems and facilitating efficient contracts—is possible only within the confines of a formal economic theory. Only within a hypothetical world can we isolate the valuation role of accounting and examine what reporting would look like in the absence of a stewardship role (e.g., Armstrong, Taylor, and Verrecchia, 2015). Purely empirical explorations are inherently difficult because the stewardship role of accounting is always present in the data—making valuation and stewardship difficult to empirically distinguish.
Second, in addition to identifying a causal channel, formal economic theory makes transparent all of the intermediate steps in a causal chain. During the conference, Nancy Cartwright illustrated causal chains using the famous “Rube Goldberg machines” from early 20th century comic strips. Rarely is the causal channel in the real world as simple as “X causes Y” (e.g., corporate governance causes future performance). To illustrate this point, consider the following hypothetical causal chain:

\[
\begin{align*}
X & \rightarrow Y \rightarrow Z \rightarrow W \\
& \downarrow \quad \downarrow \quad \downarrow \\
& A \quad B \rightarrow C
\end{align*}
\]

Each arrow describes a causal effect: X causes Y, which causes both A and Z, Z causes B, which causes C, which is also caused by W. Similarly, via a sequence of deductions, this chain implies X causes A, and X causes C.

By simultaneously considering multiple links in the causal chain, formal economic theory can offer clear guidance on which factors researchers should and should not control for in their empirical tests. In the above example, even though A and C may be correlated, neither A causes C, nor the reverse. Instead, both are caused by X. Another implication of the causal chain is that both X and W cause C. Consequently, an empirical researcher may want to control for W before testing whether X incrementally causes C. On the other hand, controlling for B when testing for a causal relation between X and C would mask the true causal relation between X and C. In this regard, formal theory can also provide insight on what empirical researchers should not control for.

---

4 The machines involve several inter-connected devices, where each device activates an adjacent device so as to create a cause-and-effect link between two ostensibly unrelated devices.
Within accounting, there is significant emphasis on including variables as controls to rule out confounding effects, and less emphasis on what variables should be excluded from the list of controls. Researchers appear more concerned about potential bias from the erroneous exclusion of control variables than potential bias from the erroneous inclusion of control variables. For example, a recent trend in empirical research is the use of large numbers of fixed effects (e.g., both firm and year). While the inclusion of such effects can be useful in ruling out alternative explanations and tightening empirical identification, depending on the form of the causal effect and any measurement error, including such fixed effects can bias regression coefficients and remove variation predicted by the respective theory being tested (see Gormley and Matsa, 2014 for technical details and examples).

We refer to this practice as the kitchen sink approach to control variables—include as many controls as possible with the hope that the correlations of interest remain statistically significant. As the above diagram illustrates, the kitchen sink approach, i.e., controlling for Y, Z, A, and B when estimating the causal relation between X and C, can potentially confound attempts to estimate causal effects. This highlights one clear advantage of formal theory: by simultaneously considering multiple links in the causal chain, it can provide clear predictions on the constructs that should be excluded from the list of controls ex ante (as opposed to ex post). To further illustrate these points, consider the following two studies.

Kelly and Ljungqvist (2012, KL) offers a formal economic theory linking information among investors to asset prices, and tests the model by examining the effect of an exogenous reduction in analyst coverage on asset prices. KL conjectures that an exogenous reduction in analyst coverage causes an increase in information asymmetry, which in turn causes a decrease in prices. KL examines price changes around reductions in analyst coverage without controlling for
changes in information asymmetry. The causal chain implied by the formal economic theory in KL suggests that if one controlled for (a perfect measure of) information asymmetry, there would be no relation between a reduction in analyst coverage and prices. Thus, formal economic theory provides a reasonable basis to exclude information asymmetry as a control. In the above causal chain, analyst coverage is X, information asymmetry is Y, and prices are A. X is conjectured to affect A only through an affect on Y.

Contrast this with LaFond and Watts (2008, LW). LW offers some intuition for why conservatism might cause a reduction in information asymmetry among shareholders. Consistent with this intuition, LW finds that conservatism is negatively related to information asymmetry. However, LW notes that when controlling for firm size, there is no incremental relation between conservatism and information asymmetry. Since size might measure information asymmetry, LW argues that this result is not inconsistent with the hypothesis that conservatism causes a reduction in information asymmetry. On the one hand, if size were predicted to affect conservatism only through information asymmetry, then controlling for size would constitute over-controlling and would not be appropriate. On the other hand, if size were predicted to affect conservatism through channels other than information asymmetry, controlling for size would be appropriate. In this case, because the theory that relates size to information asymmetry and conservatism is ambiguous, we cannot draw strong causal inferences.

Of course, causal effects can entail multiple channels. We could modify the hypothetical causal chain by drawing an arrow connecting X with A.
In this circumstance, X causes A directly, and also causes A indirectly through Y. In principle, some channels could be reinforcing, others countervailing. Formal economic theory can provide insights into the nature of each channel and offer a distinction between multiple channels. Such insights might be overlooked if one simply estimates the causal effect of X on A. This estimate would reflect the net effect of the two channels.

To illustrate this point, consider the effect of an increase in private information on expected returns. Conventional wisdom in the empirical literature is predicated on the notion that an increase in private information causes an increase in expected returns due to information asymmetry. However, in the model of Lambert, Leuz, and Verrecchia (2012) an increase in private information increases expected returns via the “information asymmetry effect” (informed investors know disproportionately more) but decreases expected returns via the “average precision effect” (on average, there is less uncertainty). Given the conventional wisdom, empirical studies that find an increase in private information causes a decrease in expected returns—while potentially consistent with the average precision channel—would likely be dismissed as anomalous and potentially abandoned. Distinguishing between countervailing causal channels provides a way to reconcile these opposing directional effects. In this regard, formal economic theory can often provide rich empirical predictions that distinguish one causal channel from another.

2.2 Formal statistical theory

We define formal statistical theory as using mathematics to specify a probabilistic process that links one variable with another. In this sense, specifying a linear regression model is offering a statistical theory that assumes a specific dependent variable is a linear combination of specific explanatory variables and noise. For example, Rust (2014) quotes Tom Sargent:
“A [statistical] model is a probability distribution over a sequence (of vectors), usually indexed by some fixed parameters.”

Similar to formal economic theory, formal statistical theory still retains the precision of using mathematics to describe assumptions and constructs. Unlike formal economic theory however, human decisions or choice problems are outside the context of these models—there is no optimization problem being solved by investors, managers, or any other party. The primary advantage of statistical models is that assumptions are made specifically with data in mind—and hence provide fewer challenges for researchers attempting to identify appropriate empirical proxies.

Within accounting, statistical models are often used to formalize assumptions in an attempt to provide insights on how to estimate (unobserved) latent variables as a function of observables. Indeed, statistical theory has been helpful in guiding accounting researchers on issues related to construct validity and test specification. We discuss three popular examples within the literature: valuation models, implied cost of capital models, and properties of earnings.

*Valuation models.* Ohlson (1995) and Feltham and Ohlson (1995) are two celebrated statistical models that link accounting amounts (e.g., earnings and book values) with cash flows, dividends, and market price. These models use assumptions about the earnings and book value generating processes, and formation of asset prices, to solve for prices as a function of earnings and book values. An entire stream of accounting research, the “value-relevance literature,” uses the statistical models of Ohlson (1995) and Feltham and Ohlson (1995) to guide their empirical tests (see Beaver, Barth, and Landsman, 2001; Holthausen and Watts, 2001).
Implied cost of capital models. Models of the implied cost of capital are effectively valuation models that are reverse engineered to solve for the unobserved discount rate that investors apply to the firm’s equity. These models often start with the discounted dividend model of stock prices, assume clean surplus accounting, and show that prices can be expressed as an infinite sum of discounted residual income. Additional assumptions are then made about the time-series evolution of discount rates, earnings, and terminal value, and the equation for price is inverted to solve for the unobserved discount rate as a function of observables (e.g., Gebhardt, Lee, and Swaminathan, 2001).

Properties of earnings. Statistical models have also been employed to recover latent variables related to properties of firms’ earnings. Consider the asymmetric timeliness coefficient of Basu (1997), hereafter ATC, as a measure of conservatism. While the measure has great intuitive appeal, its construct validity is unclear. Ball, Kothari, and Nikolaev (2013) offers a mathematical definition of conservatism, makes assumptions about the earnings, cash flows, and return generating processes, and derives the ATC in the context of these assumptions. The statistical model offered in Ball, Kothari, and Nikolaev (2013) provides guidance on the assumptions and conditions under which the ATC is (and is not) a valid measure of conservatism. Similarly, Nikolaev (2014) offers a mathematical definition of accounting quality, makes assumptions about the earnings and cash flow generating process, and uses a statistical model to show how one can use higher order moment conditions to calculate measures of accounting quality.

As these examples illustrate, statistical models make assumptions about unobserved latent variables, which can then be used to develop an empirical identification strategy. The success of
statistical models within accounting is owed mainly to using assumptions to narrow the field of investigation away from questions about the endogenous nature of accounting information and prices, which are the focus of formal economic models. This is perhaps both the greatest strength and the greatest weakness of statistical models—they rely on assumptions to bypass nuanced issues related to choice. Consequently, the mechanical nature of statistical models allows for only limited causal statements.

For example, statistical models operate in much the same way that stepping on a car’s brake causes the car to slow down. Pressing on the brake depresses a piston; this pressurizes the brake cylinder with fluid, squeezes the brake pads together, generates friction and slows the rotation of the wheel—the system is mechanical. If we want to understand, say, how inattention may cause the driver to miss a stop light, this mechanical component will play a part, but it is probably secondary to the choice to press the brake.

Consider Ohlson (1995) and Feltham and Ohlson (1995): in what sense does a change in earnings cause a change in price? The accounting information system and price formation process are exogenously specified and outside the context of the model. Thus, the mechanisms that might cause this change—investors updating beliefs—are outside the context of the model. Consider an implied cost of capital model: in what sense does an increase in uncertainty about future cash flows cause a change in the cost of capital? Here, too, the price formation process is exogenous, and thus the causal mechanism—investors’ beliefs and risk aversion—is outside the context of the model. In the same way, conservatism and accounting quality are exogenous in Ball, Kothari, and Nikolaev (2013) and Nikolaev (2014)—and thus the economic determinants of these properties (e.g., agency issues) are outside the context of these models.5

---

5 For a given model, any process or construct that is not explicitly modeled as a choice variable or outcome of a choice variable, e.g., a fixed parameter, is considered exogenous with respect to that model.
The message here should be clear: statistical theories are helpful in addressing issues of empirical measurement and test specification. Statistical models force one to precisely define the theoretical construct of interest and associated assumptions. Within the scope of the model, there is no ambiguity in assumptions or constructs. Further, due to their mathematical precision, they provide researchers with obvious directions as to how to improve the assumptions—either by manner of weaker restrictions or incorporation of more primitive latent variables. However, because the choice problem is outside the context of statistical models, it should be recognized that they are of limited use in identifying causal mechanisms. For this reason, statistical models are often used in settings where causal inferences are not the objective.

2.3 Verbal theory

We define verbal theory as a narrative exposition of assumptions, theoretical constructs, and associated causal mechanisms. In the terminology of Heckman (2005), verbal theory uses words to describe a hypothetical world and identify the causal channel in that world. Importantly, all types of theory have some verbal core; as such, the difference between verbal theory and formal theory is more a difference in the degree of precision and formalization than it is a difference in essence.

There are a number of advantages to relying on verbal theory to motivate hypotheses and describe the causal channel. First, verbal descriptions do not require substantive training in mathematics. While verbal descriptions are less precise than mathematical descriptions, verbal descriptions can seemingly be more intuitive, easier to develop, and easier to follow than mathematical descriptions. Second, reliance on verbal descriptions to motivate hypotheses and tests does not constrain empirical work to cover topics and settings for which there is a pre-existing
body of formal theory. Verbal theory can easily be shaped around a rich set of institutional details and nuances; few existing formal theories are institutionally rich enough to suggest specific measures of the construct being studied. Third, verbal theory and associated intuition may more closely describe the decision-making process followed by accountants and practitioners than formal optimization problems; few real world decisions of interest to accounting researchers are made by solving optimization problems and we may never be sure about the intricate details about a preference or utility function required to fully specify an economic model.

These advantages notwithstanding, there are two significant disadvantages of verbal theory. First, because verbal theory lacks mathematical formalism, it is necessarily restricted to intuitive predictions and reasoning. In other words—in the absence of ex post justification—verbal theory cannot generate a prediction that is unexpected or counterintuitive. This inherent limitation prevents verbal theory from delivering insights that would significantly change or reverse one’s ex-ante prior about a topic. The depth of insights provided by verbal theory are necessarily restricted to the depth of the researcher’s intuition (see also, Wolpin, 2013).

This disadvantage stands in stark contrast to formal theory, which can provide predictions that, at first blush, are not intuitive. In many cases, formal structure can help extent one’s reasoning and intuition. Consider the following examples: In the absence of the CAPM, would intuition suggest not only that the covariance between a firm’s stock return and the market return is a measures of risk, but the only risk that is priced by the market? That providing more precise public information can increase incentives to collect private information (McNichols and Trueman, 1994)? That conservatism can lead to less efficient debt contracts (Gigler, Kanodia, Sapra, and Venugopalan, 2009) or more earnings manipulation (Bertomeu, Darrough, and Xue, 2015)? That shareholders and firms can benefit from earnings manipulation (Arya, Glover, and Sunder, 2003)?
That increasing the sensitivity of the manager’s wealth to changes in stock price can reduce his incentives to adopt risky, positive net present value projects (e.g., Lambert, Larcker, and Verrecchia, 1991; Ross, 2004)? Most of the major advances in the social sciences have come through the creation of knowledge that was not, at least initially, intuitive—knowledge that helped change a widely-held belief.6

The second disadvantage is that, because of the lack of formalism and precision, theoretical constructs and assumptions are often not well-defined. In the absence of the disciplining force of mathematics, internal consistency is not assured—either in assumptions or logic. Because violations of assumptions or logic are often less transparent, ex post rationalization of empirical results is often easier with verbal theory (as opposed to formal economic theory or formal statistical theory). In the extreme, verbal theory can incorporate any empirical finding into the narrative, and in doing so, become entirely non-falsifiable.

To see how these two disadvantages might affect an empirical study interested in drawing causal inferences, consider the following hypothetical example. Suppose we want to study the causal effect of conservatism on the market reaction to earnings news. We offer a verbal theory: conservatism mitigates agency problems, which reduces managerial rent extraction and hence results in a higher valuation per-dollar of earnings. Cognizant of the statistical issues associated with drawing causal inferences, we patiently search for an ideal quasi-natural experiment. Suppose that such a setting presents itself; we find an exogenous shock that affects the market reaction to earnings only through accounting conservatism, and the setting appears to satisfy the assumptions behind our chosen quasi-natural experimental method (e.g., instrumental variables, regression discontinuity, etc.). Suppose, further, that we use the ATC to measure conservatism. We have

---

6 This theme appears central to the philosophy of science. For example, while recognizing the social nature of scientific research, Thomas Kuhn refers to the possibility of a reversal in beliefs as a unique aspect of science.
offered a verbal theory and have tested this theory in an ideal quasi-natural experimental setting. Have we identified the causal effect of conservatism on the market reaction to earnings?

First, a definition of the theoretical construct of interest, conservatism, is not provided. What is meant by conservatism? One approach might be to define conservatism according to the adage “anticipate no profits but anticipate all losses.” However, this definition still lacks precision of a mathematical model and thus still opens the door for controversy over what is meant by “conservatism” (e.g., Armstrong, Taylor, and Verrecchia, 2015). Ambiguity in the construct of interest is clearly undesirable from the standpoint of understanding its causal effects. An alternative approach might be to define conservatism mathematically—this makes the definition and underlying assumptions explicit. One can then ask: Do I agree with this definition? Under this definition, what assumptions need to be satisfied for ATC to be a valid measure of conservatism?

Second, the assumptions under which the empirical proxy is a valid measure of the theoretical construct of interest, conservatism, have not been considered. Paradoxically, Ball, Kothari, and Nikolaev (2013) show that one of the assumptions that underlies the ATC as a measure of conservatism (as they define it) is the notion that earnings do not convey incremental information to the capital market—all value-relevant events are immediately impounded in stock prices regardless of how and when they are incorporate in earnings (see also Gigler, Kanodia, Sapra, and Venugopalan, 2009, p. 784, for a related discussion). As a result—even within an ideal experimental setting—it is conceptually unclear how a researcher could claim to identify the causal effect of conservatism on the market reaction to earnings when using ATC to measure conservatism.

This hypothetical example highlights the key disadvantage of verbal theory as commonly used: constructs and assumptions are not well-defined. In principle, it seems difficult to identify
the causal effect of an intervention (e.g., regulation) on a theoretical construct, if that construct is only defined at an intuitive level, or if we do not understand the assumptions that underlie our empirical measures of the construct. Of course, it makes no difference whether assumptions and the logical chain are enumerated using words or mathematical symbols, as long as they are explicit. In practice, however, Kahn and Whited (this issue) suggest that social scientists lack the skills needed to employ verbal theory with the same precision as mathematics. No theory or research design is perfect, but clarity in assumptions is key. As researchers, we must prefer known assumptions to unknown assumptions and should be skeptical of attempts to draw causal inferences when constructs and assumptions are not well-defined.

Before continuing, we close the section with a brief note. Each type of theory has its own strengths and weaknesses, and its own place in the literature. One sentiment expressed by some at the conference was “Why do I need a formal theory if I can describe it in words?” In our experience, this is a common sentiment within accounting and suggests a far more limited role for formal theory in informing empirical work than the one we outline above. We want to push back against this sentiment. Formal theory is a necessary scientific tool that can provide insights not attainable from pure verbal descriptions and empirical relations. While it is true that assumptions underlying most formal theories might be violated, the (implicit) assumptions underlying most verbal theories are also violated. The key difference is that the former makes these violations explicit. In this regard, formal theory—the processes of formalizing and understanding assumptions—assures consistency both in logic and in measurement, both of which are critical to empirical work that seeks to draw causal inference.

In addition to emphasizing that a well-defined theory is necessary to draw causal inference from observational data, conference participants also discussed and expressed skepticism about the strong, causal statements typically associated with the use of quasi-natural experimental (QNE) methods (e.g., difference-in-differences, instrumental variables, and regression discontinuity designs). We found this particularly surprising given the increasing emphasis on these methods within the accounting literature.

Many scholars argue that accounting research should move away from estimating “associations” and move toward using quasi-natural experimental methods to estimate causal effects. Attempts to push the frontiers of accounting research, and move from descriptive inferences to causal inferences, are desirable. While we agree with the essence of this sentiment—that researchers should strive to the lofty goal of establishing causality—the discussion at the conference (and in particular during the accounting panel) emphasized two points. (1) The use of QNE methods, in and of itself, does not bestow the ability to draw causal inferences. A recurring point at the conference was that, regardless of method, causal inferences rely on untestable core assumptions; if those assumptions are violated, causal inferences are not possible. (2) External pressure to make causal inferences can potentially create a circumstance where authors use QNE methods to make sweeping causal statements without an appreciation for the limitations of these methods (see related discussion in Gow, Larcker and Reiss, 2014). Indeed, it is our assessment that reliably establishing causality and generalizing those inferences to the broader population is much harder than the impression one would get from a perusal of the accounting literature.

In this section, we offer a more skeptical appraisal of QNE methods than the view often portrayed in the accounting literature. To be clear, the purpose of this section is not to suggest that such methods do not have advantages or should be used less frequently—in many cases they offer
clear advantages—but rather to highlight some conceptual limitations that appear to be underappreciated within the accounting literature.\textsuperscript{7} We organize this section around three different areas of discussion at the conference.

First, QNE methods are not a panacea to the challenges associated with drawing causal inference. To use the language of Larcker and Rusticus (2010), the use of difference-in-differences (DiD), instrumental variables (IV), and regression discontinuity designs (RDD) does not automatically “solve the endogeneity problem.” Each of these methods requires a theory that describes which shocks are exogenous, typically from detailed knowledge of a given institutional environment.

Second, evidence of a causal effect does not imply one has identified the causal mechanism(s) or causal channel. It is possible to empirically estimate a causal effect without understanding the economic mechanism that gives rise to the effect. This limitation is of course not unique to quasi-natural experiments and also applies to laboratory and field experiments. For example, it is possible to design a laboratory experiment to examine whether consuming red meat causes cancer in rats without understanding the underlying physiological mechanism.

Third, even under ideal conditions, the extent to which results obtained from QNE methods can be generalized to the broader population is often very limited. This limitation is also not unique to quasi-natural experiments; laboratory and field experiments are often conducted in non-random samples of the population. A quick perusal of the accounting literature suggests that causal effects estimated within a relatively specialized/unique (non-random) sample are simply assumed to generalize to the broader population without any discussion (see Armstrong, 2013 for a related

\textsuperscript{7} In the interest of parsimony, our discussion assumes some familiarity with the methods and their assumptions. We refer interested readers to Roberts and Whited (2013).
point). The extent to which one can generalize results obtained in a specific (non-random) sample to the broader population should be discussed, not assumed.

3.1 QNE methods are not a panacea

Estimates obtained from QNE methods do not necessarily have a causal interpretation. Close attention must be paid to whether the setting is appropriate and whether the untestable assumptions underlying the statistical method hold. For example, the exclusion restriction of IV, the parallel trends assumption of DiD, and the continuity assumption of RDD are all inherently untestable (e.g., Larcker and Rustics, 2010; Roberts and Whited, 2013). When these assumptions are violated, estimates of causal effects obtained from QNE methods are no more valid than those obtained from a regression of quantity on price.8

This reflects a recurring theme at the conference: identification does not come from statistical techniques but from untestable core assumptions. Violations of the assumptions underlying a researcher’s measure or empirical method confound attempts to draw causal inferences—and QNE methods are not immune to this concern. For this reason, it is important to understand the assumptions underlying both the empirical measures and the statistical method. We use three hypothetical examples to illustrate that the results obtained from using QNE methods do not necessarily have a causal interpretation.

Example #1. Consider the hypothetical example in Section 2.3. Suppose we want to study the causal effect of conservatism on the market reaction to earnings news, and suppose we identify an exogenous shock that affects the market reaction to earnings only through accounting

8 See also Reiss (this issue) and Rust (this issue).
conservatism, such that the setting appears to satisfy the assumptions underlying our chosen quasi-natural experimental method. If we were to use the asymmetric timeliness coefficient (ATC) as a measure of conservatism, would we be able to identify the causal effect of conservatism on the market reaction to earnings news?

Recall that the validity of the ATC as a measure of conservatism is premised on the notion that earnings do not convey incremental information to the capital market (incremental relative to what is already reflected in price). Thus, regardless of statistical method, a critical assumption underlying the measure is violated—even under ideal conditions, the use of the ATC as a measure of conservatism confounds attempts to estimate the causal effect of conservatism on the market reaction to earnings. The example illustrates that the assumptions underlying the empirical measure of the theoretical construct are just as important as the assumptions that underlie the statistical method employed.

*Example #2.* Suppose we want to study the causal effect of board independence on future firm performance. We are cognizant of the endogenous relation between board independence and firm performance. Accordingly, we search for a quasi-natural experimental setting. Suppose a government regulation that requires all firms to have a minimum of five independent directors is introduced. Can we use the change in directors required by the regulation as a quasi-exogenous shock to estimate the causal effect of board independence on future performance?

In order to estimate the causal effect of board independence on firm performance, we need a shock that induces as-if random variation in board independence. However, in this example, the
board is required to add $\text{MAX}\{0, 5-N\}$ independent directors, where $N$ is the number of independent directors prior to the regulation being introduced. As a result, the change in independent directors required by the regulation varies with the firm’s pre-existing number of independent directors. Consequently, if the number of independent directors before the regulation, $N$, is endogenous, then the effect of the regulation on the firm is also endogenous, i.e., it too depends on $N$. Note that a binary variable that indicates whether the regulation affects the firm is also endogenous—it too depends on the firm’s existing board composition. This example illustrates that while a regulatory change itself might be exogenous, the effect of the regulation on the firm is often not as-if random.

Example #3. Suppose we want to study the causal effect of management forecasts on investors’ uncertainty about firm value. Cognizant of the fact that management’s decision to issue a forecast is a choice variable, we employ a DiD design. Firms providing forecasts (treated firms) are matched based on industry and size to firms that do not provide forecasts (control firms). We then estimate a standard DiD design, and also estimate a DiD design that includes firm-fixed effects, and time-fixed effects. Plotting the level of uncertainty over time reveals that uncertainty for control firms does not change around the treated firms’ forecast dates, but uncertainty for treated firms falls markedly after the forecast dates—the graphical evidence suggest that parallel trends assumption appears to be satisfied. Have we identified the causal effect of voluntary disclosure on investors’ uncertainty?

While we employ state-of-the-art methods, and use many of the methods in the quasi-natural experimental toolkit, the variable used to assign firms to treated and control groups is likely endogenous. Firms endogenously chose whether to provide forecasts, and may very well do so
based on the expected effect of the forecast on uncertainty. For a DiD design to provide meaningful estimates of a causal effect, the variable used to assign firms into treatment and control groups must be *exogenous*. That is, to estimate a causal effect of voluntary disclosure on uncertainty, the shock (or instrument) must provide as-if random variation in whether management issues a forecast. In other words, we would have to assume that management’s decision to issue a forecast is independent of the forecast’s effect on uncertainty. This example illustrates that the core assumptions behind QNE methods are inherently untestable—state-of-the-art tests and diagnostics are not a substitute for as-if random variation.

3.2 *Estimation of causal effects versus identification of causal mechanisms*

Estimation of causal effects should not be confused with identification of causal mechanisms or causal channels. Consider the causal chain provided in Section 2.1. A hypothetical study estimating the causal effect of X on B, may find that X causes B. However, evidence of a causal effect of X on B, does not necessarily answer the question of “why” or “how” X causes B. In the example, X does not cause B directly. There are three degrees of separation; X causes B only because X causes Y, which causes Z, which causes B. These links in the chain would not be self-evident when estimating the causal effect of X on B. Thus, while one can use QNE methods and as-if random variation to estimate causal effects, evidence of a causal effect does not imply one has identified the economic mechanism that gives rise to the effect. We use two examples to illustrate this point.

Our first example is taken from the discussion surrounding Joshua Angrist’s conference presentation on the effect of charter schools takeovers on student achievement. By way of background, there is a large literature using quasi-natural experimental methods to estimate the
causal effect of charter schools on standardized test scores (see Abdulkadiroglu, Angrist, Hull, and Parhak, 2014 for a brief review). These studies find charter schools generally cause an increase in student test scores. Abdulkadiroglu, Angrist, Hull, and Parhak (2014) examine an instance in which a charter school takes over an existing public school. They find the takeover causes an increase in student test scores.

When a charter school takes over a public school, many changes occur simultaneously. There is likely substantial turnover in staff and teachers, and new discipline policies, attendance policies, curriculums, and incentives are introduced. The estimated causal effect represents the total effect of all of these changes. In this setting, knowledge of the causal mechanism can be of potentially greater importance than knowledge of the magnitude of the total causal effect. For example, if the causal mechanism were attendance or discipline policies—rather than teacher quality—presumably these policies could be replicated in non-charter schools without resorting to privatization.

Our second example is taken from the recent literature that uses exogenous variation in analyst coverage stemming from brokerage mergers and closures to estimate the causal effect of analyst coverage on firm outcomes. Kelly and Ljungqvist (2012, KL) is one of the earliest papers on this topic. Based on the model of Easley and O’Hara (2004), KL conjectures that an exogenous reduction in analyst coverage causes an increase in information asymmetry among investors, which in turn increases the risk premium that investors require to hold the firm’s shares. This, in turn, causes a decrease in prices. KL employs a difference-in-differences design and find that losing an analyst is associated with between a 2.0 and 2.4 basis point increase in the bid-ask spread (relative to a control sample), and short-window abnormal returns that range between −2.61% and −0.78%. Following KL, a large number of studies find that brokerage mergers are associated with
various firm outcomes including an increase in voluntary disclosure; a reduction in investment and financing; a reduction in reporting quality; an increase in earnings management; and an increase in agency problems.\textsuperscript{9}

Thus, as in the case with charter school takeovers, brokerage mergers affect a number of different firm outcomes. Classical asset pricing theory suggests that there are two potential channels through which these disparate effects might manifest in share prices: changes in discount rates, i.e., changes in risk, and changes in expected future cash flows.\textsuperscript{10} On the one hand, a reduction in analyst coverage might affect share prices by altering investors’ assessments of risk. This channel does not rely on agency problems, but instead relies on information asymmetry being a priced risk. On the other hand, a reduction in analyst coverage might affect prices by altering investors’ assessments of expected cash flows. This channel does not rely on information asymmetry being a priced risk, but instead relies on a reduction in monitoring leading to an increase in agency problems. Of course, both channels could be at work, and one does not need to distinguish between these two channels to estimate the causal effect of brokerage mergers/closures on prices—only to answer the question of why brokerage mergers/closures cause a change in prices.

These two examples illustrate that estimation of causal effects is not synonymous with identification of causal channels. Evidence of a causal effect does not imply one has identified the economic force(s) that gave rise to that effect. For this reason, estimation of causal effects should not be an end, in and of itself.


\textsuperscript{10} Lambert, Leuz, and Verrecchia (2007) refer to these as the direct and indirect effects of information, respectively.
3.3 Generalizability of results obtained from QNEs

Across many of the presentations and panels, conference participants expressed concerns about the extent to which one could generalize results obtained from application of QNE methods in specialized/unique (non-random) samples to the broader population. While there is a premium on results that are both generalizable and that allow for causal inferences, it is exceedingly rare to find a setting that facilitates both causal and generalizable inferences. Indeed, use of QNE methods often requires that one sacrifice generalizability. Rather than repeat the discussion at the conference, we point to Christian Leuz’s presentation which offers a summary of these concerns:

“Often, careful identification comes at the expense of external validity or generalizability. I am exaggerating but some of the best natural experiments provide essentially examples, in which a causal effect exists and has a certain magnitude. Put differently, the setting allows for the identification of what are often very local effects that may not generalize much beyond the setting.”

To illustrate these points, consider the following two studies.

Within the literature on the capital market effects of recognition versus disclosure, Michels (2014) offers an ideal natural experiment. Michels (2014) notes that the timing of natural disasters dictates whether firm must recognize or disclose the consequences of the disaster. Because the timing of natural disasters is random, the study offers a clean setting with which to estimate the causal effect of recognition on capital markets. Yet, by design, the tests are limited to unusual, extremely rare events. Consequently, it is unclear whether the causal effects estimated in this setting are generalizable outside of the sample to the broader population (for example recognition of a recurring item).

While concerns about generalizability arise regardless of the choice of statistical method, it seems that such concerns are particularly acute in QNE methods. Due to their design, QNE

methods frequently estimate local average treatment effects. Local effects may not be representative of a causal effect in the broader population. Consequently, the absence of a result from a QNE method does not imply the absence of a causal effect.

Consider the empirical experiment in Iliev (2010). Iliev examines the effect of SOX on various firm outcomes (e.g., audit fees, earnings management, and changes in firm value) using a RDD design around the $75 million public float threshold. At the risk of simplifying the design, suppose we compared firms just below the threshold, $74-$75 million, to firms just above the threshold, $75-$76 million, and that all of the assumptions that underlie RDD are met. Now suppose we found that firms with $74-$75 million float have different outcomes than firms with $75-$76 million float. We would (rightly) conclude that SOX has a causal effect on these outcomes, or alternatively, a causal effect exists. However, suppose we had found that there were no differences in firm outcomes between these two groups of firms. Would we conclude that SOX has no effect; that a causal effect does not exist? No. The inference is that—in this very tiny sliver of the population of firms, those just to the left and right of the public float threshold—SOX does not have an effect. In other words, the absence of a causal effect using a QNE design does not imply the absence of an effect in a different sample of firms. An effect could exist, and it could be economically significant, but in a different sample of firms. Perhaps SOX affected firms with $1 billion in float, such that we cannot generalize from our findings regarding firms between $74 and $76 million in float.

We are concerned that, in practice, because of the specialized/unique samples required for QNEs, there will be a tendency to dismiss counterintuitive QNE results as idiosyncratic—a symptom of the setting—rather to update beliefs about the population. In this manner, QNEs that produce results that do not comport with prior beliefs may tend to be dismissed as not
generalizable, whereas QNEs that confirm prior beliefs will tend to be interpreted as generalizable to the broader population. Generalizability is an inherent trait that is determined a priori as a result of the empirical design. Being determined a priori, generalizability does not depend on whether the results comport with one’s priors. Consequently, if the ability to generalize results from QNEs depends on the results (and whether they comport with the researchers prior) then the results are, in truth, not generalizable.

Some of these concerns stem from how QNEs are currently used in the accounting literature. First, results obtained from specific QNEs (e.g., a specific regulation) can potentially be informative even if the inferences do not generalize. In this regard, QNEs created by regulatory shifts can potentially provide useful answers to policy related questions, even in the absence of generalizability (e.g., Leuz and Wysocki, 2015). Generalizability only becomes a concern when QNEs are used as a tool to answer broader questions about the population. Second, it is often the case that a single QNE is presented as a stand-alone study. One approach to address concerns about generalizability is to offer evidence on multiple QNEs within the context of a single study, or to combine evidence from one or more QNEs with large-sample descriptive evidence that is consistent with a causal mechanism. The more evidence that is brought to bear across different samples, the stronger the case for generalizability.

The message here is not against accepting evidence from QNEs; on the contrary, we advocate for a more open reception of results from QNEs, whatever they may be, but with a clear delimitation of their scope of application. Future accounting research employing QNEs should pay close attention to whether the setting is appropriate and the untestable assumptions underlying the particular method hold; recognize that the estimation of causal effects and identification of causal
mechanisms are distinct tasks; and discuss the extent to which one can generalize results obtained in a specialized/unique (non-random) sample to the broader population.

4. Structural Estimation

4.1 Overview

Interestingly, almost all plenary speakers viewed themselves as performing some form of structural estimation—economic theory guided the structure of their statistical tests and how they interpreted the estimated correlations. This contrasts with the limited popularity of structural estimation in accounting research. For example, while Gow, Larcker, and Reiss (2014) advocate for increasing use of structural estimation within accounting and apply the technique to an interesting accounting question (e.g., the relation between equity incentives and accounting restatements), they note that the existing body of work employing structural estimation is extremely sparse.12

We believe this is partly due to the erroneous perception among many accounting researchers that structural estimation is necessarily characterized by complex math, non-linear solvers, and cumbersome numerical optimization routines. Complexity is not a necessary characteristic of structural estimation, nor is it always desirable. Excessive complexity can obscure intuition and be counter-productive (see Welch, this issue).

As we discuss, and indeed will show, structural estimation can be done in a manner accessible to anyone with no more mathematical savvy than what is sufficient to understand ordinary least squares (OLS), and yet provide helpful insights about important questions in accounting.

---

accounting. Even a simple OLS model (returns on market returns) can yield a structural estimate of risk (e.g., CAPM Beta). In this regard, it is important to point out that structural estimation is more prevalent in accounting than is commonly believed. CAPM Beta, Kyle’s $\lambda$, and PIN are commonly used in accounting, and all are structural estimates—estimates of specific parameters appearing in specific theory models. It is thus instructive to consider a few common traits shared by these estimates.

First, many of these measures are straightforward to estimate and involve no more than a simple line of algebra or an OLS regression. A common technique for estimating a firm’s Beta is to regress the firm’s return on the market return. Similarly, a common technique for estimating Kyle’s $\lambda$ is to regress the change in price between trades on signed order flow. Structural estimation is an empirical method, and will likely be most successful when it is kept simple. Of course, we recognize the value of richer, more complex models when the additional complexity better describes observed behavior. We view the engineering of structural models as an incremental process that begins with simple models that allow us to identify basic patterns in the data. Only then can we evaluate the cost and benefits of additional complexity—math for its own sake is not desirable (Romer, 2015).

Second, none of these measures is particularly intuitive in the absence of the corresponding economic theory and assumptions. For example, in the absence of the knowledge gleaned from the CAPM, it seems counterintuitive—indeed naive—to suggest that the covariance between a firm’s return and the market return is not only a measure of risk, but the only measure of priced risk. Economic theory provides the assumptions that help guide how we interpret the relations in the data—for example, how we interpret the covariance between a firm’s returns and the market return, and how we can use intra-day order flow and trade timing to recover the extent of privately
informed trade. One of the primary benefits of the structural estimation is that it allows researchers to formalize their assumptions and uncover/recover latent variables that would otherwise not be observable to the researcher. By tying unobserved variables to observed variables, structural estimation can exploit exogenous shocks in latent unobserved variables (e.g., shocks to preferences, business cycles or corporate information environment) and thus does not rely on exogenous shocks to observed variables to the same extent that quasi-experimental methods do.

Third, structural estimation requires that one take the underlying theory literally. This is considerably more demanding than the more common approach of using theory to motivate intuition for empirical tests. For example, the formula for PIN is a relatively complicated non-linear function of different aspects of intra-day order flow (Easley, Kiefer, and O’Hara, 1997). Specifying a precise functional form is much more demanding than a directional prediction (e.g., that informed trade is increasing in a particular measurable aspect of order flow). While the specific assumptions underlying any particular theory model are almost surely violated in practice, the model can nevertheless provide a useful abstraction for how to think about the world. Indeed, despite tenuous assumptions, CAPM Beta, Kyle’s λ, and PIN have guided decades of research—suggesting a usefulness for theory-based measures even when underlying assumptions are violated. Even if we do not believe the model holds literally—and thus the estimate is a noisy measure of the parameter of interest—we can ask whether the estimates are empirically descriptive. In the end, while the assumptions that underlie the models do not hold, parameters derived under those assumptions can still help frame how we think about related issues.

Finally, each of these measures—and structural estimation more generally—allows the researcher to conduct counterfactual analysis or analyze “what if” scenarios. For example, given the CAPM and an estimate of CAPM Beta, we can say precisely how much the firm’s cost of
capital would change if the market risk premium increased from 7% to 15%. Such counterfactual analyses can be particularly useful for policy related questions. For example, how much would consumer surplus and audit fees change if auditor rotation was mandatory or if one of the Big 4 auditors was to exit the industry? (e.g., Gerakos and Syverson, 2015) The only alternative to this approach is to analyze historical data for a closely related analog to the question at hand (for example, empirically estimate the change in audit fees after Andersen’s collapse). This is problematic however, as the historical data may have been generated by a different regulatory/audit environment.13

4.2 Application—Measuring Disclosure Costs

To illustrate some of the key points of the conference, we structurally estimate the parameter for voluntary disclosure costs in Verrecchia (1983), hereafter V83. We show that V83 implies a parsimonious, theoretically-grounded measure of voluntary disclosure costs that is readily estimable from archival data. This example serves to illustrate that many existing economic models have exactly what it takes to support meaningful empirical exercises, with only minimal further derivation needed.

V83 characterizes a voluntary disclosure threshold as a threshold above which the manager will choose to disclose private information, and incur the costs of voluntary disclosure. V83 formalizes the notion of disclosure costs as an explanation for the strategic withholding of information.14 Even though the model has been around for more than 30 years and is widely known

---

13 The “Lucas (1976) critique” suggests that if we want to quantify the effect of a policy experiment, we should model and estimate, the choice problem of economic agents. If the model can account for observed empirical regularities, we can use the model to predict how economic agents will respond to a change in policy.

14 A common misconception in empirical studies is that the cost parameter in Verrecchia (1983) necessarily stems from disclosure of proprietary information. The cost parameter that appears in Verrecchia (1983) is not exclusive to proprietary costs—it includes all rationally anticipated costs of disclosure to shareholders regardless of source (e.g.,
in both the empirical and theoretical literatures on disclosure, there have been no attempt to estimate the underlying parameters of the model. As a result, we do not know the answer to basic questions such as: How high are disclosure costs? Can disclosure costs provide an explanation for managers’ disclosure behavior? How much less often would we expect to observe forecasts if disclosure costs were to increase by, say, ten percent? To make causal statements concerning disclosure costs, we need a measures of such costs. This application serves as a first step toward deriving a theoretically grounded measures of such costs. We leave it to the reader to judge the pedagogical value of this exercise and future research to judge whether the measure of disclosure costs is empirically descriptive.

4.2.1 Setup of Verrecchia (1983) and Identification of Disclosure Costs

We adopt the assumptions and notation from V83. The timeline and setup of V83 is as follows. First, a manager is endowed with private information concerning the firm’s earnings. The manager maximizes stock price, and decides to disclose the information on the basis of how disclosure will affect the firm’s stock price. Second, investors observe either the disclosure or the absence of disclosure, and update their beliefs about earnings. Third, earnings are realized and paid out as a liquidating dividend.

In order to keep the estimation straightforward, we simplify V83 along two dimensions. First, we assume that the manager and investors are risk-neutral. This assumption equates price maximization with the maximization of investors’ expectation of earnings. Second, we assume the manager knows earnings prior to making the voluntary disclosure decision. This assumption equates the distribution of the manager’s private information with the distribution of earnings.
These assumptions are not necessary for our analysis, but considerably simplify the estimation of disclosure costs.\(^{15}\)

Let \(\tilde{u}\) denote the firm’s earnings gross of any potential disclosure costs \(c\), where \(\tilde{u}\) is normally distributed with mean \(y_0\) and variance \(\sigma_h^2\) (\(\sigma_h^2 = h_0^{-1}\) in V83); we use a tilde, \(\sim\), to denote a random variable. \(y_0\) represents investors’ ex ante expectation of earnings before the manager has the opportunity to disclose his information, and \(\sigma_h^2\) represents investors’ ex ante uncertainty about earnings. If the manager chooses to voluntarily disclose his private information to the capital market, the firm’s earnings are reduced by the disclosure cost \(c\).

Let \(x\) denote the disclosure threshold such that the manager discloses his private information if and only if earnings exceed the threshold, i.e., \(u > x\). The manager prefers to disclose his information if investors’ beliefs about earnings, net of disclosure costs, are higher following disclosure than they would be in the absence of disclosure, i.e., \(E[\tilde{u} \mid u > x] - c > E[\tilde{u} \mid u \leq x]\).

The disclosure threshold \(x\) is given by eqn. (9) of V83,

\[
x = y_0 + \left( c - \frac{\sigma_h g(x)}{G(x)} \right) \tag{1}
\]

where \(g(x)\) and \(G(x)\) are the probability density function and cumulative density function of earnings. Note that we can rearrange eqn. (1) to express the threshold in terms of standardized unexpected earnings

\[
\left( \frac{x - y_0}{\sigma_h} \right) = \left( \frac{c}{\sigma_h} - \frac{\sigma_h g(x)}{G(x)} \right) \tag{2}
\]

\(^{15}\) In the notation of V83, risk neutral managers and investors implies \(\beta = 0\) and perfect observability implies \(s \rightarrow \infty\).
and solve for the disclosure cost,

\[ c_{std} = x_{SUE} + \frac{\phi(x_{SUE})}{\Phi(x_{SUE})}, \tag{3} \]

where \( c_{std} \) is the standardized disclosure cost, i.e., \( c_{std} = \frac{c}{\sigma_h} \), \( x_{SUE} \) is the disclosure threshold expressed in terms of standardized unexpected earnings, i.e., \( \left( \frac{x - y_0}{\sigma_h} \right) \), and \( \phi \) and \( \Phi \) are the probability density function and cumulative density function of the standard normal distribution, respectively.

Next, let \( q \) denote the probability of observing a management forecast, i.e., the probability that standardized unexpected earnings are above the threshold,

\[ q = \Pr(SUE > x_{SUE}) = 1 - \Phi(x_{SUE}), \tag{4} \]

where \( SUE = \frac{\tilde{u} - y_0}{\sigma_h} \). Substituting into eqn. (3), we obtain a relation between the standardized disclosure cost, \( c_{std} \), and the probability of observing a forecast, \( q \),

\[ c_{std} = \Phi^{-1}(1 - q) + \frac{\phi(\Phi^{-1}(1 - q))}{1 - q}. \tag{5} \]

Using this equation, the standardized disclosure cost can be consistently estimated using the sample frequency of observing a forecast, \( \hat{q} \).

### 4.2.2 Data and Estimation Results

To empirically estimate disclosure costs, we require data on quarterly earnings-per-share from I/B/E/S for 2004-2014. We restrict attention to those firms that can be matched to CRSP and Compustat and remove firms with less than twenty quarters of data. We collect, but do not require, data on management earnings forecasts for each firm-quarter from I/B/E/S, and we remove those
firms that never provide a forecast.\textsuperscript{16} Since we estimate the model at the firm-level, it is important that the distribution of earnings is stationary. As a result, we do not use a time-series of split-adjusted EPS whose variance decreases as one goes back in time and instead use raw, unadjusted EPS numbers (e.g., Payne and Thomas, 2003). The sample consists of 58,778 firm-quarters and 1,459 unique firms which issued at least one forecast over the period 2004-2014.

For each firm $i$, we summarize the firm’s management forecasts with two statistics. First, we compute the empirical frequency of observing a forecast over the sample period, $\hat{q}$, as the fraction of quarters where managers provide at least one earnings forecast. Second, we compute the standard deviation of the firm’s unexpected earnings over the sample period, $\hat{\sigma}_h$, using lagged EPS to measure investors’ ex ante expectation of earnings before the manager has the opportunity to disclose. Panel A of Table 1 provides summary statistics for $\hat{q}$, and $\hat{\sigma}_h$.

We then use $\hat{q}$ and eqn. (5) to estimate the firm’s disclosure costs in terms of standardized earnings, $\hat{c}_{std}$, and multiply by $\hat{\sigma}_h$ to obtain disclosure costs in terms of earnings per share, $\hat{c}$. Panel B of Table 1 provides summary statistics for our two estimates of disclosure costs. For the median firm, the estimates of disclosure costs amount to a reduction in earnings-per-share of $0.18, or 1.13 standard deviations.

Next, we conduct a counterfactual analysis. Specifically, we address the remaining question we offered as motivation for this exercise: How much less often would we expect to observe forecasts if disclosure costs were to increase by, say, ten percent? Conducting this analysis requires inverting the process. Given $\hat{c}_{std}$ we can invert eqn. (5) to solve for $q$: specifically, we set

\textsuperscript{16} In the absence of at least one voluntary disclosure over the estimation window, disclosure costs are estimated to be infinite.
\( c_{std} \) in eqn. (5) equal to \( 1.10 \times \hat{c}_{std} \) and solve for \( \hat{q}_{10\%} \). Panel C of Table 1 presents the results. For a 10\% increase in disclosure costs, we find the probability of observing a management forecast for the average (median) firm is reduced from 0.37 to 0.34.

A natural implication of the model is that the manager does not disclose earnings that are less than the disclosure threshold, \( x \). Conditional on disclosure, earnings are net of disclosure costs, i.e., earnings are given by \( u - c \). Thus, any forecasts that are below the threshold net of costs, \( x - c \), are inconsistent with the model. Panel C of Table 1 provides summary statistics for the frequency of forecasts that are inconsistent with the estimated thresholds. \( \text{ThresholdViolation} \) is an indicator variable equal to one if a management forecast is below the theoretical threshold. We find that about 14\% of management forecasts are lower than the disclosure threshold implied by the estimates of disclosure costs from V83. While this is clearly inconsistent with the model, we argue that requiring exactly zero forecasts to fall below the estimated threshold—that no firm or forecasts violates V83—is too stringent of a test. We do not expect V83 to hold exactly, and we expect measurement error in each of \( \hat{q} \), \( \hat{\sigma}_h \), \( \hat{c}_{std} \), and \( \hat{c} \). In other words, whether the 14\% is considered “large” or “small” depends on what researchers do with the structural estimate of disclosure costs.

As an analogy, consider the \( R^2 \) statistic from a linear regression model. Whether the goodness of fit of the regression is a concern depends on the main purpose of the regression. In the asset pricing literature \( R^2 \) is one of the key statistics to judge whether the proposed set of factors can explain cross-sectional expected returns; the objective is explanatory power. In many other literatures however, the objective is to identify whether a particular regressor (\( X \)) is correlated with, or alternatively causes, the dependent variable (\( Y \)). In this case empirical researchers may not be concerned with a low goodness of fit, and a perfectly specified model (\( R^2 = 1.00 \)) is not a necessary
condition to draw meaningful inferences. In our context, if one wants to argue that V83 is a good model to explain why some firms disclose and others do not, then 14% may appear quite large. However, if the purpose is to use the estimates as a proxy for disclosure costs to test other economic hypotheses (e.g., whether disclosure costs are higher in more competitive or more litigious industries/times) then perhaps 14% is acceptable—as it indicates some noise in the proxy. We leave it to future research to judge whether the measure of disclosure costs is empirically descriptive.

In addition to some preliminary estimates of disclosure costs and counterfactual analysis, this example also illustrated the various choices we had to make. We relied on moments rather than maximum likelihood to estimate the model parameters; we restricted the analysis to quarterly earnings forecasts but did not impose any other restrictions such as forecasts not being bundled with previous quarters’ earnings announcements; we assumed the distribution of earnings was stationary; we assumed managers and investors were risk neutral; and we assumed managers had perfect information. Thus, while this example provides—to the best of our knowledge—the first estimates of disclosure costs as implied by V83, it also illustrates the need for subsequent papers that use different samples, different estimation techniques, different modeling assumptions, and examine construct validity. To summarize, having a structural estimate in and of itself is not necessarily an achievement; rather it is how that estimate is used that has the potential to provide new insights.

5. Conclusion
In this paper, we report on and synthesize the proceedings of the 2014 Stanford Causality Conference as relevant for accounting researchers. We illustrate the themes of the conference using examples drawn from accounting literature and provide a brief application.

We begin by discussing the role of theory in informing empirical research that seeks to draw causal inferences. A sound, well-defined theory is necessary to draw causal inferences from observational data. Theory provides a framework for making predictions and interpreting estimated correlations. We distinguish between three types of theory, verbal, formal statistical, and formal economic, and discuss the advantages and disadvantages of each, as it relates to informing empirical work and facilitating identification of causal channels. We offer a skeptical perspective on attempts to draw causal inferences in the absence of well-defined constructs and assumptions. Generally, it seems difficult to credibly identify the causal effect of an intervention (e.g., regulation) on a theoretical construct, if that construct is only defined at an intuitive level, or if we do not understand the assumptions that underlie our empirical measures of the construct.

Next, we highlight some of the conference discussion surrounding the limitations of statistical methods known as quasi-natural experimental, or QNE, methods (e.g., difference-in-differences, instrumental variables, regression discontinuity). We frame our discussion around three concerns that, in our assessment, are underappreciated within the accounting literature: The importance of assumptions. The use of QNE methods does not bestow the ability to draw causal inferences. A recurring point at the conference was that, regardless of method, causal inferences rely on untestable core assumptions. When these assumptions are violated, estimates of causal effects obtained from QNE methods are no more valid than those obtained from a regression of quantity on price. For this reason, it is important to understand the assumptions underlying the statistical method.
Identification and estimation are distinct. While one can use QNE methods and as-if random variation to estimate causal effects, estimation of causal effects is not synonymous with identification of causal channels. It is possible to empirically estimate a causal effect without understanding, or identifying, the economic mechanism that gives rise to the effect. For this reason, estimation of causal effects should not be an end, in and of itself.

Limited generalizability. Even under ideal conditions, the extent to which results obtained from QNE methods can be generalized to the broader population is often very limited. A quick perusal of the accounting literature suggest that causal effects estimated within a relatively specialized/unique (non-random) sample are often asserted to generalize without any discussion. Generalizability should be discussed, not asserted.

Finally, we discuss the role of structural estimation as an emerging method within accounting research and provide a simple application that illustrates many of the themes discussed at the conference. Specifically, we structurally estimate the parameter for disclosure costs that appears in Verrecchia (1983). We show that this model suggest a parsimonious measure of voluntary disclosure costs that is readily estimable from archival data. This example should serve to illustrate that many existing economic models have exactly what it takes to support useful empirical exercises, with only minimal further derivation needed. We leave it to future research to judge whether the measure of disclosure costs is empirically descriptive.

To conclude, there is a growing interest within accounting research to embrace new methodologies that facilitate causal inferences. Attempts to push the frontiers of accounting research, and move toward causal inferences are desirable—but must be undertaken with care. We caution against methodological extremism—the notion that one type of theory or method is unambiguously superior. Providing generalizable causal inferences is very difficult, and usually
not feasible within the context a single study—there are very few natural experiments and even fewer that are generalizable.

As a practical matter, we interpret the conference discussion as suggesting the following minimum criteria for empirical work that investigates causal mechanisms.

(1) Predictions and test need to be motivated by economic theory. Researchers interested in investigating causal mechanisms must understand and state (or reference) the assumptions of the underlying theory that motivates the analysis, and those assumptions must be internally consistent. In this regard, linking empirical analyses to formal theory may prove particularly helpful. Formal theory—the processes of formalizing and understanding theoretical assumptions—assures consistency both in logic and in measurement, and makes transparent any inconsistencies.

(2) Explicitly state empirical assumptions and how they relate to the underlying theory—distinguishing theoretical constructs from empirical measures. Researchers interested in investigating causal mechanisms must not only understand the underlying theoretical assumptions but also: (a) the assumptions that underlie empirical proxies for the theoretical constructs of interest; (b) the assumptions that underlie the specific statistical method/model being used to estimate correlations in the data (e.g., exogeneity, parallel trends, exclusion restriction); and (c) the assumptions needed to generalize inferences obtained within a non-random sample to the broader population. Careful thought is necessary to ensure the empirical specification and associated assumptions are faithful to the motivating theory.

(3) Rule out specific alternative explanations. Such explanations could be statistical in nature (e.g., a specific form of measurement bias) or economic in nature (e.g., a specific
causal channel predicted by an alternative theory). The often heard “endogeneity critique” is not a specific alternative; endogeneity is so broad and can take on so many different forms that it is not falsifiable. However, endogeneity as the outcome of a specific alternative statistical process or economic process can be falsified. In this regard, quasi-natural experimental methods may prove particularly useful.

It is our assessment that providing reliable evidence of causal mechanisms—evidence that provides a foundation for future work—is less about finding the next natural experiment or using a specific statistical method, and more about matching theory with empirical tests, carefully articulating all theoretical and empirical assumptions, and ruling out alternatives.
References


Table 1. Summary statistics and disclosure costs estimates

Panel A presents summary statistics for variables used in our analysis. \( \hat{q} \) is the fraction of quarters where managers provided at least one earnings forecast, and \( \hat{\sigma}_h \) is the standard deviation of the firm’s unexpected earnings. Panel B presents summary statistics our estimates of the disclosure cost parameter in Verrecchia (1983). \( \hat{c}_{\text{std}} \) is the firm’s estimated disclosure cost in terms of standardized earnings and \( \hat{c} \) is the firm’s estimated disclosure cost in terms of earnings-per-share. Panel C presents results from calculating the frequency of voluntary disclosure, \( \hat{q} \), under a 10% increase in disclosure costs. Panel D presents frequency of management forecasts that violate the disclosure threshold. \( \text{ThresholdViolation} \) is an indicator variable equal to one if a management forecast is below the theoretical threshold. Sample of 1,459 unique firms over the period 2004-2014.

Panel A. Summary statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of quarters with at least one forecasts</td>
<td>( \hat{q} )</td>
<td>0.37</td>
<td>0.23</td>
</tr>
<tr>
<td>Standard deviation of unexpected earnings</td>
<td>( \hat{\sigma}_h )</td>
<td>0.23</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Panel B. Estimates of disclosure costs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disclosure cost–standardized earnings per share</td>
<td>( \hat{c}_{\text{std}} )</td>
<td>1.17</td>
<td>1.13</td>
</tr>
<tr>
<td>Disclosure cost–earnings per share</td>
<td>( \hat{c} )</td>
<td>0.38</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Panel C. Counterfactual analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of quarters with at least one forecast for a 10% increase in standardized disclosure cost</td>
<td>( \hat{q}_{10%} )</td>
<td>0.34</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Panel D. Frequency of forecasts below theoretical threshold

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{ThresholdViolation} )</td>
<td>( \Pr(\bar{u} \leq x) )</td>
<td>0.14</td>
<td>0.00</td>
</tr>
</tbody>
</table>