



Theory, research design assumptions, and causal inferences

Christopher S. Armstrong*, John D. Kepler

The Wharton School, University of Pennsylvania, 1300 Steinberg-Dietrich Hall, Philadelphia, PA 19104-6365, United States

ARTICLE INFO

Keywords:

Causal inference
Quasi-natural experiments
Disclosure theory

JEL classification:

C10
C50
M40
M41

ABSTRACT

Ferri, Zheng, and Zou test Fischer and Verrecchia's (2000) prediction that a reduction in investors' uncertainty about managers' financial reporting objectives leads to an increase in the valuation-relevance of earnings reports. They use mandatory CD&A disclosures as an arguably exogenous "shock" that provided investors with more precise information about managers' contractual incentives and find that these enhanced disclosures increased the relation between firms' unexpected earnings and stock returns. Using Ferri et al. as a backdrop, we discuss the implicit assumptions invoked in natural experimental research designs and the fundamental role of theory in drawing causal inferences from empirical evidence.

1. Introduction

Ferri, Zheng, and Zou (this issue) seek to document evidence about whether investors' uncertainty about managers' financial reporting incentives has a causal effect on the relation between their firm's unexpected earnings and stock returns. The authors motivate this research question based on the theoretical model of Fischer and Verrecchia (2000) in which a manager privately observes the firm's actual earnings and subsequently issues a potentially biased earnings report that investors use to price the firm's shares. Fischer and Verrecchia show that the more investors know about the manager's financial reporting incentives (e.g., details of the manager's compensation contract), the better they are able to "back out" any bias in the manager's earnings report so that it becomes a more informative signal of firm value. Ferri et al. use the 2006 compensation discussion and analyses (CD&A) disclosure rules implemented by the U.S. Securities and Exchange Commission (SEC) as a "shock" or instrument that arguably (i) lead to a reduction in investors' uncertainty about managers' financial reporting incentives (based on the additional details provided about managers' compensation contracts), and (ii) was exogenous with respect to the relation between firms' unexpected earnings and stock returns measured using earnings response coefficients (ERCs).

At first glance, this research setting appears to be well suited for testing the theoretical predictions of Fischer and Verrecchia (2000). However, as we explain below, there are several related theories that might better capture the salient features surrounding the 2006 rule change. To the extent that a different theory provides a more accurate description of the key features of Ferri et al.'s research setting, their empirical results may lead to considerably different inferences. As we explain, it is not clear whether Ferri et al.'s research setting "speaks to"—and therefore constitutes a proper test of—Fischer and Verrecchia (2000) rather than some alternative theory.

Nevertheless, we commend the authors for their combination of sharp empirical predictions derived from a specific theoretical model that seems to comport closely with their quasi-experimental research setting. Although we may not necessarily agree with all of their inferences, their paper is salutary example of how to generate credible causal inferences from observational data interpreted

* Corresponding author.

E-mail addresses: carms@wharton.upenn.edu (C.S. Armstrong), jkepler@wharton.upenn.edu (J.D. Kepler).

through the lens of a sound theory that captures the salient features of the research setting. It also provides an opportunity to discuss a number of subtle, but important issues that frequently arise in studies that seek to draw causal inferences.

Our discussion is organized in four main sections. First, we discuss the theory that Ferri et al. aim to test—i.e., [Fischer and Verrecchia, 2000](#)—as well as several related, alternative theories that also offer plausible explanations of Ferri et al.'s empirical findings. We argue that it is important to ensure that there is a sufficiently close correspondence between any empirical study's research setting—and the associated research design—and the theory that it purports to test. Thus, in the context of Ferri et al., to the extent that another theory more accurately captures the economic forces surrounding the 2006 CD&A disclosures, then Ferri et al. might be better viewed as a test of that theory.¹

Second, we assess how these alternative theories capture the salient features of Ferri et al.'s research setting that features the 2006 CD&A disclosure rules. We also discuss several institutional features of this setting that seem to have important implications for the authors' study and, in particular, the nature and credibility of their inferences.

Third, we discuss several issues that arise in Ferri et al.'s research design that are common in the broader literature. We explain how drawing causal inferences in quasi-experimental research settings relies on more than simply satisfying the well-known “parallel trends” assumption. It is also important to consider the less well known—or, at least, less commonly acknowledged—Stable Unit Treatment Value Assumption (or SUTVA), which has two components: (i) no interferences between units, and (ii) homogeneous treatment. We also discuss several examples of how Ferri et al.'s research design seems to condition on “post-treatment” outcomes. We explain how doing so can confound causal inferences and discuss how the use of theory to guide the development of empirical research designs and model specifications can preclude inadvertent conditioning on post-treatment outcomes. The subtle nature of these two issues should be viewed as an encouraging sign of the literature's progress in addressing causality in a thoughtful and rigorous manner.

We conclude with a brief discussion of the fundamental role of theory in drawing causal inferences from empirical evidence. We argue that causal inference is a form of knowledge creation that seeks to develop and test (i.e., falsify) explanatory theories, which are conjectures—that may take the form of a formal, analytical model, as in Ferri et al.'s use of [Fischer and Verrecchia \(2000\)](#), but need not—about the relationships between phenomena of interest (e.g., investors' precision about managers' reporting incentives and firms' earnings and returns). Argument and experimentation serve to identify errors in the theories so that they may be refined, modified, or ultimately abandoned in favor of better explanatory theories.

2. Alternative theories

2.1. [Fischer and Verrecchia \(2000\)](#)

[Fischer and Verrecchia \(2000\)](#) consider a setting in which a risk-neutral manager has an exogenously-specified contract that is a function of the firm's stock price. The manager privately observes the firm's true earnings—i.e., a noisy signal of firm value—and subsequently issues a public earnings report that risk-neutral investors use to price the firm's shares in a rational expectations equilibrium (REE). By incurring a personal cost, the manager can introduce *reporting bias*, which causes reported earnings to diverge from true earnings. An important feature of this setting is that investors have uncertainty about the manager's incentives to bias earnings, which prevents them from perfectly anticipating—and therefore “backing out”—any reporting bias. As the quality (or precision) of investors' information about the manager's incentives improves, they know more about the manager's intentions, and can better anticipate (and “back out”) any reporting bias. In the limiting case where investors have perfect information about the manager's reporting incentives, they can perfectly see through any reporting bias and recover the firm's true earnings as privately observed by the manager.

Given this setup, [Fischer and Verrecchia \(2000\)](#) derive two key results. First, a reduction in investors' uncertainty about the manager's reporting incentives leads to a higher earnings response coefficient (ERC). The intuition for this result is that since investors are better able to “back out” reporting bias in higher quality information environments, they place greater weight on reported earnings when setting price. Second, a reduction in investors' uncertainty about the manager's reporting incentives leads to greater reporting bias. The intuition for this result is that the manager stands to benefit more from inflating earnings when investors place more weight on earnings.

2.2. [Peng and Röell \(2008, 2014\)](#)

[Peng and Röell \(2008, 2014\)](#) generalize [Fischer and Verrecchia \(2000\)](#) by introducing a standard moral hazard agency problem so that the manager not only issues a (potentially biased) earnings report, but also takes a “real” productive action that is (imperfectly)

¹ It is important to note that a particular theory or model obviously does not have to describe all of the features of the research setting. However, it should at least capture those that are first-order and the empirical study's research design should be able to credibly “abstract away from” (e.g., control for) the features of the research setting that are not described by the theory. [Pearl \(2009, p. 420\)](#) explains this relationship between theory and research design as follows: “[i]n most cases the scientist carves a piece from the universe and proclaims that piece *in*—namely, the focus of investigation. The rest of the universe is then considered *out* or background and is summarized by what we call boundary conditions... In economics, circumscription amounts to deciding which variables are deemed endogenous and which exogenous, in the model or external to the model” (emphasis added).

reflected in earnings. Therefore, the manager now has two ways—both of which are personally costly—to influence earnings and, in turn, his contractual payment. Importantly, the manager's compensation contract—which provides incentives to take the productive action and to introduce bias in reported earnings—is *endogenous* and is the solution to the principal's contract design problem.²

Although this model loses much of the parsimony of Fischer and Verrecchia (2000), it is arguably more realistic in that the reason the manager's compensation is tied to stock price is to provide the manager with incentives to take a “real” productive action. However, a second-order consequence that accompanies this first-order incentive consideration is that the manager's compensation contract can also encourage reporting bias. This characterization is consistent with the expansive literature in accounting and finance that examine whether managers' contractual incentives (e.g., stock and option holdings, earnings-based bonuses, the anticipation of new stock and option grants, etc.) encourage managers to bias earnings (e.g., Armstrong et al., 2010; Armstrong et al., 2013). Studies in this literature tend to characterize financial misreporting as an unintended, but presumably tolerable, by-product of tying managers' compensation to stock price, earnings, and other potentially manipulable performance measures. In contrast, since it is not the focus of their paper, Fischer and Verrecchia (2000) abstract away from managers' productive actions to focus on managers' financial reporting decisions.

Peng and Röell (2008, 2014) find that an increase in investors' uncertainty about the manager's incentives usually calls for the manager's compensation to be *more* sensitive to stock price. The intuition for this somewhat counterintuitive finding is that because the manager can take a “real” *productive* action in addition to an *unproductive* action (i.e., biasing reported earnings), the principal has to provide the manager with incentives to encourage the productive action. Greater investor uncertainty makes stock price less sensitive to reported earnings, regardless of whether the earnings are the result of the productive action or simply reporting bias. Thus, in order to encourage the productive action, the principal provides the manager with stronger incentives.

Since an exogenous change in investors' uncertainty about a manager's incentives should also cause the principal to change the manager's incentives—as Peng and Röell (2008, 2014) show theoretically—the new CD&A disclosure rules should presumably produce a change in *both* managers' incentives *and* investors' uncertainty about these incentives, and not just the latter which Ferri et al. assume when purporting to test Fischer and Verrecchia (2000). To the authors' credit, they explicitly acknowledge this possibility and test for changes—or a lack thereof—in managers' compensation contracts following the new disclosure rules to validate the crucial identifying assumption that the 2006 rule change represented an exogenous shock that reduced *investors' uncertainty about* these contracts, but that left the *structure* of these contracts unchanged.³

Despite the clear insight provided by Fischer and Verrecchia (2000) and Peng and Röell (2008; 2104), these models—like all others—necessarily require some abstractions that may be consequential when confronted with data. For instance, these (and other) models place restrictions on the contract space (e.g., the shape of the contract and the performance measures that can be contracted upon). If the principal is limited in the performance measure available to encourage (or discourage) multiple managerial actions (e.g., productive effort *and* truthful financial reporting), the principal's problem is generally more difficult to solve.⁴ There is ample evidence that managers are compensated on the basis of *both* stock price and earnings (as well as other objective and subjective performance measures). Moreover, earnings-based compensation contracts frequently make adjustments to reported GAAP earnings (Curtis et al., 2017; Black et al., 2017), which suggests that boards might deliberately attempt to create separate performance measures for “stewardship” and “valuation” purposes, which are different problems. In other words, the fact that boards make adjustments to GAAP earnings for compensation purposes suggests that they might be trying to decouple managers' financial reporting incentives from their job of producing and reporting valuation-relevant earnings (e.g., Gjesdal, 1981; Paul, 1992; Lambert, 2001).

2.3. Which theory is being tested?

The particular theory that an empirical paper relies on, or “tests,” depends on the features of the research setting. In the case of Ferri et al., if managers' compensation contracts—or, more generally, their financial reporting incentives—also change, as recent evidence in Gipper (2017), Bloomfield (2018), and others suggest, then Peng and Röell is arguably the more appropriate theory—i.e., the one that more closely corresponds to the research setting. On the other hand, if managers' compensation contracts do *not* change,

² From an agency-theoretic perspective, it is natural to think about the manager's objective in Fischer and Verrecchia (2000) as an incentive compatibility (IC) constraint in a larger optimization problem in which the principal (e.g., the board, acting on behalf of shareholders) endogenously designs the manager's compensation contract as a function of stock price to encourage a productive action, knowing that the resulting contract might also encourage earnings manipulation as an unintended, but tolerable side effect. In this sense, Fischer and Verrecchia (2000) can be viewed as a special case of Peng and Röell's models in which the manager's IC constraint that describes his reporting incentives is the objective function rather than a constraint embedded in the principal's higher-level objective.

³ Ferri et al. attempt to alleviate concerns that the change in disclosure rules led to changes in managers' compensation contracts in two distinct analyses. First, the authors test for changes in several common incentive-compensation measures across treatment and control firms and find no detectable difference. Second, the authors argue that although the new CD&A disclosure rules may have affected the structure of managers' compensation contracts, the rules are unlikely to have led to a *differential* change at treatment and control firms. Because the new rules were proposed and approved before the control firms' fiscal year end, both treatment and control firms should have begun their fiscal years knowing the new rules and with enough time to adjust their managers' compensation contracts for the following year.

⁴ As Lambert (2001, p. 4) explains: “In the single-action case, the only incentive problem concerns the intensity of the agent's effort. Any variable sensitive to the agent's effort can be used to increase that effort. However, in the multi-action case, the contract must not only motivate the overall intensity of the agent's effort, but also its allocation across its dimensions.”

as Ferri et al. claim, then Fischer and Verrecchia may indeed be the more appropriate theory that better corresponds to the research setting.

Recall the two key results from Fischer and Verrecchia (2000) are that a decrease in investors' uncertainty about managers' financial reporting objectives leads to an increase in both (i) ERCs, and (ii) the amount of reporting bias. However, in light of Ferri et al.'s finding of no change in reporting bias following the new disclosure, their study may be better viewed as testing whether increased disclosure *in general*—rather than increased disclosure about managers' reporting incentives *in particular*—leads to an increase in ERCs. In this case, Holthausen and Verrecchia (1988), who theoretically show that general improvements in disclosure quality lead to an increase in ERCs, may be a more relevant theory.

3. Correspondence between the theory and the research setting: 2006 SEC disclosure rules

The theoretical models discussed in the previous section abstract away from a number of potentially important features of Ferri et al.'s research setting. Although this is not a problem *per se*—since “theories derive their value precisely because they distort reality” (Leamer, 2007, p. 4590)—it is nevertheless important to determine which abstractions are innocuous and which are consequential. To the extent that a theoretical model abstracts away from an important aspect of the research setting it becomes important to either (i) ensure that the research design controls for (and therefore *also* abstracts away from) the important institutional features that are not contemplated by the model, or (ii) extend the theory—typically by invoking auxiliary assumptions and introducing verbal arguments and reasoning—so that it more closely corresponds to the research setting. Rubin (2004), p. 345 explains the role of such assumptions for drawing causal inferences as follows:

[a]ll causal inference relies on assumptions that restrict the possible potential outcomes so that we can learn something about causal effects from observable data. Nothing is wrong with making assumptions; on the contrary, such assumptions are the strands that join the field of statistics to scientific disciplines. The quality of these assumptions and their precise explication, not their existence, is the issue.

It is therefore important to assess whether the key features of Ferri et al.'s research setting are captured by Fischer and Verrecchia (2000), or whether some other model (e.g., Holthausen and Verrecchia, 1988; Peng and Röell, 2008; 2014) more closely approximates their setting.⁵

The 2006 SEC rules changed many components of firms' annual proxy statements, including (i) director compensation tables, (ii) CD&A, (iii) option grant disclosures, (iv) retirement plan and perquisite information, and much more. A visible symptom of the extensive nature of these changes is that the average proxy statement word (page) count increased from approximately 16,500 (45) during 2004 to 2006 to roughly 19,000 (63) from 2008 to 2009.⁶ This raises the question of whether the 2006 SEC disclosure rules were exclusively a shock to investors' uncertainty about managers' reporting incentives, or whether alternative causal mechanisms may have also been at work.

Ferri et al. assume that the regulation was a “shock” that increased the precision of investors' information in the Fischer and Verrecchia (2000) sense. However, as discussed above, the rule change could have also triggered changes in executive compensation contracts (e.g., boards renegotiating managers' contracts), which could also alter ERCs through changes in managers' incentive alignment (Peng and Röell, 2008; 2014). Several recent studies present evidence consistent with this alternative mechanism: Bloomfield (2018) and Gipper (2017) examine the same research setting as Ferri et al. and find that boards re-contracted with executives following the new disclosure rules.⁷ Therefore, a test of Fischer and Verrecchia (2000) requires “controlling” for this mechanism in order to draw credible inferences about the causal effect of a change in the precision of investors' information about managers' incentives on their firm's ERCs.⁸

4. Research design: implicit assumptions

Ferri et al. develop a difference-in-differences research design using the 2006 SEC CD&A rules as an instrument for investors' uncertainty about managers' reporting incentives. As in any quasi-natural experimental research design, the ability to draw credible causal inferences from observational data relies on inherently *untestable* assumptions. For example, conference participants were quick to question the validity of Ferri et al.'s research design in light of an apparent violation of the “parallel trends” assumption. Specifically, several participants noted that the treatment firms' ERCs do not appear to have changed while the control firms' ERCs

⁵ See also, Pearl (2009, p. 358), who notes that “[m]athematics deals with ideal situations, and it is the experimenter's job to make sure that the experimental conditions approximate the mathematical ideal as closely as possible.”

⁶ The 2006 CD&A rules also closely coincided with the mandatory adoption of SFAS 123R stock option expensing. While SFAS 123R began to take effect in 2005 for voluntary adopters, 2006 was the first full year that it became mandatory. Prior studies have documented numerous changes to executive compensation contracts as firms responded to the new financial reporting requirements. Hayes et al. (2012), Murphy (2013), and Core and Packard (2017) discuss firms' shift away from stock options toward performance vesting restricted stock.

⁷ It would be useful for future research to investigate the source of the seemingly discrepant findings in Bloomfield (2018) and Gipper (2017) who, on the one hand, find that executives' compensation contracts changed following the 2006 rule change, and Ferri et al. who, on the other hand, find no change in the same setting. One obvious candidate for the apparent discrepancy is their focus on different compensation measures (e.g., Bloomfield examines revenue-based performance metrics and Gipper focuses on pay levels).

⁸ Ferri et al. (p. 4) note how their research design “relies on a maintained assumption that the other parameters potentially affecting the ERC did not change differentially for treatment and control firms around the introduction of the CD&A rules.”

appear to have decreased following the new disclosure rules, which accounted for a *relative*, but not an *absolute*, increase in treated firms' ERCs. However, this may not necessarily be a problem. As Gelman and Hill (2006), p. 228 explain:

The assumption needed with [a difference-in-differences] strategy is weaker than the (unconditional) ignorability assumption because rather than assuming that potential outcomes are the same across treatment groups, one only has to assume that the potential gains in potential outcomes over time are the same across groups. (emphasis added)

In the context of Ferri et al., this assumption requires that the *difference* in ERCs over time would have been the same across the treatment and control firms, but for the treatment (i.e., the new CD&A rules). This could occur if, for example, the ERCs of both treatment and control firms would have declined following the 2006 rule change (e.g., because of macroeconomic conditions that attenuated the ERCs of all firms in the economy), but treatment firms' increased disclosure offset the decline in their ERCs that would have otherwise occurred.

The apparent lack of a change in the treatment firms' ERCs coupled with a change in the control firms' ERCs highlights Ferri et al.'s maintained assumption that the treatment firms' ERCs also *would have changed* in the same way that the control firms' ERCs *did change* following the 2006 disclosure rules, but for their new disclosures. The conference participants' focus on this point might be a reflection of the fact that Ferri et al. do not provide a compelling explanation for what accounted for the change in the *control firms'* ERCs following the new disclosure rules and, in turn, why the treatment firms' ERCs also would have changed.

Although this observation does not seem to provide conclusive evidence of a violation of the “parallel trends” assumption, this discussion highlights the importance of evaluating the validity of not only the well-known assumptions that are *explicitly* stated (e.g., parallel trends), but also any assumptions that are left *implicit* that are also necessary for drawing causal inferences. We now discuss—and assess the validity of—several important implicit assumptions.

4.1. No interference

The so-called Stable Unit Treatment Value Assumption (SUTVA) is an important—but frequently overlooked and therefore implicit—assumption that excludes the possibility of: (i) units interfering with each other, and (ii) multiple versions of the treatment. The former requires that each firm's treatment status has no effect on the potential outcomes of other firms, which acknowledges the “partial equilibrium” nature of natural experiments by ruling out any general equilibrium effects. The latter precludes “non-homogeneous treatments,” which could result from, *inter alia*, differential compliance with the disclosure regulation.

It is important to recognize that SUTVA is a type of exclusion restriction that, like other exclusion restrictions, cannot be verified from the data, but is instead based on previous subject matter knowledge. The restriction embodied by the “no interference” assumption requires that subjects' treatment status does not affect either the treatment status of other subjects, or either the potential or actual outcome of other subjects. In the context of Ferri et al., if this assumption is violated, then firms' potential outcomes are not uniquely defined, but instead differ depending on whether or not other firms received the treatment. Second, the disclosure of treatment firms should have no effect on the predetermined outcomes of control firms that are “close to” (e.g., competitors or those with the same compensation consultant) treatment firms.

Ferri et al.'s inferences may be confounded if *treatment* firms' proxy statements under the new disclosure regime provide investors with information about the incentives of *control* firms' managers. For example, CBS, Time Warner, and Viacom all have a December fiscal year-end and are presumably in Ferri et al.'s “treatment” sample. All three firms include each other *and* Disney, which has an October fiscal year-end and is therefore a control firm, in their respective compensation peer groups. In addition, all four companies retained Towers Perrin as their compensation consultant and operate in the same (Fama–French 48) industry. Finally, and perhaps not surprisingly, all four companies include similar performance measures (i.e., operating income and cash flow) in their executives' bonus plans. Therefore, knowledge of managers' compensation contracts at CBS, Time Warner, and Viacom likely allows investors (and analysts) to form reasonable expectations about Disney's managers' compensation contracts. To the extent that CBS, Time Warner, and Viacom's CD&A disclosures following the new reporting rules were also informative about Disney's managers' incentives, then the no interference requirement of SUTVA would seem to be violated.⁹

Although identifying treatment firms may be relatively straightforward (i.e., those with more informative disclosure), it is not obvious what firms should form the control group against which to compare the treatment group. For example, should the control group consist of firms with no close competitors that provide more informative disclosure? Alternatively, should it include firms that have a close competitor that do provide more informative disclosure? This is an important research design choice because different control conditions imply different counterfactual states and, in turn, lead to different estimates of the causal effect. Returning to our example, a potential solution—or “sensitivity analysis”—would be to repeat the analysis assuming that Disney is a “treatment” firm to determine whether the same results obtain.

In the preceding example of the CD&A disclosures for compensation peer group firms with different fiscal year-ends, the potential

⁹ We also note that there are arguments that militate against the possibility of such spillovers and, in turn, a violation of SUTVA. For example, even if peer firms use similar performance metrics and rely on the same compensation consultant, it may nevertheless be difficult for investors to predict the specific weights and performance targets for a particular firm based on its *peers'* proxy statements. The extent to which investors learn about managers' incentives at control firms based on the disclosures of treatment firms has crucial implications for the validity of Ferri et al.'s research design and the credibility of their inferences. Moreover, the plausibility of this countervailing argument highlights the importance of acknowledging identifying assumptions and providing persuasive arguments and evidence to justify their validity.

(and actual) outcome of Disney depends only on whether Disney received the treatment and not on whether CBS or any other compensation peer received the treatment. However, if Disney's potential outcome depends not only on whether Disney received the treatment, but also on whether CBS (or any of Disney's other compensation peers) received the treatment, the number of potential outcomes grows exponentially as the number of firms that “spill over” to Disney increases.

One way to address a potential violation of SUTVA is to redefine—and, in particular, “coarsen”—the unit of analysis to a broader grouping (e.g., compensation peer groups, industries, etc.) that captures any spillovers. In the case of Ferri et al., this might be accomplished by examining the effect of an increase in investors' information about the incentives of managers within peer groups—induced by the new disclosure rule—on, say, the average ERC of firms in the peer group. In this case, the coarser unit of analysis (i.e., peer group rather than firm) allows for the possibility that investors learn about the incentives of managers within the peer group either directly from the firm's own disclosure, or indirectly from the disclosures of its peers. However, it is important to note that doing redefining the unit of analysis inherently alters the nature of the resulting inferences, which would relate to the net effect of the new CD&A disclosures on the broader group—inclusive of any spillovers—rather than its effect on individual firms.

4.2. Homogeneous treatment

The second element of the SUTVA in the context of Ferri et al. is related to the SEC's Intent and Operation of the Compensation Discussion and Analysis. As discussed in the ruling, “The Compensation Discussion and Analysis requirement is principles-based. [...] [O]verall we designed the proposals to state the requirements sufficiently broadly.” The “principles-based” nature of the rules suggests that firms have considerable discretion in complying with the new disclosure requirements. Consequently, it is not obvious whether the new disclosure rules represented a single homogeneous “treatment” that applied uniformly to all (December year-end) firms, or whether it is better characterized as providing non-homogeneous treatments that depend on how particular firms endogenously chose to interpret and implement the rules.

A cursory examination of proxy statements following the rule change suggests that there was, in fact, differential compliance with the new rules. In other words, different firms provided different amounts of detail—and therefore different degrees of precision—about their executives' compensation contracts. Consequently, the rules introduced multiple version of treatment. Pearl (2009), p. 260 explains the inferential problems that result from differential—or imperfect—compliance:

“[i]mperfect compliance poses a problem because simply comparing the fractions in the treatment and control groups may provide a misleading estimate for how effective the treatment would be if applied uniformly to the population. In the absence of additional modeling assumptions, treatment effectiveness cannot be estimated from the data without bias, even when the number of subjects in the experiment approaches infinity and even when a record is available of the action and response of each subject.”

One potential solution for dealing with non-homogeneous treatments in observational studies is to “coarsen” the definition of the treatment. In Ferri et al.'s setting, this might be accomplished by defining the “treatment” as the 2006 SEC Regulations rather than a reduction in investors' uncertainty about managers' reporting incentives *per se*. Describing the treatment in this way would make transparent that the study examines the effect of the regulation more broadly on investors' response to reported earnings, rather than only through the specific channel of a reduction in investors' uncertainty about managers' reporting incentives. This approach is analogous to the previously discussed solution for a potential violation of the first SUTVA element (i.e., no interference), which is to “coarsen” the unit of analysis. It is also similar in that it alters the research question, the nature of the inferences, and may result in a substantial departure from the theoretical predictions of any particular model (e.g., Fischer and Verrecchia, 2000).

4.3. Conditioning on post-treatment variables

Another potential concern with the research design in Ferri et al. is conditioning on—or controlling for—variables that may themselves be affected by treatment (see related discussion in Bertomeu et al., 2016). Angrist and Pischke (2009) refer to post-treatment control variables as “bad controls,” and Pearl (2009) refers to conditioning on post-treatment variables that lie on the causal path between the treatment and outcome as “conditioning on a collider.” These and other authors explain the inferential problems associated with this practice and discuss the importance of devoting careful attention to the choice of control—or partitioning—variables and the timing of their measurement.¹⁰ This issue also arises when researchers wish to estimate causal effects for different subsamples, but identify the different subsamples based on a measure determined *after* the intervention. Failure to specify the empirical model in a way that conforms to the underlying causal theory can result in—perhaps inadvertently—conditioning on post-treatment outcomes. Ultimately, the question of whether to include a particular control—including various types of fixed effects—should be answered by consulting the underlying theory.

One possible instance of conditioning on a post-treatment variable occurs when Ferri et al. test for moderating effects of disclosure by partitioning their treatment firms according to whether they received an SEC comment letter because of perceived deficiencies in their disclosure.¹¹ The intuition for this test is that treatment firms that received letters from the SEC presumably provided investors

¹⁰ As Rosenbaum (2005) explains: “A variable measured prior to treatment is not affected by the treatment and is called a covariate. A variable measured after treatment may have been affected by the treatment and is called an outcome. An analysis that does not carefully distinguish covariates and outcomes can introduce biases into the analysis where none existed previously.”

¹¹ The fact that some treatment firms received SEC comment letters suggests that there was differential compliance with the new CD&A disclosure

with less information about their managers' financial reporting incentives. Consequently, these firms' ERCs should exhibit an attenuated response to the new disclosure rules.

Although this research design that partitions the sample based on a variable that the underlying theory suggests should moderate the main effect is common in the causality literature, there is a crucial difference. Typically, the conditioning variables are measured *before* the treatment and, unless the measure is influenced by expectation of the treatment, such variables are presumably not influenced by the treatment. However, given that the comment letters were necessarily issued *after* firms disclosed under the new rules and that the likelihood of receiving an SEC comment letter depends on a firm's compliance to the new disclosure rules, this partitioning variable is almost mechanically influenced by the treatment. In particular, given the discretion in implementation, managers who expected to benefit the most from biasing earnings may have provided less informative disclosures when deciding how to comply with the new disclosure rules.

A second potential instance where Ferri et al. seem to condition on post-treatment variables is more subtle. Ferri et al. define unexpected earnings (UE) as the difference between a firm's realized earnings and the most recent analyst consensus earnings forecast scaled by the firm's stock price immediately before the earnings announcement. However, the most recent analyst consensus earnings forecast occurs *after* firms disclose details of their managers' compensation contracts. Similar to how investors are expected to use the disclosure to "back out" bias in reported earnings, analysts would also presumably use the disclosure to anticipate managers' reporting bias and incorporate this into their earnings forecasts. Since the consensus analyst earnings forecast is formed *after* the disclosure and is likely affected by the disclosure, it also appears to be a post-treatment variable. Moreover, if the disclosure affects firms' stock price, then scaling firm's earnings surprise by stock price measured *after* the disclosure also seems to be an instance of conditioning on a post-treatment variable. Although Ferri et al.'s descriptive statistics show that pre-treatment variable means are largely similar across the treatment and control groups, unexpected earnings—which, as explained above, seems to be constructed using post-treatment variables—differs across the two groups.

5. The role of explanatory theory in causal inference

The foregoing discussion highlights the central role of explanatory theory in causal inference.¹² Explanatory theories describe certain structural properties of the world and allow us to deduce predictions about relationships between the phenomena that we seek to understand. No amount of empirical evidence—regardless of whether it is from an observational study, a quasi-natural experiment, or randomized controlled trial—can "prove" the causality of a relationship.¹³ This point is particularly salient in light of researchers' apparent desire to draw ever stronger causal inferences. Instead, causality is an abstraction that exists in the context of a (conjectured) theoretical explanation about the relationship between some phenomena of interest. These conjectured explanations are subjected to criticism aimed at detecting and eliminating errors. A great deal of this criticism takes the form of argument that identifies logical inconsistencies and areas where the theory conflicts with other extant explanations, which is to say those theories that have withstood criticism and have not yet been—but may eventually be—falsified. The existence of multiple viable, but conflicting theories is the occasion for empirical testing, which can involve both observation and experimentation.

In the case of Ferri et al., Fischer and Verrecchia (2000) is not the only viable theory in the sense of offering a plausible explanation for a relationship—that may or may not be causal—between increased disclosure about the parameters of managers' compensation contracts and their firm's ERCs. In particular, two other theories immediately come to mind. First, Holthausen and Verrecchia (1988) show that increased disclosure *in general* leads to increased ERCs. Second, Peng and Röell (2008; 2014) show that a reduction in investors' uncertainty about managers' incentives should lead to changes in managers' incentives themselves, which also affect firms' ERCs.

An important question that arises from Ferri et al. is how to discriminate among these—and possibly other—related theories. In order to properly test Fischer and Verrecchia's (2000) theory, Ferri et al. necessarily make a host of auxiliary assumptions—many of which are implicit—to ensure that their research design "tests" (i.e., has the potential to falsify) Fischer and Verrecchia's theory.¹⁴ It is important to acknowledge and justify such assumptions that are necessarily inherited by the research design of any empirical study that purports to test a particular theory.

6. Conclusion

Ferri et al. examine an interesting, theoretically-motivated research question to empirically document whether investors'

(footnote continued)

rules. If so, this would be *prima facie* evidence of non-homogeneous treatment and would constitute a failure of the second component of SUTVA.

¹² Norwood Russell Hanson's claim that all observation is "theory-laden" underscores the indispensable role of theory in empirical work. It also serves as a reminder that data cannot even be obtained without a theory—perhaps largely unstated and implicit—that tells us what to look for, where and when to look for it, how to identify and measure it, and why.

¹³ For instance, Rosenbaum (1999, p. 260) states that: "To discuss the quality of evidence provided by an empirical study one must first recognize that evidence is not proof, can never be proof, and is in no way inferior to proof. It is never reasonable to object that an empirical study has failed to prove its conclusions, though it may be reasonable, perhaps necessary, to raise doubts about the quality of its evidence."

¹⁴ Samuelson (1947, pp. 4-7), *The Foundations of Economic Analysis*, writes: "[O]ur theory is meaningless in the operational sense unless it does imply some restrictions upon empirically observable quantities, by which it could conceivably be refuted... By a meaningful theorem I mean simply a hypothesis about empirical data which could be refuted, if only under ideal conditions. A meaningful theorem may be false."

uncertainty about managers' financial reporting incentives has a causal effect on the relation between their firm's unexpected earnings and stock returns. The authors develop a creative research design to isolate arguably exogenous variation—and, specifically, a reduction—in investors' uncertainty about managers' reporting objectives to show that it led to larger stock price reactions to managers' earnings disclosures. We wholeheartedly commend the authors for their combination of specific theory and rigorous empirical tests in a novel research setting. Although we may not necessarily agree with all of their inferences, their paper provides a nice template for how to generate credible causal inferences from the combination of sharp predications derived from a specific theoretical model and rigorous empirical tests developed in a quasi-experimental research setting.

Acknowledgment

We thank S.P. Kothari (editor), Mirko Heinle, Dan Taylor, David Tsui, Ro Verrecchia, and the authors of Ferri, Zheng, and Zou for constructive comments and suggestions. Our discussion has also benefitted from the questions and comments from the 2017 *Journal of Accounting and Economics* Conference participants. We gratefully acknowledge financial support from the Wharton School of the University of Pennsylvania. Armstrong thanks EY and Kepler thanks the Deloitte Foundation for their research support.

References

- Angrist, J., Pischke, J., 2009. Mostly harmless econometrics: an empiricist's companion. Princeton Univ Press.
- Armstrong, C., Jagolinzer, A., Larcker, D., 2010. Chief executive officer equity incentives and accounting irregularities. *J. Account. Res.* 48 (2), 225–271.
- Armstrong, C., Larcker, D., Ormazabal, Taylor, D., 2013. The relation between equity incentives and misreporting: the role of risk-taking incentives. *J. Financ. Econ.* 109 (2), 327–350.
- Bertomeu, J., Beyer, A., Taylor, D., 2016. From casual to causal inference in accounting research: the need for theoretical foundations. *Foundations Trends Account.* 10 (2–4), 262–313.
- Black, D., Black, E., Christensen, T., Gee, K., 2017. The Use of non-GAAP Performance Measures in Executive Compensation Contracting and Financial Reporting. Working paper.
- Bloomfield, M., 2018. Compensation Disclosures and the Weaponization of Executive Pay: Evidence from Revenue-Based Performance Evaluation. Working paper.
- Core, J., Packard, H., 2017. Non-price and Price Performance Vesting Provisions and Executive Incentives. Working paper.
- Curtis, A., Li, V., Patrick, P., 2017. The Use of Adjusted Earnings in Performance Evaluation. Working paper.
- Fischer, P., Verrecchia, R., 2000. Reporting bias. *Account. Rev.* 75 (2), 229–245.
- Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge university press.
- Gjesdal, F., 1981. Accounting for stewardship. *J. Account. Res.* 19 (1), 208–231.
- Gipper, B., 2017. The Economic Effects of Mandating Expanded Compensation Disclosures. Working paper.
- Hayes, R., Lemmon, M., Qiu, M., 2012. Stock options and managerial incentives for risk taking: evidence from FAS 123R. *J. Financ. Econ.* 105 (1), 174–190.
- Holthausen, R., Verrecchia, R., 1988. The effect of sequential information releases on the variance of price changes in an intertemporal multi-asset market. *J. Account. Res.* 82–106.
- Lambert, R., 2001. Contracting theory and accounting. *J. Account. Econ.* 32 (1–3), 3–87.
- Leamer, D., 2007. Chapter 67 Linking the Theory with the Data: That is the Core Problem of International Economics. In: Heckman, James J., Leamer, Edward E. (Eds.), *Handbook of Econometrics* 6, Part A. Elsevier, pp. 4587–4606.
- Murphy, K., 2013. Executive compensation: Where we are, and how we got there. *Handbook of the Economics of Finance* 2. Elsevier, pp. 211–356.
- Paul, J., 1992. On the efficiency of stock-based compensation. *Rev. Financ. Stud.* 5 (3), 471–502.
- Pearl, J., 2009. Causality: Models, Reasoning and Inference. Cambridge University Press, New York, NY, USA.
- Peng, L., Röell, A., 2008. Manipulation and equity-based compensation. *Am. Econ. Rev.* 98 (2), 285–290.
- Peng, L., Röell, A., 2014. Managerial incentives and stock price manipulation. *J. Finance* 69 (2), 487–526.
- Rosenbaum, P., 1999. Choice as an alternative to control in observational studies. *Stat. Sci.* 259–278.
- Rosenbaum, P., 2005. Observational Study. In *Encyclopedia of Statistics in Behavioral Science* (eds B. S. Everitt and D. C. Howell).
- Rubin, D., 2004. Teaching statistical inference for causal effects in experiments and observational studies. *J. Educ. Behav. Stat.* 29, 343–367.
- Samuelson, P., 1947. Some implications of linearity. *Rev. Econ. Stud.* 15 (2), 88–90.