Journal of Accounting and Economics 000 (2017) 1–8

Contents lists available at ScienceDirect

Journal of Accounting and Economics

journal homepage: www.elsevier.com/locate/jacceco

Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017)☆

Stephen Glaeser, Wayne R. Guay*

The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

ARTICLE INFO

Article history:
Received 7 August 2017
Accepted 8 August 2017
Available online xxx

JEL classification:
C12
C51
M40
M41

Keywords:
Causal inference
Accounting research
Quasi-experimental methods
Generalizability

ABSTRACT

Christensen et al. (2017) provide evidence that the dissemination of mine safety information in SEC filings has real effects on mine safety. We discuss the extent to which Christensen et al.'s results generalize to a research question that we consider of broader interest to accounting researchers, specifically where and when mandated disclosure in SEC filings can increase the dissemination of information. We also discuss identification of causal effects and generalizability concerns more broadly in the context of large sample studies and quasi-natural experiments, as well as potential ways authors might address these concerns in accounting research.

1. Introduction

Identification of causal effects has become a front-and-center issue in accounting research in recent years. This is not surprising given its growing prominence in economics and finance research, and is largely warranted given the large number of research topics where convincingly documenting causal relationships has proven challenging. Because the perfectly identified and generalizable research design is rare, if not unattainable, identifying causal relationships for important accounting research questions is most likely to come from a mosaic of studies that collectively update our priors. We refer to this mosaic as a "Bayesian approach to causal inference," in the sense that each well-executed study on a particular topic offers evidence that researchers use to update their priors on the applicability and generalizability of the theory being tested. This approach, discussed in the context of the research questions addressed by Christensen et al. (2017), will be the focus of our comments.

Issues of generalizability feature prominently in our discussion since research studies that are perhaps the best identified are often the least generalizable, and vice versa. Our discussion is meant to complement the methodological literature that discusses the frequent tradeoff between internal and external validity faced by researchers (e.g., Shadish et al., 2002; Leuz and Wysocki, 2016). We believe that the most important research questions in accounting require a general understanding

☆ We thank Chris Armstrong, Terrence Blackburne, John Core (editor), Bob Holthausen, Rick Lambert, Christian Leuz, and Dan Taylor for helpful comments and suggestions.

* Corresponding author.
E-mail addresses: sglaeser@wharton.upenn.edu (S. Glaeser), guay@wharton.upenn.edu (W.R. Guay).

http://dx.doi.org/10.1016/j.jacceco.2017.08.003
0165-4101/

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

of the phenomenon, and there is ample room for both broad sample and narrow sample research designs to update the priors of the academic community.

Christensen et al. ask the interesting question of whether the mandated disclosure of social responsibility information in SEC filings can have real effects on the disclosing firms, even when the information has been previously disclosed. They motivate this question with the observation that U.S. policymakers appear to be increasingly interested in mandating disclosure of social responsibility information in SEC filings. To explore this question, the authors identify a regulatory change in the mining industry where safety violations were mandated to be disclosed in SEC filings, even though this information was already required to be disclosed on a government website in a timelier manner. Specifically, following the tragic explosion at Massey Energy's Upper Big Branch mine in April 2010, policymakers required SEC-registered firms to disclose mine-safety violations in their periodic 10-K and 10-Q reports (hereafter referred to as "MSD"), and to report immediately imminent danger orders in an 8-K filing. These violations, however, were already required to be publicly disclosed on the government's Mine Safety and Health Administration (MSHA) website, typically within 24 h. The authors argue that mandated disclosure can nonetheless influence the behavior of mining firms, investors, and other interested parties if the disclosure in SEC filings disseminates mine safety information to a wider audience.

To facilitate convincing identification in this setting, the authors take advantage of a control group of private mining firms that were not subject to the SEC disclosure requirements. To the extent that these private mining firms are similarly affected by all of the other economic factors that affected public mining firms, the authors can compare the change in the behavior between the groups following the regulatory shock to estimate a causal treatment effect. Thus, this regulatory shock provides a plausibly well-specified quasi-natural experiment (QNE) to explore whether safety violations disclosed in mining firms' SEC filings reach a broader audience than the same safety violations disclosed on the MSHA website.

The authors provide convincing evidence that mining firms' SEC filings disseminate mine safety violation information to a wider audience than does the MSHA website, and that this dissemination has real effects on public mining firms. Specifically, they document negative announcement returns and negative effects on holdings by mutual funds that identify as socially responsible, as well as increased media and analyst attention, when safety violations are disclosed in the 8-K. Further, because working conditions and worker safety are important corporate social responsibility (CSR) issues, the authors also explore whether this greater dissemination of safety violation information induces public mining firms to improve safety conditions for workers. Consistent with this "real effect," the authors find that public mining firms do improve safety, as evidenced by a lower incidence of safety violations and injury rates for public mines relative to private mines following MSD. Safer mines undoubtedly save lives, and Christensen et al.'s findings provide valuable insights into the SEC's role in effecting change through mandated disclosure requirements.

Although the real effects documented by Christensen et al. are interesting, we do not view these tests as the most important aspect of the paper for a general accounting audience. As noted by the authors in their literature review, a large body of prior research documents real effects on firm behavior when information is conveyed to investors, creditors, employees, suppliers, customers, and regulators. And, in light of this prior work, the authors acknowledge that their work, "contributes to this prior work primarily because mine-safety records are already publicly available outside of a firm's financial reports, which allows us to isolate and estimate the incremental effect of including information in financial reports as opposed to the effects of disclosing information not previously publicly released elsewhere." We agree with this assessment, and would argue that the most interesting and novel question that Christensen et al. address relates to advancing our understanding of how information is disseminated via SEC filings, as opposed to whether dissemination of information can have real effects.

Corporations have a number of mediums at their disposal to disseminate information, including financial reports, conference calls, press releases, social media, and the corporation's own website. Prior research has documented that some mediums disseminate information more broadly than others.[1] However, our current understanding of the dissemination "pecking order" for disclosure mediums, and how this order varies across firms and economic contexts, is at best limited.

The results in Christensen et al. are intriguing, and identify a setting where SEC filings appear to disseminate a specific type of information more widely than the website-based medium that was previously used to provide that same information. At the same time, it is reasonable to ask how far the study has advanced the broad and provocative question that is likely to be of interest to a general accounting audience; that is, when and where mandated disclosure in SEC filings can increase the dissemination of information that has been previously disclosed publicly? Although the authors argue that their findings contribute to the broad literature on the real effects of disclosure and financial reporting, how far has their work advanced our understanding of the pecking order of disclosure mediums in terms of their ability to disseminate information? To answer this question, a natural starting point is to assess the generalizability of the authors' inferences from their QNE research setting.

2. Assessing generalizability

To assess the generalizability of inferences, one might begin by considering the specific features of a research setting. In this case, specific features include the following facts: 1) the public dissemination of the safety disclosures is via the gov-

[1] E.g., Bushee et al. (2010), Engelberg and Parsons (2011), Drake et al. (2012), Blankespoor et al. (2013), Peress (2014), Drake et al. (2015).

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

ernment's MSHA website; 2) the sample includes only mining firms; 3) the disclosure relates specifically to mine safety. As discussed below, the combination of these three facts would seem to be a very limiting aspect of this study's generalizability.

Beginning with the MSHA website, one might ask how much traffic this website receives. As of November, 2016, the MSHA website ranked #62,976 for U.S. web traffic. In our conference discussion, we benchmarked this ranking as being comparable to some rather obscure websites (a hobby store website in Blue Ridge, Georgia; a Chrysler Minivan Fan Club website, and; a website posting stories about mother-in-laws). Although this portion of the presentation was partly theatrical for the benefit of audience attention, the exercise was intended to emphasize that the MSHA website likely receives most of its traffic from the relatively small group of the most interested and invested followers of mining companies and mining safety. Thus, it is perhaps not surprising that disclosing mine safety violations in SEC filings would capture the attention of a broader audience that includes retail investors, the general public, and the media.

Building on this point, it is reasonable to ask how the results might be different if the safety violations information had been publicly disclosed via a different medium. For example, what if the information had been publicly disclosed on each mining firm's corporate website? Or, perhaps publicly disclosed in press releases or news articles? On a mine safety watchdog website? On a social media platform? It is not hard to imagine that disclosure in SEC filings might disseminate information more broadly than some of these mediums, but less broadly than others. Because the effect of dissemination is likely a function of the degree to which the information was observed prior to being disseminated, and the degree to which it was disseminated, it is reasonable to question whether the treatment effects documented by Christensen et al. would generalize to these other settings.

Compounding the complexity of this issue, consider that the specific industry being examined in the paper is mining, and the specific piece of information being disclosed is mine safety violations. Mine safety is a very specific type of CSR issue. Although worker safety is an issue of universal concern, the keen followers of miner safety are likely to be fairly localized to mining communities. Therefore, one might ask whether SEC filings would have different dissemination effects when examining CSR issues with wider appeal and following, such as climate change, political issues, human rights, worker equality, etc. Further, because mining itself is somewhat of a niche industry, would the dissemination effects of SEC filings be different when studying previously disclosed information about the consumer products, technology, banking, or retail industries (which contain widely followed firms such as Apple, Google, Procter & Gamble, Morgan Stanley, Nike, and Wal-Mart)?

Finally, and perhaps most importantly, what if the previously disclosed information was not about CSR issues, but rather about financial, strategic, investing, or operational issues that provide an even clearer link to expected cash flows and risk? In other words, is the dissemination role of SEC filings specific to CSR information, or does it extend more broadly to information relevant to investors, creditors and analysts in making pricing, investment, and contracting decisions? Again, the broad accounting question of interest to researchers would seem to be, whether, when and where mandated disclosure in SEC filings can increase the dissemination of information that has been previously disclosed publicly. Although the Christensen et al. results provide insight into one small slice of that question, we emphasize that it is wholly premature to deem the question to be answered, and that researchers should use caution when using the Christensen et al. results as support for the dissemination role of SEC filings in other settings.

3. Identification of causal effects

So, in light of these generalizability issues, one might ask why the authors choose this mining industry empirical setting to explore the dissemination role of SEC filings. Although mining is a somewhat interesting and important industry in its own right, it is not a natural or obvious place for accountants to study disclosure issues. Rather, the authors appear to choose the mining industry because it provides a research setting that they view as having favorable properties for the identification of causal effects. In particular, the passage of MSD regulations provides a QNE that the authors believe allows them to make relatively strong causal statements about whether mining safety violation disclosures in SEC filings increase the dissemination of information to investors and the general public.

Sorting out causation from association is unquestionably one of the most topical and important econometric issues in accounting, finance, economics and many other disciplines. Before discussing this issue in the context of Christensen et al., however, it is useful to work through some of the challenges of identifying causal effects more generally.[2] Endogeneity is the term most researchers ascribe to the issues that can confound identification. In most empirical accounting studies, endogeneity can take two general forms: correlated omitted variables and reverse causality (also frequently referred to as "simultaneity"). These are two different problems, with potentially very different solutions. Failure to determine which of these problems is likely to be present in a given study can lead to the use of inappropriate econometric techniques. To guide our discussion, we discuss two illustrative research settings, one that attempts to address concerns about correlated omitted variables, and the other that relates to concerns about reverse causality (note that we do not hold out these illustrations to be ideal research designs, but rather as being instructive in their approach, as well as being familiar to the authors of this discussion).

[2] Our discussion of these issues is intended to complement a growing body of literature exploring issues related to causality within accounting research. For example, see Bertomeu et al. (2016); Chen and Schipper (2016), Gow et al. (2016), Leuz and Wysocki (2016).

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

Guay et al. (2016) study the question of whether financial statement complexity induces managers to resolve informational problems by increasing voluntary disclosure. They examine a broad sample of public corporations, and find that firms with more complex financial statements engage in greater voluntary disclosure, as measured by management forecasts, 8-Ks, and firm initiated press releases. Given this research design, one can envision a relatively straightforward correlated omitted variables concern that more complex firms might have both greater financial statement complexity, as well as greater demand for voluntary disclosure.

There are a variety of ways to mitigate concerns about correlated omitted variables.[3] As a start, if the omitted variables are known and measureable, the researcher can include the variable(s) as controls in the empirical specification. Another useful technique is to include fixed effects (industry, firm, year, etc.), which can control for correlated omitted variables that do not vary within the level of the fixed effect. A changes specification can also help alleviate concerns about omitted variables by differencing out variables that do not vary over time. Finally, cross-sectional predictions can identify conditional relations between the dependent and independent variables that narrow the ways in which a conjectured correlated omitted variable would explain all of the findings. For example, Guay et al. predict and find that the positive relation between financial statement complexity and voluntary disclosure varies with the change in liquidity around the 10-K announcement, the scrutiny of external monitors such as analysts and institutional investors, and both firm performance and the level of earnings management. Although these techniques cannot definitively rule out the possibility that correlated omitted variables influence the findings, in some cases these methods can do a great deal to assuage concerns and allow researchers to draw very convincing inferences.

Reverse (or joint) causality is a substantially different type endogeneity problem than a correlated omitted variables problem. The concern here is that the dependent variable might have a causal effect on the independent variable(s) of interest (in this regard we treat reverse causality as synonymous with simultaneity). As an illustration, consider the prediction in Armstrong et al. (2014) that independent directors will be responsible monitors of managers' accounting and disclosure choices, and therefore that firms with a greater proportion of independent directors on the board will ensure that the firm has higher accounting quality and more transparent disclosure. A competing explanation for this relation, however, is that independent directors have a limited amount time to become informed about the firm's business activities, and are more likely to accept (and be offered) board positions when corporate transparency is strong. Both explanations predict a positive relation between board independence and corporate transparency, but in one case it is strong transparency that causes board independence, and in the other case, it is board independence that causes transparency.

Control variables, changes analyses, and cross-sectional predictions are unlikely to significantly mitigate concerns about reverse causality. The problem here is not ruling out the presence of correlated omitted variables. That is, the problematic variable here is not "omitted", but is actually the dependent variable, which is simultaneously determined with the independent variable. A common technique to address this issue is to find a source of variation in the independent variable of interest that is exogenous with respect to the dependent variable. Researchers can then use this exogenous source of variation in an appropriate research design to estimate causal treatment effects (e.g., regression discontinuity, instrumental variable, or difference-in-differences research designs). Armstrong et al. attempt such an approach using mandated changes in board independence and a QNE research design to explore causal effects of independent directors on firm transparency.[4]

Christensen et al. also take such an approach, estimating a difference-in-differences research design in the context of their QNE setting. They argue that their QNE generates "plausibly exogenous" variation in the independent variable (i.e., dissemination of mine safety information), while leaving other variables of concern unaffected, or at least similarly affected for both public and private mining firms. By other variables of concern, we mean the dependent variable of interest, as well as any omitted variables that covary with both the independent and dependent variable.

QNEs are typically not, however, a panacea for resolving issues of endogeneity. We use the term QNE to refer to differences-in-differences, instrumental variables, and regression discontinuity designs that do not feature random variation, but instead rely on "plausibly exogenous" or "conditionally exogenous" variation. In this regard, the "quasi" in QNE distinguishes QNEs from natural experiments that feature random variation.[5] In the absence of random variation, researchers examining QNEs must identify and eliminate all alternative explanations (i.e., potential correlated omitted variables and

[3] The correlated omitted variables problem might also take the form of sample self-selection. For example, below we discuss IFRS studies, and it is well-known that some firms selectively adopted IFRS early. The determinants of this self-selection have been carefully studied and the failure to control these determinants when comparing the outcomes of firms that adopt IFRS voluntarily to the outcomes of firms that do not would result in potentially serious correlated omitted variables problems.

[4] Roberts and Whited (2011) describe reverse causality/simultaneity with the following example: "Simultaneity bias occurs when y and one or more of the x's are determined in equilibrium so that it can plausibly be argued either that xk causes y or that y causes xk. For example, in a regression of a value multiple (such as market-to-book) on an index of antitakeover provisions, the usual result is a negative coefficient on the index. However, this result does not imply that the presence of antitakeover provisions leads to a loss in firm value. It is also possible that managers of low-value firms adopt antitakeover provisions in order to entrench themselves." Obviously, one cannot resolve this issue with control variables—i.e., the dependent variable cannot also be included as a control variable. Thus, to address this issue the researcher must identify a source of variation in the independent variable (antitakeover provisions in this example) that is not subject to concerns about simultaneity (that is not correlated with firm valuation in this example).

[5] "Plausibly" or "conditionally" exogenous variation is the result of an intentional process that classifies firms into treatment and controls groups based on some non-random variation (e.g., regulation that classifies firms into treatment and control groups based on state borders). The researcher uses this variation to estimate causal effects under the assumption that assignment is exogenous with respect to the outcome of interest, conditional on the empirical model's controls. In contrast to plausibly exogenous variation, random variation is the result of intentionally classifying firms into treatment control groups based on randomly generated variation (e.g., a researcher rolling a die to classify test subjects).

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

reverse causality), in much the same way as in broad sample studies. When viewed through this lens, the benefit of non-random QNEs is that they may allow the researcher to narrow the set of plausible alternative explanations for the findings.

One challenging aspect of many QNEs, however, is that the institutional settings examined are often new to both the researchers and the readers, making it difficult to identify plausible alternative explanations for the findings (for example, most accounting researchers are likely to be unfamiliar with the institutional details surrounding mine safety regulations). A potential benefit of highly generalizable QNEs, such as state antitakeover laws, is that many researchers and critics will examine them. Consequently, there is a greater likelihood that key institutional details and potential confounding factors will be identified.[6] This is also a benefit of presenting papers featuring new QNE settings at workshops and conferences, as there is a greater likelihood that a participant familiar with the idiosyncratic features of the new setting will be present and guide researchers through potential confounding factors.

In the context of Christensen et al., conference participants raised several such concerns about potential confounding factors. As one example, outrage about mine safety and the Upper Big Branch disaster plausibly affected not only differential MSD disclosure requirements between public and private mining firms, but also differential enforcement of safety violations between public and private firms.[7] Specifically, prosecutors investigating the Upper Big Branch explosion targeted, for the first time, senior management of a mining company in the aftermath of a disaster. In October, 2011, Massey's security chief, Hughie Stover, was convicted on felony obstruction of justice charges.[8] Further, in November, 2014, Massey CEO Don Blankenship, a long standing opponent of increased environmental and safety regulations, was charged on four criminal counts, including two counts of securities fraud.[9] The securities fraud charges were built around allegations related to Blankenship's role in false and misleading disclosures to Class A Common Stock shareholders, as well as the filing of such false and misleading disclosures with the SEC.[10]

Although the charges against Blankenship were not formally filed until 2014, the Upper Big Branch disaster received substantial attention from the press beginning in April, 2010, and it was well known as early as 2010–2011 (both within the mining industry and more broadly) that Massey Energy and its senior managers were the target of a continuing investigation. For example, in December, 2010, eighteen Massey executives (including Blankenship) invoked their Fifth Amendment rights and refused to be interviewed in the federal investigation.[11] Further, Massey's security chief was indicted on felony charges in March, 2011, and in February, 2012 criminal conspiracy charges were filed against a top-ranking Upper Big Branch mine supervisor.[12] And, at the time the latter charges were filed, the United States Attorney involved in the case stated that he was "absolutely not" finished with the investigation (note also that such criminal charges have a 5-year statute of limitations, which can allow time for prosecutors to build their case).[13]

If the Upper Big Branch mining disaster prompted expectations of stronger mine safety enforcement at public mining companies (via the use of securities laws) as compared to private mining companies, then a differential change in enforcement is a potential correlated omitted variable. In this case, it becomes challenging to disentangle the effects driving Christensen et al.'s results. Our main point here is to emphasize that in QNE settings that do not have random variation, correlated omitted variables are likely to be cause for concern, and researchers must give these issues serious consideration.[14]

As a final point, we have sometimes heard critiques about correlated omitted variables and reverse causality levied as a vague conjecture; as in, "I don't believe your inferences because your results could be driven by endogeneity." Although researchers should think carefully about potential endogeneity issues, a vague and non-specific concern should not be the residual claimant on otherwise competently executed results. A critique about endogeneity should be clear about whether it refers to reverse causality or potential correlated omitted variables. Such a critique should also be well thought out and reasonably specific, as in "existing theory predicts that X is likely to be associated with both your dependent variable and your independent variable of interest, but is currently omitted from your analysis" The same point applies to critiques about generalizability, which should make clear why the result does not generalize, and why generalizing is important for purposes

[6] For example, Karpoff and Wittry (2017) highlight institutional details of state antitakeover laws, a frequently examined QNE setting, and argue that the inferences of several prior studies change substantially when considering these details. Their work emphasizes the challenges faced by researchers in sorting out the institutional features of a given setting, that settings examined in a large number of papers are perhaps more likely to be vetted, and that non-random QNEs are not necessarily a panacea for resolving endogeneity concerns.

[7] We thank Terrence Blackburne for providing details and background documents on this point.

[8] "Some Justice At Upper Big Branch," *New York Times*, October 28, 2011 (Tavernise, 2011).

[9] "Ex-Executive Is Indicted in Disaster at Coal Mine," *New York Times*, November 14, 2014 (Grabirel, 2014).

[10] For example, Blankenship's federal indictment states, "At the time BLANKENSHIP approved the release of and filing of the UBB [Upper Big Branch] Shareholder Statement, he knew that the statements…were materially false, fraudulent, fictitious, and misleading; that the UBB Shareholder Statement contained untrue statements of material fact and omitted to state material facts necessary in order to make the statements made, in light of the circumstances under which they were made, not misleading; that it employed devices, schemes, and artifices to defraud; and that it would operate as a fraud and deceit upon purchasers and sellers of Massey Class A Common Stock (*United States of America v. Donald L. Blankenship, Superseding Indictment, United States District Court for the Southern District of West Virginia, Charleston Grand Jury* 2014)."

[11] "Mine Owners Misled Inspectors Before Blast, Investigators Say," *New York Times*, June 30, 2011.

[12] "Justice for Upper Big Branch," *New York Times*, February 28, 2012.

[13] "Justice for Upper Big Branch," *New York Times*, February 28, 2012.

[14] As a side note, institutionally-specific QNEs, such as the Upper Big Branch mining disaster, also create challenges for editors and referees with respect to developing a keen understanding of the research setting to ensure that tests are well specified, and that the authors have adequately controlled for potential correlated omitted variables.

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

of answering the research question. More precise critiques allow the researcher an opportunity to identify and potentially address the omitted variable, reverse causality issue, or concern about generalizability.

4. Other identifying assumptions

Although researchers often focus on the exogeneity assumption when motivating a QNE setting, it is important to note that even in the presence of exogenous variation, natural experiments do not necessarily produce accurate estimates of treatment effects. QNEs rely on additional identifying assumptions, beyond just conditional exogeneity (i.e., beyond the parallel trends assumption in difference-in-differences, the exclusion restriction in instrumental variables, or the local continuity assumption in regression discontinuity designs). For example, all experimental methods also require the stable unit value treatment assumption (SUTVA). Difference-in-differences methods additionally require the perfect compliance assumption (Bundell and Dias, 2009). Christensen et al. discuss the exogeneity assumption in detail, but do not discuss the SUTVA or the perfect compliance assumption.[15]

The SUTVA requires that the treatment status of the treated group does not affect the outcomes of the control population and vice versa. In the context of the Christensen et al. setting, the SUTVA would be violated, for example, if public mine safety has spillovers on private mine safety because of competition for workers, customers, etc. If these spillovers are positive, which would be the case if private mines compete on safety and emulate the behavior of public mines, the authors' results would underestimate the dissemination effect of MSD. If these spillovers are negative, which would be the case if private mines gravitate away from the behavior of public mines because the mining industry is segmented with respect to safety, the authors' results would overestimate the dissemination effect of MSD. Our point here is not that SUTVA is necessarily violated in the Christensen et al. setting, but rather that researchers should, in general, discuss the SUTVA and consider how a violation of the assumption would affect the interpretation of results.

Similarly, researchers estimating difference-in-differences models should consider the perfect compliance assumption. Perfect compliance is the assumption that no firms received the treatment in the pre-treatment period and that all firms in the treatment group—and only those firms—received the treatment in the post-treatment period. In Christensen et al., perfect compliance would occur if no public mines voluntarily disseminated mine safety violations in their accounting reports prior to the MSD regulation, no private mines voluntarily disseminated mine safety violations in accounting reports at any time, and no public mine owners opted out of MSD (say, by delisting). Although Christensen et al. do not discuss the perfect compliance assumption, a strength of their setting is that it appears to be characterized by near perfect compliance.

Nonetheless, perfect compliance is rare; most QNEs feature imperfect compliance. To illustrate issues that arise with imperfect compliance, we use the International Financial Reporting Standards (IFRS) adoption literature as an example. Some firms adopted IFRS prior to their country's mandate, and some firms adopted IFRS voluntarily in countries that did not mandate IFRS (Daske et al., 2008). We adopt the language commonly used in the econometrics literature and refer to these firms as "always compliers." Some firms in countries that required IFRS likely delisted or used loopholes to avoid the IFRS mandate (Pownall and Wieczynska, 2016). We refer to these firms as "never compliers." Finally, there are firms that adopted IFRS as the result of a mandate, but would not have done so otherwise. We refer to these firms as "marginal compliers." The presence of always compliers or never compliers means that compliance with IFRS mandates was imperfect (to be clear, the presence of either always compliers *or* never compliers is a sufficient condition for imperfect compliance).

Researchers interested in the effects of IFRS can compare the outcomes of firms that voluntarily adopt IFRS to those that do not. If there are no endogeneity issues, doing so will accurately estimate the effect of IFRS on always compliers. Alternatively, researchers can compare the change in outcomes for firms in countries that mandated IFRS to the change in outcomes for firms in countries that did not mandate IFRS. However, always compliers and never compliers do not change their IFRS behavior in response to mandates. As a result, the estimated treatment effect will be the causal effect of IFRS, but only for marginal compliers.[16]

The imperfect compliance aspect of IFRS is potentially a concern to the extent that IFRS conveys differing benefits to always compliers, marginal compliers, and never compliers. For example, to estimate the effects of IFRS on always compliers, authors have examined firms that adopted IFRS voluntarily. These authors typically find higher estimated benefits for these firms than for firms that adopted IFRS because of a mandate (e.g., Daske et al., 2008). One explanation for these differences is that firms adopt a variety of other mechanisms to improve transparency simultaneously with their adoption of IFRS, and these other mechanisms are correlated and omitted variables (e.g., Daske et al., 2008, Christensen, 2012).

[15] The parallel trends assumption is the assumption that the change in outcomes for the control group and the treatment group *would have been* the same in the absence of treatment. It is frequently "tested" by examining for differences in the pre-treatment trends (e.g., Fig. 1 of Christensen et al.). Such analyses can bolster confidence that the parallel trends assumption is valid, but they cannot formally test the assumption because the parallel trends assumption is an assumption about the unobservable counterfactual.

[16] When there is imperfect compliance, authors can estimate causal treatment effects as long as the exogeneity, SUTVA, and an additional monotonicity assumption are satisfied (e.g., Angrist et al., 1996; Heckman et al., 2006; Blundell and Dias, 2009). The monotonicity assumption requires that while the shock might not change the behavior of some affected firms, all of those who do change their behavior because of the shock change their behavior in the same direction (i.e., that there are no "defiers"). When there is imperfect compliance, but no endogeneity issues, instrumental variables estimates treatment effects for marginal compliers and difference-in-differences estimates a weighted average of the treatment effect for marginal compliers and no effect for always and never compliers (Angrist et al., 1996; Blundell and Dias, 2009). If the monotonicity assumption is violated, estimates of the treatment effect will be attenuated if the effects of ceasing and adopting the behavior are symmetric (Heckman et al., 2006).

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

Another plausible (non-mutually exclusive) explanation, however, is that the effects of IFRS are heterogeneous, and the benefits are greater for voluntary adopters. That is, economic theory suggests firms that expect the greatest net benefit from IFRS will be the most likely to voluntarily adopt IFRS. If this is the case, the benefits of IFRS for these always compliers will likely be greater than the benefits of IFRS for marginal adopters (Heckman et al., 2006).[17] The main point here is that QNEs with imperfect compliance create additional challenges for researchers attempting to generalize treatment effects beyond the effect on marginal compliers.

As a final note, our discussion in this section is not meant to criticize QNEs, but rather to encourage researchers to think carefully about the underlying assumptions of the research design. Further, the common critique that QNEs' estimate only the marginal treatment effect may not be a concern in settings where the marginal treatment effect is the most informative treatment effect. For example, regulators considering implementing IFRS in the United States are probably not very concerned with the effect of IFRS on always compliers. Rather, they want to understand the behavior and prevalence of never-adopters, the effect of IFRS on marginal compliers, and any general equilibrium effects. Similarly, treatment estimates are often most valuable to individuals and firms on the margin of a decision.

5. Concluding remarks

Readers familiar with empirical archival accounting research will no doubt appreciate that researchers often face tradeoffs between credible identification and generalizability. That is, broad sample studies that are the most generalizable often have thorny identification problems, and studies that are chosen specifically to address identification issues (e.g., QNEs) often draw conclusions that are not readily generalizable.

So, how should the prudent researcher proceed? Our view is that the literature is likely to be best served by addressing causal inference in a Bayesian manner, whereby multiple studies, using a variety of research designs and sample selection, are used to update researchers' priors on important topics and theories. Specifically, it seems uncontroversial that both broad sample empirical studies and more narrowly focused QNEs provide valuable information on a wide range of research questions, to say nothing of survey, structural, and other methods beyond the scope of this discussion. Broad sample studies provide information about pervasive associations between economic variables of interest, and as noted above, can use an array of econometric techniques to narrow concerns about correlated omitted variables, or at least narrow the list of correlated omitted variables that are likely to be of concern. And, importantly, broad sample studies often explore samples of firms or individuals that are of general interest to a wide audience.

QNEs, on the other hand, can sometimes achieve more convincing identification of causal effects, in part because they tend to be more narrowly focused, but often at the cost of generalizability. Researchers examining QNEs can often broaden the generalizability of their results, in much the same way that authors of broad sample studies can narrow concerns about correlated omitted variables. For example, they can examine additional features of their data, such as what type of firms comply with a shock and whether treatment estimates vary in the cross-section. They can also provide insight into the mechanisms that link cause and effect by testing key assumptions or additional predictions of the theoretical framework that motivates their analysis. Researchers can also pair broad sample evidence with a QNE in the same study, as in Guay et al. (2016). These efforts to address generalizability are extremely important, in large part because most accounting researchers are interested in understanding pervasive economic and behavioral phenomenon across large sectors of the economy.

So, returning now to Christensen et al., how could researchers interested in generalizing their results proceed? One possibility would be to explore similar information dissemination questions using broader samples. For example, one might consider examining certain types of press releases, press articles, or corporate website disclosures that contain information that is subsequently systematically disclosed in accounting reports. Obviously, one would need to be clever in choosing the types of information to examine (similar to the ways that researchers typically make clever choices about QNE settings), but doing so could allow one to examine the dissemination of financial, strategic, investing, or operational information in SEC filings across a heterogeneous set of firms.

Researchers might also identify additional QNEs that examine unrelated settings where SEC filings have the potential to increase the dissemination of information. The robustness of the Christensen et al. findings to other settings would increase confidence that the results are generalizable. However, follow-up studies that attempt to broaden the generalizability of previously published work can be extremely difficult to publish. Similarly, it can also be very difficult to publish follow-up studies that examine research settings that facilitate sharper identification of causal effects to explore the robustness of inferences in previously published broad sample studies. As discussed by conference participants, many journals are reluctant to publish papers that replicate previously documented relations in novel samples. Yet, without such replications, it is often impossible to have confidence that the findings in a single study are pervasive economic effects.

As a final point, we caution accounting researchers against being too quick to deem a research question definitively answered by a single broad-sample study with thorny identification issues, or by a single QNE study examining a narrow setting. Beyond simply encouraging a more open response to multiple papers that tackle important research questions from

[17] The key difference between reverse causality and imperfect compliance is that the former is the result of non-exogenous treatment, while the latter is the result of non-mandatory compliance with an exogenous treatment. When there is reverse causality the estimated treatment effect is biased because treatment and control groups are dissimilar. When there is imperfect compliance, the estimated treatment effect is an unbiased estimate of the effect on marginal compliers because the treatment and controls groups are similar.

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003

differing perspectives, some possible ways to further address this concern might include: 1) development of a journal, or perhaps a peer-reviewed website, dedicated to identification and generalizability extensions of previously published studies; 2) periodic special issues dedicated to identification and generalizability of previously published studies (championed by journals that would like to become recognized for contributing to this worthwhile effort), or 3) journals and authors of original studies agreeing to a process whereby published papers might later include addenda that offer identification and generalizability extensions written by other authors.

References

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91 (434), 444–455.
Armstrong, C.S., Core, J.E., Guay, W.R., 2014. Do independent directors cause improvements in firm transparency. J. Financial Econ. 113 (3), 383–403.
Bertomeu, J., Beyer, A., Taylor, D.J., 2016. From casual to causal inference in accounting research: the need for theoretical foundations. Found. Trends® Account. 10 (2–4), 262–313.
Blankespoor, E., Miller, G.S., White, H.D., 2013. The role of dissemination in market liquidity: evidence from firms' use of twitter™. Account. Rev. 89 (1), 79–112.
Blundell, R., Dias, M.C., 2009. Alternative approaches to evaluation in empirical microeconomics. J. Hum. Resour. 44 (3), 565–640.
Bushee, B.J., Core, J.E., Guay, W., Hamm, S.J., 2010. The role of the business press as an information intermediary. J. Account. Res. 48 (1), 1–19.
Chen, Q., Schipper, K., 2016. Comments and observations regarding the relation between theory and empirical research in contemporary accounting research. Found. Trends® Account. 10 (2–4), 314–360.
Christensen, H.B., 2012. Why do firms rarely adopt IFRS voluntarily? Academics find significant benefits and the costs appear to be low. Rev. Account. Stud. 17 (3), 518–525.
Christensen, H.B., Floyd, E., Liu, L.Y., Maffett, M.G., 2017. The real effects of mandated information on social responsibility in financial reports: evidence from mine-safety records. J. Account. Econ. forthcoming.
Daske, H., Hail, L., Leuz, C., Verdi, R., 2008. Mandatory IFRS reporting around the world: early evidence on the economic consequences. J. Account. Res. 46 (5), 1085–1142.
Drake, M.S., Roulstone, D.T., Thornock, J.R., 2012. Investor information demand: evidence from Google searches around earnings announcements. J. Account. Res. 50 (4), 1001–1040.
Drake, M.S., Roulstone, D.T., Thornock, J.R., 2015. The determinants and consequences of information acquisition via EDGAR. Contemp. Account. Res. 32 (3), 1128–1161.
Engelberg, J.E., Parsons, C.A., 2011. The causal impact of media in financial markets. J. Finance 66 (1), 67–97.
Gow, I.D., Larcker, D.F., Reiss, P.C., 2016. Causal inference in accounting research. J. Account. Res. 54 (2), 477–523.
Grabirel, T., 2014. Ex-executive is indicted in disaster at coal mine. New York Times A12.
Guay, W., Samuels, D., Taylor, D., 2016. Guiding through the fog: financial statement complexity and voluntary disclosure. J. Account. Econ. 62 (2), 234–269.
Heckman, J.J., Urzua, S., Vytlacil, E., 2006. Understanding instrumental variables in models with essential heterogeneity. Rev. Econ. Stat. 88 (3), 389–432.
Justice for the Upper Big Branch. [Editorial]. (2012, February 28). New York Times, p. A24.
Karpoff, J.M., Wittry, M.D., 2017. Institutional and legal context in natural experiments: the case of state antitakeover laws. J. Finance forthcoming.
Leuz, C., Wysocki, P.D., 2016. The economics of disclosure and financial reporting regulation: evidence and suggestions for future research. J. Account. Res. 54 (2), 525–622.
Peress, J., 2014. The media and the diffusion of information in financial markets: evidence from newspaper strikes. J. Finance 69 (5), 2007–2043.
Pownall, G., & Wieczynska, M. (2016). Deviations from the mandatory adoption of IFRS in the European Union: implementation, enforcement, incentives, and compliance. Working paper.
Roberts, M.R., Whited, T.M., 2011. Endogeneity in empirical corporate finance. In: Constantinides, G., Stulz, R., Harris, M. (Eds.), Handbook of the Economics of Finance. Elsevier, Amsterdam, the Netherlands.
Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Wadsworth Cengage Learning.
Some Justice at Upper Big Branch. [Editorial]. (2011, October 28). New York Times, p. A30.
Tavernise, S., 2011, June 30. Mine owners misled inspectors before blast, investigators say. New York Times A14.
United States of America v. D.L. Blankenship, Criminal Action No. 5:14-cr-00244 (S.D.W. Va., 2014).

Please cite this article as: S. Glaeser, W.R. Guay, Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), Journal of Accounting and Economics (2017), http://dx.doi.org/10.1016/j.jacceco.2017.08.003