



# CFS Working Paper Series

No. 703

Christian Leuz

## Towards a Design-Based Approach to Accounting Research

## The CFS Working Paper Series

presents ongoing research on selected topics in the fields of money, banking and finance. The papers are circulated to encourage discussion and comment. Any opinions expressed in CFS Working Papers are those of the author(s) and not of the CFS.

The Center for Financial Studies, located in Goethe University Frankfurt's House of Finance, conducts independent and internationally oriented research in important areas of Finance. It serves as a forum for dialogue between academia, policy-making institutions and the financial industry. It offers a platform for top-level fundamental research as well as applied research relevant for the financial sector in Europe. CFS is funded by the non-profit-organization Gesellschaft für Kapitalmarktforschung e.V. (GfK). Established in 1967 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia. GfK members comprise major players in Germany's financial industry. The funding institutions do not give prior review to CFS publications, nor do they necessarily share the views expressed therein.

# Towards a Design-Based Approach to Accounting Research\*

Christian Leuz

*University of Chicago Booth School of Business, CEPR & NBER*

August 2022

## Abstract

Armstrong et al. (2022) review the empirical methods used in the accounting literature to draw causal inferences. They document a growing number of studies using quasi-experimental methods and provide a critical perspective on this trend as well as the use of these methods in the accounting literature. In this discussion, I complement their review by broadening the perspective. I argue for a design-based approach to accounting research that shifts attention from methods to the entire research design. I also discuss why studies that aim to draw causal inferences are important, how these studies fit into the scientific process, and why assessing the strength of the research design is important when evaluating studies and aggregating research findings.

*Keywords:* Endogeneity, Causal inferences, Research design, Empirical methods, Natural experiments, Accounting research

*JEL classification:* C4, D8, M4

\*I thank Matthias Breuer, Anna Costello, Joao Granja, Delphine Samuels, Andrew Sutherland, Dan Taylor as well as the conference participants for helpful comments and discussions. All remaining errors are of course my own.

## 1. Introduction

To very briefly summarize, Armstrong, Kepler, Samuels and Taylor (2022; henceforth AKST) start from the observation that there is a rising trend towards studies drawing causal inferences in the accounting literature. AKST review the empirical accounting literature in the top-3 journals to document this trend and provide descriptive evidence of the methods that accounting researchers use to draw causal inferences. Their paper provides a critical perspective of this trend and the methods used in quasi-experimental studies in accounting research. AKST remind us of the assumptions of these methods and discuss practical considerations and tradeoffs in applying these methods. AKST highlight the challenges of drawing causal inferences. They emphasize the importance of theory and the need to triangulate methods and designs.

AKST make many important and well-taken points about the use of empirical research methods in accounting research. They also provide useful descriptive statistics on the use of quasi-experimental methods. What sets their review paper apart from prior surveys that the Journal of Accounting and Economics has published in recent years is that AKST do not focus on a particular topic but instead review the methods that accounting studies use. This “method lens” is a welcome addition. Overall, I agree with most of the authors’ points. Many of them have been made previously in applied econometrics textbooks, discussion papers on causal inferences or in prior literature reviews for particular topics. But it is nevertheless useful to put these points together in one review focused on the accounting literature and its practices. My discussion therefore tries to amplify AKST’s important messages, in part by recasting them from different angles, with the intention of helping them to have the impact they deserve.

I also have a few quibbles. In parts of their discussion, AKST draw stronger contrasts and distinctions than I see or find necessary. I worry that some of the negative undertones with respect

to studies attempting to draw causal inferences could have the effect of turning back the clock for accounting research, especially relative to other social sciences, such as economics and finance, which have more wholeheartedly embraced quasi-experimental methods. Of course, accounting is a distinct field with a focus on information, measurement and understanding of institutional detail and hence it faces distinct methodological tradeoffs. But we cannot ignore the fact that many of the advances in research designs and best practices in empirical research originate in economics and finance. More importantly, aiming for research designs that allow us to draw causal inferences is important because many theories are causal (or about cause and effect) and, as a result, require causal evidence to test them. I would also submit that many important accounting questions, especially the policy relevant ones, are indeed causal questions. In addition, causal estimates are important if we want to learn about effect magnitudes or elasticities.

Next, AKST state that they offer a framework on causal inferences that is rarely present in econometrics textbooks. My sense is that they do not really offer a new framework but instead are firmly building on the advances that I just mentioned. I see them as advocating for sound and theory-motivated empirical research that recognizes the challenges of drawing causal inferences. With their paper, AKST kick off an important discussion about our research practices, and I am confident that the paper will be very useful to Ph.D. students in accounting.

In my discussion, I broaden AKST's method lens and advocate for a design-based approach to empirical accounting research. This approach is common in microeconomics (Card, 2022). It puts front and center the question of how a given research design allows us to answer the research question in a convincing fashion. It recognizes that research design needs to be fit for purpose. Empirical methods are an important part of the research design, but causal inferences require much more than finding a "shock" or applying a quasi-experimental method out of the toolbox. The

design-based approach starts with carefully thinking about relevant counterfactuals and then focuses on isolating variation in the variable of interest through careful research design. The chosen design essentially defines the counterfactuals and the variation used to identify the estimated effect. The design-based approach also entails specification tests and bringing auxiliary data to bear to substantiate that the research design works in the intended way, which connects with a point that AKST make about the need for triangulation. Simply put, the design-based approach shifts attention from methods (e.g., the use of difference-in-differences or instrumental variables) to the entire research design. The aim is to design studies that can credibly answer the questions they pose.

An important point of my discussion is that the design-based approach not only requires theory but also a deep understanding of the institutional setting. Our institutional knowledge is critical for identification. It is not sufficient to find a setting that generates some variation in the variable of interest. We also need the institutional knowledge to evaluate how well the setting fits the theory being tested, to guide measurement choices, to recognize threats to our research design, and to evaluate and perhaps even rule out potential alternative explanations.

My discussion also provides another perspective on the trend towards more quasi-experimental designs that AKST document. In my mind, the increasing number of studies that attempt to draw causal inferences reflects not only advances in the quasi-experimental methods, but also the life cycle of research ideas. The relevant standards for research are dynamic and evolve as the literature matures. Early on studies are often more descriptive and simpler in their design. Later studies tend to use more sophisticated designs. It is hard to come up with completely novel questions and at the same time answer them in a credible way. Thus, the rising trend in quasi-experimental methods also reflects where we are in this life cycle in many areas of the accounting literature.

In light of this rising trend documented by ASKT, I discuss whether the underlying push for causal inferences creates distortions in the accounting literature. Consistent with arguments in AKST, I see the potential for some distortions (e.g., with respect to topics and questions). Still, we cannot abandon the aim for causal inferences. They are hard to achieve but necessary for scientific progress. A tight and convincing research design allows us to learn because it rules out alternative explanations or rejects theories.

That said, in practice, the quest for identification is not a binary matter, as AKST point out as well. Studies differ in how convincingly they can identify a particular effect and in how large (or relevant) the group is for which they can identify an effect. Therefore, in evaluating studies, it is important to consider the strength of the research design that generates the results. As List (2020) puts it: “Each study moves priors by an amount corresponding to its quality and the strength of priors.” However, our priors are not only informed by theory but also by evidence in prior studies. These studies often form the conventional wisdom, even when they have relatively loose designs, which creates the potential for research to become circular. Therefore, our priors should also reflect the strength of the design of prior work. It is a fallacy to think that “piling up” many studies with weak identification eventually allows us to draw causal inferences or leads to sound evidence that justifies strong priors against which we can gauge new work. I illustrate this issue with two examples. I further note that this issue is particularly pernicious when research findings are used to inform policy (Leuz, 2018). In this context, recovering causal relations is crucial.

## **2. Research trends: The life cycle of research and the credibility revolution**

AKST document a rising trend of studies with quasi-experimental methods in the top-3 accounting journals. They note that this trend is connected to an increase in studies attempting to draw causal inferences. With regard to the latter, they make two important points. First, they

emphasize the role of theory to draw causal inferences. Theory helps researchers to identify what likely are spurious correlations and to interpret meaningful correlations. Second, they point out that empirical studies can be useful even when they do not attempt to draw causal inferences. I agree wholeheartedly with both points and recast them for emphasis below. AKST also provide a citation analysis examining whether studies with quasi-experimental methods are cited as frequently as one would expect given their prevalence. They find that the proportion of highly cited papers using quasi-experimental methods is lower than their proportion among all empirical studies. In other words, quasi-experimental studies are less often highly cited. In my mind, both findings in AKST, i.e., the increase in quasi-experimental studies and their lower citation count, reflect two forces: (i) scientific progress and the life cycle of ideas and (ii) advances in empirical research designs. Let me discuss both forces.

The scientific process starts with an idea. This idea could be a new theory, a new research setting or a new data set. In the end, however, we need both theory and empirical analysis in order to make scientific progress. The two are symbiotic. AKST emphasize that descriptive evidence that does not make causal claims can very useful. I fully agree; not every empirical study needs to test a theory. However, most theories are about causes and effects. Testing these theories requires causal evidence.<sup>1</sup> We learn through falsification and the rejection of theories (Popper, 1959). In order to test (or reject) a theory, we need research designs that isolate the causal link and rule out other explanations, typically by holding all other factors constant. Thus, theories actually motivate empirical studies with causal inferences and the desire to draw causal inferences is largely driven

---

<sup>1</sup> One might object here that not all theories require causal tests. For instance, tests of the CAPM are typically performed without the use of quasi-experimental methods. However, the underlying theory is clearly causal in that it describes the link between beta and expected returns. This link in turn could be tested with a study that cleanly isolates variation in beta and observes corresponding changes in expected returns.



by the desire to make scientific progress.

Moreover, research follows a natural life cycle for a given idea or theory. Typically, one study introduces a new idea (or an observation) to the literature. Other studies follow. These follow-on studies might challenge the idea or corroborate it. Eventually, the literature matures and we (hopefully) settle the matter. Along the way, studies typically become more sophisticated and their designs tighter.<sup>2</sup>

Let me illustrate this life cycle for the proprietary cost theory, which states that proprietary costs reduce firms' incentives to disclose information voluntarily. This idea (theory) was formally introduced into the accounting literature by Verrecchia's seminal 1983 paper. Subsequently, many empirical papers have tested this idea. Let me highlight three of them: one from each of the subsequent decades. One of the first to empirically analyze the proprietary cost theory is Harris (1998). This paper provides a standard cross-sectional design and uncovers associations between measures of competition and managers' segment reporting choices that are consistent with the theory (or inconsistent with the null hypothesis that proprietary costs play no role for disclosures). The next paper I highlight in the life cycle is Berger and Hann (2007). They exploit a change in the segment reporting rules (SFAS 14 to SFAS 131) to examine two motives for managers to conceal segment profits: proprietary costs and agency costs. Their results are more consistent with agency costs being a factor in managers' reporting decisions. They provide only mixed evidence for the proprietary cost motive. The third study is Li, Lin and Zhang (2018). They exploit the staggered adoption of the inevitable disclosure doctrine (IDD) by U.S. state courts, a legal change that generates variation in firms' proprietary costs and is plausibly exogenous to their disclosure

---

<sup>2</sup> The same life cycle can also be observed for analytical papers in that later papers are often more refined or become more general, relaxing previous assumptions.

choices. Li et al. (2018) emphasize the use of a quasi-experimental design and make an explicit causal claim for the role of proprietary costs for the supply of disclosures. Thus, their paper is the type of study that AKST focus on in their review.

There are many more papers that test the proprietary cost theory and provide relevant evidence. However, these three empirical papers are sufficient to illustrate the natural progression of research designs in a literature. Tighter or more sophisticated designs come later in the life cycle and eventually the design is sufficiently tight for the authors to make (or for reviewers and editors to “accept”) causal claims. Obviously, these claims can later be revised by studies that have even better designs.

This life cycle and the corresponding natural progression of methods can explain AKST’s results in the citation analysis. The lower count of highly cited, quasi-experimental studies quite likely reflects their relative position in the life cycle of the proprietary cost theory, along with the fact that studies cite all key papers that came before them. Consistent with this explanation, Verrecchia (1983) boasts 4,676 Google Scholar cites, Harris (1998) has 694, Berger and Hann (2007) has 541, and Li et al. (2018) is cited 148 times.<sup>3</sup> We would expect such differences in the citation counts between the first word and the last word on a matter. I do not find this particularly surprising. To learn more from the citation analysis, the authors would have to adjust for age of the idea by the time an article is published and the natural citation patterns that arise due to the described life cycle of successful research ideas.

Importantly, the lower citation count does not imply that we learn less from the later papers or

---

<sup>3</sup> The respective Google Scholar citations were taken as of August 23, 2022.

from studies with causal inferences.<sup>4</sup> The later papers are important in several ways. They tighten our posterior beliefs and help us settle questions in the literature. For this reason, there is typically a close connection between a later paper's research design and its incremental contribution to the literature. The standards for research designs evolve and tend to go up as the literature matures. Moreover, later studies often allow us to learn about the effect magnitudes. Without a tight design that affords a causal estimate, the magnitude of the estimated effect is of limited relevance.

A second force behind the rise of quasi-experimental methods and causal claims documented by AKST are advances in economics and econometrics. Angrist and Pischke (2010) referred to these advances and the widespread application of quasi-experimental methods (or natural experiments) as the “credibility revolution” in applied economics. In fact, the evidence in AKST on the rise of quasi-experimental methods mimics the trends that Panhans and Singleton (2017) document with their bibliometric analysis of the economics literature, except that there, the trend starts and accelerates much earlier, around 1990 and after 2000, respectively. Panhans and Singleton (2017) attribute these trends to individual pioneers like Theodore Schultz and Orley Ashenfelter but also the influential work by Joshua Angrist, David Card and Alan Krueger, all of whom are closely associated with the credibility revolution.

In his commentary on the 2021 Nobel Prize, awarded to Angrist, Card and Imbens, Pischke (2021) states: “Important questions in economics are causal questions.” I would make the same statement for accounting, subject to two clarifications. First, I do not believe *all* important questions are causal.<sup>5</sup> Second, we do not necessarily need causal evidence before we can tackle an

---

<sup>4</sup> In fact, some early papers have high citation counts because later papers have identified flaws in the approach that the early paper proposed. Jones (1991) and the estimation of discretionary accruals are a good example here, which also illustrates that citation counts are an imperfect measure of how much we learn from a paper.

<sup>5</sup> For instance, with the rise of machine learning, prediction has recently become an important goal of research as well. Prediction can be aided by an understanding of causal links but prediction models do not seek to identify causal

important question.<sup>6</sup> Answering important questions convincingly is hard and hence it often takes time for researchers to find ways or research designs to provide convincing answers to such questions. Moreover, given the causal nature of most theories, we need an empirical toolkit that allows us to test these theories. For this reason alone, quasi-experimental methods are here to stay and we should welcome their rise in accounting.

Of course, the widespread application of these methods has not been without debate and criticism (Sims, 2010; Deaton, 2010; Heckman and Urzua, 2010). Consistent with this more critical debate, AKST emphasize that quasi-experimental methods are not a panacea. Theory is still important in guiding the empirical analysis and the methods alone are not sufficient for causal inferences. In the next section, I argue that the design-based approach in economics encompasses these important points. Moreover, the design-based approach shifts attention from a narrow focus on methods to a broader perspective that focuses on designing studies that can credibly answer the questions they pose.

Importantly, however, the methodological advances in estimating causal effects and the rise of studies that can credibly draw causal inferences do not mean that we need only such studies, or that we should always strive to estimate causal effects or draw causal inferences. As AKST point out, descriptive studies can be important. For instance, documenting correlations in the data can lead to new theories that seek to explain these correlations. Descriptive evidence can give us a sense for how important or common a certain phenomenon (e.g., a corporate practice, a contract,

---

links. See Kleinberg et al. (2015), Mullainathan and Spiess (2017), and Athey (2018). In fact, fraud prediction has a long tradition in accounting research.

<sup>6</sup> For example, Ball and Brown (1968) tackle one of the most fundamental questions in accounting but do not seek to show a causal relation between earnings and returns. That said, it is based on a causal link. The core idea is that both earnings and returns reflect information about a firm's fundamentals, i.e., changes in fundamentals cause both changes in earnings and stock returns. Moreover, Ball and Brown (1968) use research design to rule out other explanations for the contemporaneous associations between earnings and returns.

etc.) is. Thus, it matters again where we are in the life cycle of a particular literature. Early on, we need to know key facts and the lay of the land. Descriptive papers can provide such evidence for a broad sample or the population and allow other studies to build on these facts. In contrast, quasi-experimental designs often (have to) focus on narrower samples and particular settings in order to achieve identification.

At the same time, we have to be careful that we do not over-interpret the early papers in a literature, especially if they provide primarily descriptive evidence. These studies still appeal to causal theories for the interpretation of their estimated associations and hence imply that we are learning about a causal effect. As an editor, I sometimes see reviewers stating that a new paper does not offer an incremental contribution because we already “know” from prior studies that X has an effect on Y. In making such a statement, it is important to evaluate the strengths of the research designs of the prior studies as well as the new paper. A new paper can make an important contribution by tightening our posteriors about an effect or by providing estimates for the magnitude of an effect. Having several prior studies finding evidence consistent with an effect does not necessarily mean we “know” that the effect exists. This is an important point to which I come back.

### **3. The Design-Based Approach to Identification: The Role of the Institutional Setting and Fixed Effects**

Fundamentally, AKST debate the role that identification plays in the accounting literature. Conceptually, identification is about isolating an explanation or an effect. It amounts to ruling out alternative explanations, so that we can home in on one explanation and draw a causal inference. However, sometimes, this is not feasible and we end up with more than one plausible explanation. Depending on the remaining set of explanations, such a study can still be insightful, provided we

acknowledge and discuss these potential explanations. In this sense, identification is more a goal and, in practice, not necessarily a binary matter.

There are many identification strategies. The ideal (from the perspective of identification) is to use randomization, as in a randomized control trial or a lab experiment. Sometimes we have truly random variation (e.g., weather, lotteries) and we can “run” a natural experiment with archival data. But these instances are rare. Quasi-experimental settings (e.g., regulatory changes, variation in regulation across firms, stock index inclusions, etc.) typically do not give us truly (or as if) random variation in the variable of interest. The latter means that most of the time we have to do additional work in order to draw causal inferences. For instance, we add controls and fixed effects, use econometric techniques such as instrumental variables, conduct additional analyses to explore the mechanism or to rule out alternative explanations, or we use theory to differentiate between more and less plausible explanations. All this amounts to what David Card (2022) calls a “design-based approach.” I adopt this term because it emphasizes that identification is not just about using a particular method or a “shock.”<sup>7</sup> The term emphasizes that we need a careful research design to rule out alternative explanations. At the heart of this approach are (i) thinking about counterfactuals and (ii) asking what (experimental or quasi-experimental) variation would provide a credible answer to the research question.

A key element of the design-based approach that deserves special mention is institutional knowledge.<sup>8</sup> From the perspective of identification, the institutional setting serves essentially as an “instrument” – it provides the variation in the variable of interest that we exploit for

---

<sup>7</sup> Card coined this term for studies that use simplified one-equation (or reduced-form) models and uses it “in contrast to model-based studies that specify a data generating process for all factors determining the outcome.” I use this term to denote archival studies that use research design for identification with the aim to credibly answer the research questions they pose. The more relevant contrast in accounting are probably studies that do not seek identification and are more descriptive in nature (e.g., establish facts), rather than model-based or structural studies.

<sup>8</sup> The following two paragraphs rely heavily on earlier discussion of this matter in Leuz and Wysocki (2016).

identification. Our institutional knowledge in turn tells us which explanations are more plausible or which threats to identification we face (e.g., the existence of concurrent events). It is critical for judging whether the exclusion restriction or the parallel trends assumption are satisfied, as they cannot be formally tested. Thus, institutional knowledge is critical for ruling out alternative explanations. A good design-based study invests heavily in understanding the intricacies of the institutional setting that it uses. As Leuz and Wysocki (2016) put it, “[a]rticulating why a particular setting and design provides proper identification of the economic effect as well as appreciating the potential threats to identification requires a deep understanding of the institutional setting.” This is something that should come naturally to accounting researchers as our field studies institutions and often focuses on changes in regulation or accounting standards.

Understanding the institutional setting is also important when interpreting empirical results. The setting is not only what often affords identification, but it also determines the (treatment) group for which a causal effect is estimated. The latter is a key factor for generalizability of the results. The group for which we can identify the effect could be quite narrow, making the estimated treatment effect fairly local (Angrist and Imbens, 1994). A good example is the regression discontinuity design, which estimates the effect using firms or individuals that are (very) close to the threshold that determines treatment. The estimated effect is therefore very local and could be quite different than the effect for firms or individuals that are far from the threshold.

Similarly, the estimated effect of mandatory IFRS reporting on the information available to investors could be quite different in countries that chose to adopt IFRS versus countries that chose not to adopt IFRS. Importantly, and in contrast to some discussion in AKST, the choice to adopt IFRS by country X does not impede our ability to estimate a causal effect for firms in country X on which the reporting mandate is imposed. With proper design, we can still use the mandate to

recover a causal effect of IFRS reporting in adopting countries.<sup>9</sup> The fact that countries endogenously chose to adopt IFRS does not per se bias the estimated effect for firms on which IFRS reporting was imposed, but it limits the group for which we can recover an effect and hence the interpretation of the results (Christensen et al., 2013). For instance, the estimated effects might not generalize to firms in countries that chose not to adopt. Our institutional understanding (e.g., the reasons some countries adopted and others did not) can help us evaluate the extent to which the results can be extrapolated to other countries or other settings. Theory is very important in this regard as well; it is a key tool when generalizing results. Furthermore, theory and institutional knowledge can help assess issues like selection into treatment (e.g., through voluntary adoption). In essence, the design-based approach combines theory, institutional knowledge and quasi-experimental methods to provide an answer to the research question posed.

Another topic that features prominently in AKST and deserves further discussion are fixed effects. They are a very powerful tool because they allow us to control for unobserved or unobservable confounding factors. That is, we can rule out alternative explanations without necessarily having the right data or proxies for them.<sup>10</sup> AKST point out that fixed effects are not a panacea and that they have to be used judiciously. I concur. I also do not object to their suggestion to report variance-inflation factors or the amount of variation that is absorbed by the fixed effects.

However, my take on fixed effects and their role in the design-based approach is a bit different from AKST. In my mind, the amount of variation that is absorbed by the fixed effects is of secondary importance. From a design perspective, it is okay to “throw out a lot of variation” when

---

<sup>9</sup> As Angrist and Imbens (1994) point out, there are many different treatment effects (e.g., the effect on the treated, the average treatment effect or local average treatment effects).

<sup>10</sup> The logic is that the fixed effects absorb potentially confounding variation across groups and once the confounding factor is constant within a group, then we no longer have to control for it. See also Breuer and deHaan (2022).



this variation does not afford identification because it is confounded by a number of factors or variables. The design perspective asks which variation would allow us to isolate the effect or provide a test of our theory. If the identification strategy suggests high-dimensional fixed effects and it turns out that this structure absorbs most of the variation and leads to the multicollinearity problem that AKST describe, then the conclusion should not be that a coarser fixed effect structure is in order, but that there is not enough non-confounded variation to learn from. Put differently, the setting does not offer enough information to identify the effect.<sup>11</sup>

The design perspective also implies that we need theory and institutional knowledge to determine which fixed effects should be included in the model. The choice of the fixed effects is an integral part of our research design. The goal when choosing the fixed effects is to home in on variation that we can interpret in the context of the theory or hypothesis. Thus, the study's research design dictates which structure we should use. For this reason, it is not obvious that we should encourage researchers to (mechanically) explore a host of fixed effect structures for robustness.<sup>12</sup> For the same reason, it is not clear that it is necessarily better to have a tighter (more saturated) fixed effects structure. It could make matters worse as AKST point out.

Let me illustrate the point of the previous paragraph with an example. Christensen et al. (2013) perform a liquidity analysis around IFRS adoption using country-specific quarter-year fixed effects. The underlying rationale is that these fixed effects exploit variation in firms' fiscal year ends and hence homes in on reporting effects. They eliminate the influence of time-period effects (e.g., country level shocks and other concurrent regulatory changes) that apply to all firms in a

---

<sup>11</sup> For this reason, sufficiently granular data are often the limiting factor with respect to identification. See Leuz (2018) for more discussion.

<sup>12</sup> This practice could help with respect to concerns about p-hacking and knife-edge results. However, if the results differ across fixed effect structures, it is again theory and institutional knowledge that tell us which results matter.

given country and month. Such confounding effects are not necessarily observable or known and hence difficult to control for directly. Thus, Christensen et al. (2013) use quarter-year fixed effects to absorb this variation and instead focus on variation that comes from when firms release their financial statements, which are plausibly reporting effects.

This difference in perspectives on fixed effects (i.e., remaining variation vs. the role of variation in identification) is also discussed by Breuer and deHaan (2022). On one hand, fixed effects reduce the available variation and hence statistical power. Therefore, the inclusion of fixed effects can lead to wider confidence intervals and increase the likelihood of false negatives as AKST point out. On the other hand, fixed effects reduce concerns about spurious explanations and hence decrease the likelihood of false positives. They can therefore even increase the power to detect true positives. From a design perspective, ruling out spurious explanations is likely a bigger deal than the concerns about multicollinearity raised in AKST.

In sum, good research design and drawing causal inferences requires a lot more than just a shock or an extensive fixed effects structure. The design-based approach to causal inferences highlights this point. It employs theory and institutional knowledge and considers mechanism tests, placebo or falsification tests, triangulation, and more. Ultimately, good design is more art than science. There is no cookbook or checklist that one can follow. The appropriate design depends on the research question and the research setting.

#### **4. Is there a distortion from the rise of quasi-experimental studies?**

An important question that AKST pose with their critical perspective on the rising trend of quasi-experimental studies is whether the emphasis on methods and causal inferences creates distortions in the accounting literature. In debating this question, we first have to recognize that there is a natural and well-known tradeoff between internal and external validity. Restricting the

relevant variation in the variable of interest for the purpose of identification often comes at the cost of generalizability. However, this tradeoff is not a distortion.

A distortion could arise if the quest for identification crowds out descriptive studies or broad-based sample studies without a clear identification strategy. It could also distort the questions that researchers seek to answer.<sup>13</sup> In this regard, AKST note that it is better to have a stream of imperfect papers on an important causal question than to ignore such a question for lack of quasi-experimental evidence. Angrist and Pischke (2010) posed essentially the same question: “Critics [...] argue that in pursuit of clean and credible research designs, researchers seek good answers instead of good questions.” This question has been heavily debated in economics as the caricature by Scheiber (2007) illustrates.<sup>14</sup>

Considering that the rise of quasi-experimental methods came much later in accounting than in neighboring fields, I am not yet concerned. My sense is that, on balance, the push for identification and the rise of quasi-experimental studies has been a positive development. As argued earlier, these trends are a sign of scientific progress and a natural development considering that the accounting literature has matured substantially in a number of areas. I also do not think that accounting researchers have stopped asking good or hard questions. If anything, we have broadened the domain of accounting research in recent years by studying a wide array of non-traditional settings and questions.<sup>15</sup> I would argue that this broadening has made accounting research more exciting and is likely to connect accounting with a larger set of related fields. Moreover, as Leuz and Wysocki (2016) point out, non-traditional settings can also offer the

---

<sup>13</sup> Yet another concern is the repeated use of the same experimental setting for several outcome variables (e.g., Heath et al., 2020).

<sup>14</sup> For perspective, see also the answers to a poll of the IGM’s US Economic Experts Panel on the 2021 Nobel prize (<https://www.igmchicago.org/surveys/natural-experiments-in-labor-economics-and-beyond-2/>).

<sup>15</sup> To provide a few, non-systematically selected examples: Christensen et al. (2017), Sutherland (2018), Kim and Valentine (2021), Samuels (2021), She (2021), Fiechter, Hitz, and Lehmann (2022), Tomar, 2022.

opportunity to study links that are relevant to accounting but are difficult to study (or isolate) in more traditional settings.

Pushing back on the broadening of the domain and the use of quasi-experimental methods could undo these positive developments and decouple accounting research from other fields. Moreover, we see similar trends towards quasi-experimental methods not just in economics and finance, but also in the political and social sciences more generally. Thus, turning back the clock on these advances would be a mistake. At the same time, I agree with AKST that we have to become sophisticated users of quasi-experimental methods and appreciate the challenges of causal inferences. It is for this very reason that I advocate for a broader design-based approach.

To be sure, embracing these developments does not mean that we should only pursue studies with tight research designs. As discussed earlier, we can learn a great deal from descriptive papers. A good example to illustrate the usefulness of descriptive evidence is the fair value accounting debate during and after the financial crisis. Initially, many were concerned that the use of fair value accounting accelerated the crisis through fair value losses leading to fire sales. However, descriptive evidence on the use of fair value on banks' balance sheets, the magnitude of their fair-value losses as well as relatively simple analyses of banks' sales of securities was able to cast serious doubt on this explanation (Laux and Leuz, 2010; Badertscher et al., 2012).

## **5. Model Uncertainty, the Conventional Wisdom and the Perils of Piling Up Studies**

AKST highlight that the data generation process is unknown and that therefore we should be more Bayesian in how we evaluate empirical studies and interpret empirical evidence (see also Glaeser and Guay, 2017). As discussed earlier, I very much agree with this point. What I would like to add to the discussion is that we have to recognize that we face model uncertainty. In essence, we are uncertain about the theory (i.e., how the world works) but also the correct empirical model

to test the theory (e.g., Ausperg and Brüderl, 2021).<sup>16</sup> The latter implies that the plausibility of the empirical model is a key ingredient in drawing inferences and updating our priors. Tighter research designs require weaker identifying assumptions for the empirical model. Conversely, weaker designs require stronger identifying assumptions. For instance, a simple linear regression with a few controls and industry fixed effects estimated for a broad sample might yield tight confidence intervals. But the tight confidence intervals do not mean much if the model is likely misspecified because it requires implausible assumptions for identification (i.e., that any confounding variation is captured by the controls and the industry fixed effects).<sup>17</sup> A tight research design that uses less variation and a narrower sample likely yields larger confidence intervals, but it can still have more power to identify the true effect because it uses cleaner variation and requires fewer assumptions.

Another key ingredient in evaluating empirical studies are our priors. If we have strong priors, empirical evidence generated by a relatively weak design is unlikely to move our priors much. Thus, in this instance, we need tight designs to make progress and learn something. This point connects with my earlier discussion that late in the life cycle of a research idea the tightness of the research design and the incremental contribution are closely connected.

Furthermore, our priors are informed not just by economic intuition and theory but also by basic correlations in the data and prior empirical studies. Therefore, if a descriptive study provides new evidence based on a broad sample that largely confirms the conventional wisdom about a relation, we should not update our beliefs much, given the circularity between our priors and the data. Similarly, we have to be careful how much weight we give to our priors that are formed based on this study. Imagine that later a tight, design-based study comes along and its evidence goes against

---

<sup>16</sup> I thank Matthias Breuer for helpful discussions of this issue.

<sup>17</sup> See Breuer and Schütt (2021) for an illustration in the context of accrual models.

the conventional wisdom. If we now give a lot of weight to our priors, we might reject the new study as implausible, which can obviously be problematic.

My point is that it is perfectly fine to let studies without tight designs into the literature, as long as we are willing to update our priors based on later studies with more convincing designs, even if this evidence challenges the conventional wisdom. Remember the conventional wisdom and our underlying assumptions can be wrong.<sup>18</sup> The promise of tight, design-based studies is that they can provide compelling evidence on a relation or an effect, which in turn allows us to learn (either by further tightening our priors or by rejecting the conventional wisdom). We learn from these studies precisely because they rule out the alternative explanations.

As discussed earlier, tight designs can come at the cost of generalizability, as they often focus on specific settings or narrower samples. But in the above context, it is actually helpful when studies consider new quasi-experimental settings (e.g., a new legal change). Different settings bring new information, i.e., new experimental variation to the table. In this case, the uniqueness of the setting is actually a strength (List, 2020). In contrast, revisiting previously studied links in the same or a similar broad sample without careful identification cannot really challenge the conventional wisdom and is unlikely to move our priors by much.

This discussion brings me to my last point. AKST argue that “[i]f five decades of observational evidence on a particular theory exist, then the reader likely has well-defined priors” and furthermore that “[a]lternative explanations can be ruled out over time through collection of studies.” These statements invoke the notion that over time studies can reinforce each other and solidify our evidence. I do not disagree with these statements and frankly some of my earlier points

---

<sup>18</sup> Two good examples are Card’s studies using the Mariel boatlift (Card, 1990) or the New Jersey minimum wage setting (Card and Krueger, 1994; Card and Krueger, 2000).

on identification are quite similar. However, it is important to see the perils of aggregating studies without causal evidence (or what AKST call imperfect studies). I provide two examples to illustrate the fallacy that having many studies with similar results implies sound evidence.

The first example is the famous J-curve describing the relation between alcohol consumption and health outcomes, in particular mortality. The curve shows that mortality at first declines for occasional and moderate drinkers, relative to abstainers, before it then increases steadily with the amount of alcohol consumption. In other words, the J-curve suggests health benefits to occasional and moderate alcohol consumption. There are over 2,600 observational studies on this link as well as a series of meta-analyses. Much of the evidence is consistent with the J-curve and its above interpretation. In fact, the epidemiological and the physiological evidence was considered sufficiently compelling to consider recommending abstainers to drink (Ronksley et al., 2011; Brien et al., 2011). However, more recent studies question this curve. For instance, Stockwell et al. (2016) point out that many observational studies insufficiently account for the selection issue related to abstaining from alcohol consumption. People who do not drink often have (other) health reasons for doing so. In addition, there are large life-style differences (e.g., eating and exercise) between abstainers, moderate drinkers and those that drink more. Once Stockwell et al. (2016) adjust for abstainer bias and various quality-related study characteristics, the health benefits from moderate drinking disappear and the J-curve straightens to a linear relation, with mortality risk increasing as alcohol consumption increases. The point is that, in this case, even over 2,600 observational studies did not solidify our evidence because essentially all studies suffer from the same design challenges and very similar selection problems.

My second example makes a similar point but in the context of accounting. We easily have more than 100 studies on the effects of mandatory IFRS adoption around the world. Many of these

studies document considerable benefits associated with IFRS adoption and their overall message of their results is quite consistent (De George, Li, and Shivakumar, 2016). However, all these studies face the same challenges: IFRS adoption by most countries is highly clustered in time. In 2005, the majority of the adopting countries were member states of the European Union, which passed a series of financial market directives around the same time IFRS reporting became mandatory (Daske et al., 2008; Christensen et al., 2013). These other directives are significant confounders in most IFRS studies (Leuz and Wysocki, 2016). Thus, aggregating over a large number of studies in this context will again be of little help.

Meta-analysis and aggregation can be helpful if we have a series of randomized control trials with different populations or different designs, as is often the case in medical studies. Similarly, triangulation is valuable, but it requires using different settings and designs, with different design challenges, to be useful. But in the absence of causal estimates, the number of studies with consistent results can be quite misleading and overstate how confident we can be about the evidence. Thus, in aggregating studies and evidence, it is important to not only look at the results (i.e., the estimates for the relation of interest) but also to consider the strength of the designs of these studies.<sup>19</sup>

As the first example aptly illustrates, we have to be particularly careful about the aggregation of research findings when it comes to policy advice or policymaking. Causal inferences are critical for policy, not only because policymakers often care about effect magnitudes. Without defining counterfactuals properly, it is difficult to evaluate policies. Moreover, before we recommend a

---

<sup>19</sup> In this regard, it is interesting to consider Cochrane Reviews used in medicine to aggregate evidence on a particular question. Such a review follows a protocol that starts with searching for all the existing studies and identifying the relevant ones. The review then evaluates each relevant study and provides “risk of bias judgments.” The summary of the findings for each outcome or effect comes with a grade on the “certainty of evidence.” See Normansell et al. (2018) as an example.



policy, we want to be sure that the policy lever X truly matters for the intended outcome Y. We would not want to issue health recommendations without causal evidence and a plausible causal mechanism. Although the stakes are likely higher in medicine, the same principle applies to reporting regulation and accounting standard setting.

## **6. Conclusion**

There is little doubt that we have seen a marked shift in methods and research designs in the accounting literature. Similar developments have taken place in economics and finance. The key question is what we make of these developments. As AKST point out, there are many tradeoffs and hence the developments are unlikely to be unequivocally good or bad. It is easy to see problematic outcomes in either direction. For instance, on one hand, there is the risk of misinterpreted or misused correlations from studies without identification, especially in policy decisions. On the other hand, we might have studies chasing “cute” settings that deliver precise estimates for fairly small questions. But if we want to make progress as a field, then there is no question in my mind that it is worth pushing for tight designs and causal inferences, while recognizing – as AKST point out – how challenging the latter can be.

The late Alan Krueger once said: “The idea of turning economics into a true empirical science, where core theories can be rejected, is a big, revolutionary idea.” This ideal is obviously very hard to achieve in economics and in accounting, but it is nevertheless worth trying.

## REFERENCES

- Angrist, J. D., Imbens, G. W., 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62 (2), 467–475.
- Angrist, J. D., Pischke, J. S., 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *J. Econ. Perspect.* 24 (2), 3-30.
- Armstrong, C., Kepler, J. D., Samuels, D., Taylor, D., 2022. Causality redux: The evolution of empirical methods in accounting research and the growth of quasi-experiments. *J. Account. Econ.* 101521.
- Athey, S., 2018. The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*. University of Chicago Press, Chicago, pp. 507-547.
- Ausperg, K., Brüderl, J., 2021. Has the credibility of the social sciences been credibly destroyed? Reanalyzing the “many analysts, one data set” project. *Socius* 7, 1-14.
- Badertscher, B. A., Burks, J. J., Easton, P. D., 2012. A convenient scapegoat: Fair value accounting by commercial banks during the financial crisis. *Account. Rev.* 87 (1), 59-90.
- Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. *J. Account. Res.* 6 (2), 159-178.
- Berger, P. G., Hann, R. N., 2007. Segment profitability and the proprietary and agency costs of disclosure. *Account. Rev.* 82 (4), 869-906.
- Breuer, M., deHaan, E., 2022. Using and interpreting fixed effects models. Working paper.
- Breuer, M., Schütt, H., 2021, *Accounting for Uncertainty: An Application of Bayesian Methods to Accruals Models*, *Review of Accounting Studies*, forthcoming.
- Brien, S. E., Ronksley, P. E., Turner, B. J., Mukamal, K. J., Ghali, W. A., 2011. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *Bmj* 342, d636.
- Card, D., 1990. The impact of the Mariel boatlift on the Miami labor market. *ILR Rev.* 43 (2), 245-257.
- Card, D., 2022. Design-Based Research in Empirical Microeconomics. *Am. Econ. Rev.* 112 (6), 1773-81.
- Card, D., Krueger, A. B., 1994. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *Am. Econ. Rev.* 84 (4), 772-793.
- Card, D., Krueger, A. B., 2000. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: reply. *Am. Econ. Rev.* 90 (5), 1397-1420.
- Christensen, H. B., Hail, L., Leuz, C., 2013. Mandatory IFRS reporting and changes in enforcement. *J. Account. Econ.* 56 (2-3), 147-177.
- Christensen, H. B., Floyd, E., Liu, L. Y., Maffett, M., 2017. The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records. *J. Account. Econ.* 64 (2-3), 284-304.
- 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.
- Deaton, A., 2010. Instruments, randomization, and learning about development. *J. Econ. Lit.* 48 (2), 424-55.
- De George, E. T., Li, X., Shivakumar, L., 2016. A review of the IFRS adoption literature. *Rev. Account. Stu.* 21 (3), 898-1004.

- Fiechter, P., Hitz, J. M., Lehmann, N., 2022. Real effects of a widespread CSR reporting mandate: Evidence from the European Union's CSR Directive. *J. Account. Res.* 60 (4), 1499-1549.
- Glaeser, S., Guay, W. R., 2017. Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017). *J. Account. Econ.* 64 (2-3), 305-312.
- Harris, M. S., 1998. The association between competition and managers' business segment reporting decisions. *J. Account. Res.* 36 (1), 111-128.
- Heath, D., Ringgenberg, M., Samadi, M., Ingrid M. Werner., 2020. Reusing natural experiments. Working paper.
- Heckman, J. J., Urzua, S., 2010. Comparing IV with structural models: What simple IV can and cannot identify. *J. Econom.* 156 (1), 27-37.
- Jones, J. J., 1991. Earnings management during import relief investigations. *J. Account. Res.* 29 (2), 193-228.
- Kim, J., Valentine, K., 2021. The innovation consequences of mandatory patent disclosures. *J. Account. Econ.* 71 (2-3), 101381.
- Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Prediction policy problems. *Am. Econ. Rev.* 105 (5), 491-95.
- Laux, C., Leuz, C., 2010. Did fair-value accounting contribute to the financial crisis?. *J. Econ. Perspect.* 24 (1), 93-118.
- Leuz, C., 2018. Evidence-based policymaking: promise, challenges and opportunities for accounting and financial markets research. *Account. Bus. Res.* 48 (5), 582-608.
- 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.
- Li, Y., Lin, Y., Zhang, L., 2018. Trade secrets law and corporate disclosure: Causal evidence on the proprietary cost hypothesis. *J. Account. Res.* 56 (1), 265-308.
- List, J. A., 2020. Non est disputandum de generalizability? A glimpse into the external validity trial (No. w27535). National Bureau of Economic Research.
- Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. *J. Econ. Perspect.* 31 (2), 87-106.
- Normansell, R., Sayer, B., Waterson, S., Dennett, E. J., Del Forno, M., Dunleavy, A., 2018. Antibiotics for exacerbations of asthma. *Cochrane Database Syst. Rev.* 6.
- Panhans, M. T., Singleton, J. D., 2017. The empirical economist's toolkit: from models to methods. *Hist. Poli. Econ.* 49 (Supplement), 127-157.
- Pischke, J. S., 2021. Natural experiments in labour economics and beyond: The 2021 Nobel Laureates David Card, Joshua Angrist, and Guido Imbens. <https://cepr.org/voxeu/columns/natural-experiments-labour-economics-and-beyond-2021-nobel-laureates-david-card/> (Accessed 25 August 2022).
- Popper, K. (1959). *The logic of scientific discovery*, Routledge, New York.
- Ronksley, P. E., Brien, S. E., Turner, B. J., Mukamal, K. J., Ghali, W. A., 2011. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *Bmj* 342, d671.
- Samuels, D., 2021. Government Procurement and Changes in Firm Transparency, *The Accounting Review*, 96 (2021), 401-430.
- Scheiber, N., 2007. Freaks and geeks: How freakonomics is ruining the dismal science. *The New Republic.* 2, 27-31.

- She, G., 2021. The real effects of mandatory nonfinancial disclosure: evidence from supply chain transparency. *Account. Rev.*
- Sims, C. A., 2010. But economics is not an experimental science. *J. Econ. Perspect.* 24 (2), 59-68.
- Stockwell, T., Zhao, J., Panwar, S., Roemer, A., Naimi, T., Chikritzhs, T., 2016. Do “moderate” drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. *J. Stud. Alcohol. Drugs.* 77 (2), 185-198.
- Sutherland, A., 2018. Does credit reporting lead to a decline in relationship lending? Evidence from information sharing technology. *J. Account. Econ.* 66 (1), 123-141.
- Tomar, S., 2022. Greenhouse gas disclosure and emissions benchmarking. SMU Cox School of Business Research Paper. 19-17.
- Verrecchia, R. E., 1983. Discretionary disclosure. *J. Account. Econ.* 5, 179-194.

## Recent Issues

All CFS Working Papers are available at [www.ifk-cfs.de](http://www.ifk-cfs.de).

No.	Authors	Title
702	Christian Leuz, Anup Malani, Maximilian Muhn, and Laszlo Jakab	Do Conflict of Interests Disclosures Work? Evidence from Citations in Medical Journals
701	Christian Andres, Dmitry Bazhutov, Douglas J. Cumming, and Peter Limbach	Does Speculative News Hurt Productivity? Evidence from Takeover Rumors
700	Douglas J. Cumming and Pedro Monteiro	Sovereign Wealth Fund Investment in Venture Capital, Private Equity, and Real Asset Funds
699	Douglas J. Cumming and Pedro Monteiro	Hedge Fund Investment in ETFs
698	Dirk Krueger and Harald Uhlig	Neoclassical Growth with Long-Term One-Sided Commitment Contracts
697	Harold L. Cole, Dirk Krueger, George J. Mailath, and Yena Park	Trust in Risk Sharing: A Double-Edged Sword
696	Amy Whitaker And Roman Kräussl	Art Collectors as Venture Capitalists
695	Roman Kräussl, Kalle Rinne, and Huizhu Sun	Does Family Matter? Venture Capital Cross-Fund Cash Flows
694	Roman Kräussl, Tobi Oladiran, and Denitsa Stefanova	A Review on ESG Investing: Investors' Expectations, Beliefs and Perceptions