Out of Control: The (Over)Use of Controls in Accounting Research

Robert L. Whited*
North Carolina State University
rwhited@ncsu.edu

Quinn T. Swanquist
University of Alabama
qtswanquist@cba.ua.edu

Jonathan Shipman
University of Arkansas
jshipman@walton.uark.edu

James R. Moon, Jr.
Georgia Institute of Technology
robbie.moon@scheller.gatech.edu

April 2021

* Corresponding Author

This study has benefited from discussions with faculty and PhD students at the University of Tennessee, the University of Alabama, and the 2018 ATA Midyear Meeting. We thank Lisa Hinson, Yupeng Lin, Shuqing Luo, Chris McCoy, Linda Parsons, Michael Ricci, Daniel Street, Jennifer Tucker, Sally Widener, the Tax Reading Group at University of California–Irvine, and workshop participants at Clemson University, National University of Singapore, and the University of Florida for valuable input. We also thank Chelsea Anderson, Justin Blann, Nathan Groff, and Charley Irons for valuable research assistance. Robbie Moon gratefully acknowledges financial support from the Hubert L. Harris Early Career Professorship at the Georgia Institute of Technology, Jonathan Shipman gratefully acknowledges financial support from the Garrison/Wilson Endowed Chair in Accounting at the University of Arkansas, and Quinn Swanquist gratefully acknowledges financial support from the Ernst and Young Professorship at the University of Alabama.

Out of Control: The (Over)Use of Controls in Accounting Research

Abstract: In the absence of random treatment assignment, the selection of appropriate control variables is essential to designing well-specified empirical tests of causal effects. However, the importance of control variables seems underappreciated in accounting research relative to other methodological issues. Despite the frequent reliance on control variables, the accounting literature has limited guidance on how to select them. We evaluate the evolution in use of control variables in accounting research and discuss some of the issues that researchers should consider when choosing control variables. Using simulations, we illustrate that "more control" is not always better and that some control variables can introduce bias into an otherwise well-specified model. We also demonstrate other issues with control variables including the effects of measurement error, use of controls with interactions, and complications associated with fixed effects. Lastly, we provide practical suggestions for future accounting research.

KEYWORDS: accounting research methods, controls, measurement error, interactions, fixed effects

JEL Classifications: M40, M41, C18, C52

1. Introduction

A large body of empirical accounting research attempts to draw causal links between treatments (X) and outcomes (Y). Studies using non-experimental data face the difficult task of ruling out alternative (non-causal) explanations for observed relations between variables. In the absence of random (or as-if random) treatment assignment, researchers often use control variables (Z) to empirically adjust for factors that confound estimates of the causal relation between X and Y.[1] Failure to include a confounding control results in omitted variable bias (OVB), an issue widely recognized in accounting research (e.g., Bloomfield, Nelson, and Soltes (2016), Speklé and Widener (2018), and Ittner (2014)). To mitigate OVB, studies often include an extensive list of control variables; Bertomeu, Beyer, and Taylor (2016) refer to this as the "kitchen sink" approach to control selection. This approach, which is potentially an outcome of the publication process, generally presumes that more controls improve model specification. While proper controls help alleviate OVB, more controls do not necessarily translate to "better" models. In fact, some controls can lead to "included variable bias" by isolating or opening unwanted paths from X to Y (Ayres 2005). Moreover, even a "good" control must be accurately measured and reliably capture the intended construct to effectively address OVB.

In this study, we offer guidance on the use of control variables to isolate causal effects. We begin by reviewing the use of control variables in top accounting journals over time. We then discuss the conditions under which a control variable is and is not appropriate using causal diagrams to illustrate the expected relations among variables (Gow, Larcker, and Reiss 2016; Pearl 1995). Causal diagrams demonstrate how controls can distort causal interpretation and

[1] We use the terms "controls" and "control variables" to indicate attempts to disentangle an alternative, or confounding, explanation to make a causal inference. The extent to which control variables effectively control for alternative explanations serves, in part, as a motivation for this study.

force the researcher to consider the nature of the underlying data and the proposed theoretical relations. This exercise is helpful because statistical software cannot inform whether X causes Y, Y causes X, or whether Z variables confound these relations. Rather, proper model design requires theory to inform the underlying direction of causality and the interpretation of statistical estimates. Consistent with this, correlation between a potential control, Z, and X or Y does not necessarily justify inclusion of Z in a regression or improve the causal interpretation of the coefficient estimate on X. Instead, the researcher should control for factors (Z) relating to X and Y that are not caused by X or Y (i.e., outcomes of X or Y) when investigating whether variation in X *causes* variation in Y. Controlling for variables which are affected by X or Y leads to biased estimates of the causal effect of X on Y. To provide context for these issues, we use simulated data on the returns to a certified public accountant license (CPA) and archival data on auditor-client characteristics. In each setting, we discuss the process of identifying confounding controls and demonstrate empirical consequences of including/excluding these "good" controls. Likewise, we discuss the conditions under which a variable impairs causal interpretation of the resulting regression and demonstrate the bias introduced by including these "bad" controls.

We next cover several other issues related to improving control variable selection and measurement. First, we discuss why control variables need to accurately and precisely capture the intended construct to be effective. While a common refrain is that measurement error in a treatment variable "biases against" findings, we illustrate how measurement error in control variables dilutes control effectiveness. Second, explanatory variables that relate only to X or Y can be useful in some settings even though they are not true confounding controls. We discuss and illustrate the effects of including these types of variables. Third, fixed effects are simply "dummy" controls that adjust for the effects of categorical groups and isolate within-group variation in the data. Depending on the type of fixed effect and the nature of treatment, within-group variation may differ substantially from cross-sectional variation. We discuss and illustrate the consequences of fixed effects inclusion in a variety of conditions. Fourth, we discuss the importance of controls when investigating the interactive effect of X with a moderating variable (I). We demonstrate that, in such cases, the interactive effects of I with properly selected control variables (i.e., Z interacted with I) often correlate with the interaction of interest, resulting in OVB if excluded.

We conclude the study by offering suggestions for future research on best practices for the use of control variables. We hope that the discussion and suggestions in this study are useful to accounting researchers in all topical areas as they continue to seek convincing empirical support for causal inferences.

2. The Importance of Control Variables

2.1 Control Variable Guidance and Use in Accounting Research

While accounting researchers understand OVB and the importance of control variables, the accounting literature and most doctoral program curricula have limited specific guidance on how to select and specify controls. For example, econometrics texts aimed at graduate students (which often serve as the first introduction to econometric analysis for accounting researchers) cover these issues at a high level. Greene (2012, p. 51-52) discusses moving from a research question to an empirical specification, noting broadly that "the underlying theory will specify the dependent and independent variables in the model." Likewise, Wooldridge (2010, p. 3) notes that "deciding on the list of proper controls is not straightforward, and using different controls can lead to different conclusions about a causal relationship." While these texts provide useful discussions of endogenous variables and the implications of OVB, they do not focus on practical and detailed guidance on how to select control variables to strengthen causal inference.[2] Some recent studies focus on broad issues with causal inference in accounting reseach and in doing so touch on some control-related issues (e.g., Gow et al. 2016; Bertomeu et al. 2016).[3] However, similar to the econometric texts, the accounting literature provides limited guidance for improving control variable selection. Our objective is to provide a more comprehensive discussion of what constitutes proper control and practical guidance for accounting researchers using controls to isolate causal effects.

To provide a longitudinal perspective on the use and importance of controls in accounting research, we reviewed studies published in *The Accounting Review, Journal of Accounting and Economics,* and *Journal of Accounting Research* from 1980 until 2020. We find that regression-based research designs now dominate the literature as the percentage of studies using regression analysis increased from 25 percent in the period between 1980 and 2000 to 68 percent between 2005 and 2020.[4] While accounting literature has experienced a trend towards identifying "natural experiments" or "exogenous shocks," researchers still rely heavily on controls for causal inference. For studies using regression analysis, our review suggests that the number of controls has increased significantly. From 1980 until 2000, the number of regressors averaged six. This number has since steadily increased to a high of 16 in 2020.[5] Advances in data analysis capabilities coupled with continuously expanding data on financial statements, management disclosures, executive compensation, audits, stock returns, and more have likely facilitated this

[2] This is not a criticism as discipline-specific model design and control variable selection are not the purpose of these resources. These texts instead aim to provide graduate students with an understanding of econometric techniques and the assumptions underlying their use.
[3] Other disciplines also provide discipline-specific advice on control variable selection (e.g., Becker 2005; Atinc, Simmering, and Kroll 2012).
[4] We examine articles appearing in the first issue of each journal in five-year increments from 1980-2020, resulting in 9 issues from each journal. In recent years, the majority of studies that did not use regressions were experiments.
[5] We identify the number of regressors based on the regression with the maximum number of regressors in a study.

growth.

Though recent studies use more controls, their prevalence in tabulated results has declined. Studies frequently suppress control variable coefficients, particularly in recent years. From 1980-2000, 13 percent of tabulated regressions suppress control variable coefficients; from 2005-2020, 22 percent suppress control coefficients (26 percent in 2020). Relatedly, econometrics texts often advocate "model building" (Greene 2012) or the delineation of "base" and "alternative" specifications (Stock and Watson 2011) to help evaluate the sensitivity of results to control inclusion. One may expect these approaches to be more common in recent literature given the aforementioned trends, but, on the contrary, we note a slight decline in specification building from 26 percent from 1980-2000 to 23 percent from 2005-2020.[6] Collectively, we observe that researchers increasingly rely on regression-based analysis and use more control variables, but we do not observe similar increases in attention to control sensitivity.

2.2 Omitted Variable Bias and Causal Diagrams

In observational studies, a univariate analysis is unlikely to yield an unbiased estimate of the *causal* effect of a treatment *X* on an outcome *Y*. Consider the following model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

The regression above yields biased estimates of the causal effect of *X* on *Y* ($\hat{\beta}_1$) if factors that determine *Y* also determine *X* (i.e., non-zero correlation between *X* and ε), an endogeneity problem commonly referred to as OVB (Stock and Watson 2011).[7] A researcher can alleviate OVB by specifying a multiple regression model that includes *Z* for common determinants of *X* and *Y*. By "extracting" the effects of *Z* from the error term (ε), the researcher can obtain unbiased

[6] We identify this type of analysis when a study presents the same analysis with different sets of controls.
[7] In fact, most problems with endogeneity, or alternative explanations, boil down to an OVB problem (see Chapter 4 of Wooldridge (2010) and Chapter 9 of Stock and Watson (2011) for discussion). In this sense, researchers must address OVB to draw reliable causal inferences from observational data.

estimates of the causal relation between X and Y. Given most treatments in accounting research are self-selected or otherwise non-randomly assigned (e.g., executive compensation, governance mechanisms, disclosure choice), these issues are ubiquitous. Absent as-if random treatment, researchers must identify and accurately specify the appropriate Z to isolate the causal effect of X on Y.[8]

To aid in the construction of well-specified causal models, researchers should consider their research question in the context of causal diagrams (Pearl 1995). These diagrams help the researcher identify sources of OVB and properly specify a model; they also help the reader understand the causal relations the researcher had in mind when designing the study. We display an example causal diagram adapted from Bertomeu et al. (2016) in Figure 1. Here, *A* causes *B*, *B* causes *C*; *C* causes *D* and *E*; *D* also causes *E*. Such models vary in terms of sophistication and complexity, but they allow a researcher to lay out the underlying theory for the relations between variables.[9] For example, if the reader were interested in the effects of *D* on *E*, this model suggests *C* represents a confounder, and therefore a necessary control. Conversely, a researcher interested in the effects of *C* on *E* would not want to control for *D*, as *D* mediates the relation between *C* and *E*. The specific causal link in question should drive model design. Importantly, a regression cannot tell the researcher whether *C* causes *E*, *E* causes *C*, or whether *D, A*, or *B* represent necessary controls. Regressions simply estimate conditional correlations. Only theory can inform the causal relations which dictate the structure of the regression and the interpretation of results. We expand on these issues in the following sections.

[8] Prior accounting research documents complications associated with estimating causal effects in accounting settings when data on a confounding variable is unavailable ("unobservable") (Lennox, Francis and Wang 2012; Larcker and Rusticus 2010; Tucker 2010), or when the functional form is misspecified (Lawrence, Minutti-Meza, and Zhang 2011; Shipman, Swanquist, and Whited 2017).

[9] To conceptualize this diagram, suppose *A* reflects workforce quality, and *B* reflects the product quality. *B* determines company performance (*C*), which determines both share price (*D*) and executive compensation (*E*). Executive compensation includes equity incentives, which generates a link between *D* and *E*.

2.3 Specifying Models: Identifying "Good" and "Bad" Controls

In the context of causal diagrams, "good" Z represent possible "alternative explanations" for the relation between X and Y (i.e., confounders). Confounders capture constructs that cause variation in both X and Y. Failure to include confounders in Z leads to biased estimates of the causal effect of X on Y (i.e., OVB). However, including control variables in Z that 1) are outcomes of X or Y, 2) capture the same construct as X or Y, or 3) are otherwise mechanically related to X or Y, can impair a model's causal interpretation. A good rule of thumb advocated by Angrist and Pischke (2009, p. 64) is that "good controls are variables that we can think of as having been fixed at the time the regressor of interest [X] was determined," and "bad controls are variables that are themselves outcome variables in the notional experiment at hand." In general, if a potential control is determined *after* the treatment variable, the researcher should seriously consider its appropriateness in the model because variables that are on the causal path between treatment and outcome cannot conceptually be "held constant" as treatment varies.[10]

Extant accounting research rarely motivates control variables using causal diagrams, instead relying on (1) determinants models from prior literature or (2) "kitchen sink" approaches with an extensive set of control variables. Approach (1) often involves selecting a model from prior literature used to test the relation between a different independent variable (X) and the same (or similar) dependent variable (Y). Assuming the prior study properly specified its model, the validity of this approach requires that the common determinants of X and Y be the same in both studies. The "kitchen sink" approach involves gathering controls from multiple studies and generally yields a lengthy model with an extensive set of control variables. Without careful

[10] A control often proxies for an unobservable construct. For instance, a test score may proxy for intelligence. In this case, the underlying construct "causes" or "determines" X and Y, but the observed Z may be measured after Y. This does not cause an issue as long as X and/or Y do not, themselves, cause variation in Z.

theoretical consideration, though, both approaches risk including control variables that impair the causal interpretation of the model and/or masking the absence of important variables with unnecessary controls.

Relatedly, accounting studies generally discuss controls in the context of predicting *Y*, a useful consideration, but often fail to consider the control's relation with *X*, at least explicitly. Regressions take the form "*Y* = ..." potentially reinforcing the focus on predicting *Y*. Consistent with this perspective, measures of predictive power such as R^2 or the area under the ROC may be used as evidence of model quality, but high predictive ability does not necessarily indicate model quality or control sufficiency for several reasons. First, the R^2 is largely a product of the overall predictability of *Y* rather than the causal relation of interest. For example, a regression of audit fees (*Y*) on an auditor trait (*X*) will have a high R^2 as long as *Z* includes a control for client size. However, despite a high R^2, OVB concerns abound due to the complex relationship between audit fees and traits of the client and auditor. Conversely, a regression of market returns (*Y*) on an event (*X*) and even the most thorough set of potential market predictors will yield a low R^2. The low R^2 occurs due to the idiosyncratic nature of returns and is neither an indication that the test likely suffers from OVB nor that the researcher should try to identify control variables that improve the R^2. In each of these settings, researchers should consider whether there are likely variables that may represent a common cause of *X* and *Y*, and therefore, present OVB concerns. Another important consideration with respect to R^2 is that bad controls may significantly increase the predictive ability of the model, while at the same time impairing causal inference. For example, control variables that are outcomes of *X* or *Y* or reflect the same construct as *X* or *Y* will increase the predictive power of the model, sometimes significantly, but will also impair causal inference.

3. Illustrations of "Good" and "Bad" Controls

To illustrate the concepts of "good" and "bad" control, we present two parallel examples of the effect of controls on estimated causal effects. The first relies on a simulated dataset of accountants' CPA certification status, skill, and income. The second uses archival data on auditor type, client size, and audit fees.[11]

3.1 Description of Simulation and Data

For our first setting, we simulate a dataset of accountants including three variables: innate accounting skill (*Skill*), CPA status (*CPA*), and earnings (*Earnings*) using the following parameters (the related descriptive statistics are presented in Table 1, Panel A):

A1) Create a dataset of 5,000 accountants[12]
A2) Half of the accountants in the sample have high skills: $P(Skill = 1) = 0.50$
A3) Approximately three in ten low-skill accountants obtain a CPA: $P(CPA = 1 \mid Skill = 0) = 0.30$
A4) Approximately seven in ten high-skill accountants obtain a CPA: $P(CPA = 1 \mid Skill = 1) = 0.70$
A5) The average earnings for low-skill accountants without a CPA is \$50,000: $E[Earnings \mid CPA = 0, Skill = 0] = \$50{,}000$
A6) The average incremental earnings for high-skill accountants relative to low-skill accountants is \$15,000, holding CPA status constant: $E[Earnings \mid Skill = 1, CPA] - E[Earnings \mid Skill = 0, CPA] = \$15{,}000$
A7) The average incremental earnings for CPAs relative to non-CPAs is \$30,000, holding skill constant: $E[Earnings \mid CPA = 1, Skill] - E[Earnings \mid CPA = 0, Skill] = \$30{,}000$
A8) Earnings contains a random 'noise' component drawn from a normal distribution with a mean of \$0 and a standard deviation of \$10,000

For our second setting, we extract data from the Audit Analytics Audit Fees dataset (AA) for fiscal years between 2003 and 2015 for non-financial/utility industries (we present

[11] These examples are meant to be simple, adaptable, and relatable for an accounting audience. We emphasize that both settings are over-simplified and in no way intended to inform the underlying research questions. We also note that our CPA simulation is adapted from the classic "returns to schooling" setting commonly used in econometrics texts (e.g., Angrist and Pischke (2015, p. 214)).
[12] Sample size in these simulations does not affect the issues related to the existence of bias that we discuss. Increasing the sample size simply increases precision but does not relate to the absence or presence of bias.

descriptive statistics in Table 1, Panel B). *Big4* represents company years with a Big 4 auditor, *ln(Fees)* represents the natural log of total fees (total_fees), and *ln(Assets)* represents the natural log of assets (matchfy_balsh_assets).

3.2 Confounders

The term confounder refers to a variable that "confounds," or provides an alternative explanation for, a causal relation between *X* and *Y*. Including confounders in *Z* helps alleviate OVB.[13] To illustrate the importance of these controls, consider advising a student who plans to enter the accounting profession on whether to pursue a CPA certification. To assist the student, you seek to answer the following research question:

RQ 1a: What is the effect of the CPA certification on earnings?

We can easily estimate the average difference in earnings between CPAs and non-CPAs without holding "all else equal" by comparing average earnings between CPAs and non-CPAs. However, this difference is unlikely to inform your student's decision since there are common determinants of both CPA licensure and earnings. What we really want to know is: how much more, on average, can a student expect to earn if they obtain a CPA certification? To estimate this effect, we must address the non-random nature of certification status. For instance, suppose the innate skill of CPAs exceeds that of non-CPAs on average, as we specify in parameters A3 and A4. Because skill also positively affects earnings, a naïve comparison of CPAs versus non-CPAs' earnings suffers from OVB and yields an inflated estimate of the effect of certification on earnings (i.e., $E[Earnings_{0i} \mid CPA_i = 1] > E[Earnings_{0i} \mid CPA_i = 0]$).

In Figure 2a, we present a causal diagram of the relations between *CPA*, *Earnings*, and

[13] Controls that do not meet the criteria of confounders are not necessarily "bad controls." For example, a predictor variable that determines *Y* but not *X* will improve estimate precision and will not impair causal inference (see Section 4.2.1).

Skill. We include marginal effects in line with the parameters, which dictate that *CPA* has a "causal" benefit of $30,000 on *Earnings* (parameter A7). If your student estimates the incremental cost of a CPA certification at, for example, $32,000, the earnings benefits of pursuing a CPA do not justify the costs. To uncover this effect, though, we need to disentangle the effect of skill on earnings since skill increases the likelihood of certification (+0.4) *and* has a direct effect on earnings (+$15,000). To illustrate, we compare a naïve model in [1a] to an expanded model that includes a control for *Skill* in [1b].

$$Earnings_i = \beta_0 + \beta_1 CPA_i + \varepsilon_i \quad [1a]$$

$$Earnings_i = \beta_0 + \beta_1 CPA_i + \beta_2 Skill_i + \varepsilon_i \quad [1b]$$

In Table 2, Panel A, we present estimations of [1a] and [1b] in columns 1 and 2 respectively. The estimate of β_1 in column 1 ($36,105) exceeds the true causal effect of $30,000 due to OVB. Thus, a naïve estimation would lead you to incorrectly advise your student that pursuing a CPA certification carried a net expected value of approximately $4,000. However, column 2 provides an unbiased estimate of the causal effect of CPA certification, $30,348, which would lead you to advise your student not to pursue a CPA.

Next, we consider confounder controls using archival accounting data:

RQ 1b: Do Big 4 auditors charge higher audit fees?

It is well established (and perhaps obvious) that: (1) Big 4 auditors charge higher fees, (2) larger clients tend to choose Big 4 auditors, and (3) larger clients cost more to audit. We present these relations in the form of a causal diagram in Figure 2b. In this case, client size is a common cause of both audit fees and auditor selection.[14] Therefore, client size represents a confounder.

[14] Causal diagrams can only be determined theoretically. However, client size predates auditor selection and auditor selection precedes audit fee determination. As such, we suggest that theory supports the causal diagram in Figure 2b. We intentionally oversimplify the model by assuming client size is the only confounding factor.

We demonstrate this by estimating the following models in Table 2, Panel B:

$$ln(Fees)_i = \beta_0 + \beta_1 Big4_i + \varepsilon_i \qquad [1c]$$

$$ln(Fees)_i = \beta_0 + \beta_1 Big4_i + \beta_2 ln(Assets)_i + \varepsilon_i \qquad [1d]$$

In Table 2, Panel B, we present estimations of [1c] and [1d] in columns 1 and 2 respectively. Column 1 reflects the difference in average fees for Big 4 and non-Big 4 clients. The coefficient estimate, 2.33, is a replication of the mean difference in Table 1 and suggests that Big 4 clients pay significantly higher audit fees (> 900 percent).[15] A glaring omission in this model is that Big 4 auditors have substantially larger clients than non-Big 4 auditors. When controlling for *ln(Assets)*, the estimated Big 4 auditor premium declines dramatically to a more realistic effect of 0.55 (73 percent). These settings illustrate the importance of including controls for confounding constructs, particularly when the confounder strongly predicts both *X* and *Y*.

3.3 Mediators

While confounder controls improve causal estimates, "mediator" controls alter the interpretation of the relation between *X* and *Y,* and, depending on the research question, can bias estimates of causal effects. This occurs because mediators "block" (or mediate) a path by which *X* affects *Y*. To illustrate, we use the CPA setting and consider the following research question:

RQ 2a: What is the total effect of accounting skill on earnings?

The causal diagram for RQ 2a, presented in Figure 3a, differs fundamentally from the diagram for RQ 1a. While *CPA* only affects *Earnings* via one direct path, *Skill* affects *Earnings* via two paths. First, *Skill* increases *Earnings* by $15,000 via a direct path, as prescribed by parameter A6. In other words, we expect a highly skilled accountant to make $15,000 more than a non-highly skilled accountant with the same certification status. Second, accounting skill

[15] The economic significance is calculated as $e^{2.33}-1 = 9.27$ or 927 percent.

improves the likelihood of certification which, in turn, increases earnings (parameters A3, A4, and A7). We refer to this as the mediated (or indirect) effect. Thus, skill increases earnings through two causal "paths" and the total effect of *Skill* on *Earnings* is the combination of these two effects.

To estimate the total effect of skill on earnings, we should not control for the effect of CPA status on earnings, as skill increases the likelihood of a CPA. Thus, *CPA* represents a mechanism through which *Skill* increases earnings. Controlling for a mediator such as *Skill* will "throw the baby out with the bathwater" by isolating *only* the direct effect of skill on earnings, a relation that does not address *RQ 2a*. Another approach is to consider the total effect of a treatment as the expected difference in *Earnings* between *Skill* = 1 and *Skill* = 0 in an experiment that randomly assigns skill to individuals prior to the decision to obtain a CPA. Conceptually, we cannot "hold constant" certification status while varying *Skill*, since certification status itself is an *outcome* of skill. To illustrate, we compare a model without *CPA* in [2a] to the expanded model that includes *CPA* in [2b].

$$Earnings_i = \beta_0 + \beta_1 Skill_i + \varepsilon_i \tag{2a}$$

$$Earnings_i = \beta_0 + \beta_1 Skill_i + \beta_2 CPA_i + \varepsilon_i \tag{2b}$$

[2a] should yield estimates for β_1 of approximately $27,000, capturing the total effect of *Skill* on earnings. This effect equals the direct effect of *Skill* on earnings ($15,000) plus the increased probability of CPA certification (0.40) multiplied by the effect of certification on earnings ($30,000). In contrast, [2b] should yield estimates for β_1 of approximately $15,000 because holding *CPA* constant blocks the path from *Skill* to earnings through *CPA*, thereby isolating the direct effect. Table 3 Panel A presents estimations of [2a] and [2b] that conform with expectations. As expected, column 1 provides an estimate of the full causal effect.

However, by controlling for a mechanism through which skill increases earnings in column 2, we only capture the direct effect of skill on earnings. This demonstrates that inappropriate controls, in this case *CPA*, can obscure causal effects even in the presence of as-if random treatment. This scenario also highlights the importance of considering coefficient estimates with respect to the research question. That is, researchers may be interested in how skill affects earnings. For example, an accounting firm, state certification board, or labor economist may wish to know whether skill influences earnings independent of certification (i.e., would two accountants of different skill make the same amount conditional on certification status?). In this case, [2b] would be the appropriate choice.

Next, we consider the case of a mediator in the audit fees setting and ask:

RQ 2b: Do larger companies pay higher audit fees?

We present a causal diagram for RQ 2b in Figure 3b. Client size affects fees through two "paths." First, client size increases audit work, thereby increasing audit fees (i.e., direct effect). Second, larger clients are more likely to select more expensive Big 4 auditors (i.e., indirect path). Similar to the simulation, we estimate the following two models in Table 3:

$$ln(Fees)_i = \beta_0 + \beta_1 ln(Assets)_i + \varepsilon_i \quad [2c]$$

$$ln(Fees)_i = \beta_0 + \beta_1 ln(Assets)_i + \beta_2 Big4_i + \varepsilon_i \quad [2d]$$

Estimations of [2c] and [2d] document a positive and significant relation between company size and audit fees in Table 3, Panel B. This is unsurprising given the effect of *ln(Assets)* on *ln(Fees)* is perhaps one of the strongest relations documented in the accounting literature. Similar to the simulated setting above, though not as dramatic, we observe a substantially smaller coefficient on *ln(Assets)* in model [2d] which includes the mediator, *Big4*, as the inclusion of this control extracts a path through which company size influences audit fees.[16] An important observation from these analyses is that [1b/d] and [2b/d] are identical specifications, but the appropriateness of each specification depends on the research question. In fact, [1b(d)] yields biased estimates of the full effect of *Skill* (*ln(Assets)*) but unbiased estimates of the effect of *CPA* (*Big4*). For this reason, researchers should use caution when borrowing models from prior literature or judging the appropriateness of a model based on the significance of control variable coefficient estimates. If researchers desire to estimate the mediated effect (i.e., direct path), then they should motivate and interpret the model accordingly.

To illustrate this issue in a plausible research setting, we consider Francis, Nanda, and Olsson (2008), who argue that earnings quality, not voluntary disclosure, has a first order effect on the cost of equity. They correctly include proxies for earnings quality when exploring the association between voluntary disclosure and the cost of equity. However, if one wished to analyze the full causal effect of earnings quality on cost of capital, voluntary disclosure likely qualifies as a mediator.

3.4 Colliders

"Colliders" are outcomes of *Y*, and generally impair causal inference. In the CPA setting, consider an accounting firm contemplating whether the CPA designation reflects prospective hires' accounting skill and asking the following research question:

RQ 3a: Are higher skilled accountants more likely to be CPAs? (i.e., is the CPA certification process selective?)

In Figure 4a, we present the causal diagram for RQ 3a. One could make a specious argument for the inclusion of an individual's earnings to answer RQ 3a since earnings correlates with *CPA* and *Skill* (and improves the R^2). In this setting, however, earnings (*Z*) is an outcome of

[16] This mediator diminishes the effect size because the relation between *X* and *Z* is the same as the relation between *Z* and *Y*. However, mediators can also magnify the effect if the variable exhibits opposite correlations with *X* and *Y*.

both the treatment (*Skill*) and outcome (*CPA*) variables. In this sense, earnings should not (and cannot) be held constant because it is counterintuitive to change skill and certification status while holding the outcome of those variables constant. However, statistical estimation tools lack this intuition and will render a coefficient estimate whether it has practical meaning. To illustrate the effects of including a "collider" control, we estimate the following regressions:

$$CPA_i = \beta_0 + \beta_1 Skill_i + \varepsilon_i \qquad [3a]$$

$$CPA_i = \beta_0 + \beta_1 Skill_i + \beta_2 Earnings_i + \varepsilon_i \qquad [3b]$$

As the empirical results in Table 4 Panel A demonstrate, including *Earnings* as a control generates seriously misleading results. [3a] produces an unbiased estimate of the effect of skill on certification status (based on parameters A3 and A4 above). Higher skill corresponds to an approximately 40 percent increase in the probability of obtaining a CPA, suggesting that the CPA designation is selective and reflects accounting skill. However, estimates from [3b] suggest a significant *negative* relation between *Skill* and *CPA*, which would incorrectly suggest the *less* skilled accountants obtain certification. This occurs because a high skill individual with the same earnings as a low skill individual is less likely to have CPA certification.

We also illustrate this issue in the Big 4 setting with the research question:

RQ 3b: Do larger clients tend to select Big 4 auditors?

Here we predict that larger clients are more likely to select a large auditor. We present a causal diagram of this relation in Figure 4b. In this setting, a researcher might include an audit fee control to proxy for client reporting complexity since, similar to including earnings in [3b] above, it correlates with both the *Y* and *X* variables and improves R^2. However, audit fees are an outcome of client size and auditor type, making fees a collider in this setting. To demonstrate, we present estimations of the following models in Table 4, Panel B:

$$Big4_i = \beta_0 + \beta_1 ln(Assets)_i + \varepsilon_i \quad [3c]$$

$$Big4_i = \beta_0 + \beta_1 ln(Assets)_i + \beta_2 ln(Fees)_i + \varepsilon_i \quad [3d]$$

While the sign does not flip, as in the CPA setting, the inclusion of *ln(Fees)* substantially biases estimates of β_1, cutting the magnitude of the estimate in half (0.10 to 0.05). Moreover, the coefficient estimate on company size has little meaning. Conceptually, it makes little sense to investigate the effect of company size on auditor selection holding fees constant as fees necessarily increase when a client hires a more expensive auditor.

Unlike mediators, colliders unequivocally impair causal inference provided they exhibit at least a modest association with *X* and *Y*. While the audit fee example may seem impractical, colliders are often more subtle. For instance, a researcher may wish to investigate the association between some executive background trait and company performance (e.g., return on assets). Since controls are often motivated based on their association with *Y*, a researcher may include an executive compensation control to adjust for "executive incentives." However, company performance (*Y*) at least partially determines executive compensation (*Z*) due to performance and stock-based incentives. Likewise, CEO background traits (*X*) may affect compensation (*Z*) via a variety of paths. Thus, executive compensation likely qualifies as a collider.

3.5 Same Construct Controls

"Same construct" controls refer to variables that are inseparable from either *X* or *Y* since they largely reflect the same underlying construct. While these controls are similar to mediators and colliders, they cannot be cleanly placed in a causal diagram since they are, by definition, determined contemporaneously (i.e., they belong in the same box as *X* or *Y*) and can significantly distort causal estimates. If *Z* reflects the same construct as *Y*, it is an outcome of *X*, but controlling for it produces the counterintuitive estimate of "the relation between *X* and *Y* holding the same construct as *Y* (i.e., *Z*) constant." In other words, the variable captures an alternative dependent variable rather than a confounding factor. A related issue occurs if *Z* reflects the same construct as *X.* Theoretically, *X* cannot vary while holding constant another measure of the same underlying construct. When this occurs, the partial derivative of *Y* on *X* does not capture the causal effect of *X* on *Y*.

In reality, variables frequently reflect a variety of constructs making these same construct issues less obvious than mediators or colliders. To avoid such controls, we suggest considering the construct underlying the observed measures of *X* and *Y*, and whether *Z* ostensibly overlaps with these constructs. Grouping *Z* variables by construct (e.g., company size, profitability, governance) can aid in the assessment of controls at the construct level since it may elucidate when *X* or *Y* falls into one of these groups. In general, researchers should not include *Z* that reflect *X* or *Y* unless intentional (e.g., horseracing the predictive ability of a new IV (*X*) against existing measures (*Z*)).[17]

To illustrate the same construct concepts, suppose a researcher wants to investigate how hiring a Big 4 auditor (*Big4*) impacts audit fees (similar to RQ 1b), holding constant the number of the audit firm's clients (*ln(Auditor Clients)*).[18] Because *Big4* is definitionally an auditor size dummy variable, the construct of auditor size underlies both *Big4* and *ln(Auditor Clients)*. As a result, it is difficult to conceptualize varying *Big4* without also varying *ln(Auditor Clients)*. We empirically demonstrate these issues using models [4a] (which is the same as [1b]) and [4b] (which includes *ln(Auditor Clients)* as an additional control):

[17] Relatedly, researchers may have multiple empirical measures of the same construct underlying *X* (e.g., multiple measures of accounting quality). In this case, variables can be included in separate specifications, included jointly but not evaluated independently (i.e., conduct an F-test to jointly test the significance), or perform a principal component analysis to create an aggregate measure of the underlying construct (see Guay, Samuels, and Taylor (2016) for an example of this).
[18] We do not simulate the same construct issue in the CPA scenario, but including a control for individual IQ when *Skill* is *X* or a control for years of education when CPA is the *Y* would lead to same construct issues.

$$ln(Fees)_i = \beta_0 + \beta_1 Big4_i + \beta_2 ln(Assets)_i + \varepsilon_i \quad [4a]$$

$$ln(Fees)_i = \beta_0 + \beta_1 Big4_i + \beta_2 ln(Assets)_i + \beta_3 ln(Auditor\ Clients)_i + \varepsilon_i \quad [4b]$$

In Table 5, we estimate a coefficient of 0.11 on *ln(Auditor Clients)* in column 2 and its inclusion dramatically reduces the coefficient on *Big4* from 0.55 (column 1) to 0.17 (column 2). However, this regression does not inform the effects of a Big 4 auditor, as [4b] splits the relevant effect of auditor size between the coefficient on *Big4* and *ln(Auditor Clients)*. Variance inflation factors (VIFs) do not necessarily diagnose the same construct issue as all VIFs, which are commonly used for assessing multicollinearity, are less than 5.0. This highlights the importance of relying on theory to identify same construct issues rather than relying on VIFs.

Building models using the "kitchen sink" approach can often lead to same construct controls since variables of interest in one study may serve as controls in others. For example, studies may proxy for the "information content" of an event with absolute price response or trading volume response. Therefore, a researcher testing the information content, measured with abnormal volume response (*Y*) of some event (*X*), should not control for absolute abnormal returns (*Z*) since it is another measure of information content. However, absolute returns frequently appear as a regressor in research when the desired construct is disagreement (e.g., Garfinkel 2009).[19] So, if a researcher wishes to investigate whether an information event produces disagreement, then absolute returns would be an appropriate control. This setting again highlights the importance of clearly identifying causal mechanisms and the constructs underlying variables, as well as interpreting coefficient estimates in light of the variables included in a model.

[19] As an aside, researchers commonly operationalize constructs using the absolute value of a measure (e.g., absolute abnormal accruals). In such cases, researchers should consider whether control variables should be specified accordingly (e.g., absolute value of operating cash flows rather than signed operating cash flows).

4. Other Considerations for Good Control

4.1 Measurement Error

Proper "control" for confounding factors relies on the ability to observe and measure those factors with precision and accuracy. However, measurement error masks the relation between the variable and the intended construct, limiting the effectiveness of control variables measured with error. As a result, measurement error in a given variable not only biases that specific variable's coefficient estimate towards zero, an issue referred to as "attenuation bias," but also generally leads to biased coefficient estimates on other variables in the model (Wooldridge 2013; Maxwell and Delaney 1990). Unlike other settings where statistical power can diminish concerns with noise (e.g., noisy Y), large sample sizes do not alleviate this issue (Westfall and Yarkoni 2016). Measurement error in X (or Y) is often considered a secondary concern since it usually biases against a statistically significant effect in X (as long as the noise is random).[20] However, measurement error in Z can bias in favor of a statistically significant effect in X as measurement error in confounding Z variables can increase OVB.[21]

In practice, measurement error arises from multiple sources. The first source relates to errors in the data. This can arise when a company misstates an account in their financial statements or the data aggregator (e.g., Compustat) inputs the wrong value. While some account balances are easily verifiable (e.g., cash balance) and are therefore unlikely to contain significant measurement error, other accounts are significantly more difficult to measure (e.g., Level III fair value assets) and thus are more likely to deviate from the 'true value.' A second source of

[20] Measurement error in Y can also introduce bias in certain instances. See deHaan, Lawrence, and Litjens (2021) for a discussion and analysis of one such setting.
[21] Contemporaneous work by Jennings, Kim, Lee, and Taylor (2021) explores this issue as well. We focus on the effect of "noise" in control variables, or measurement error uncorrelated with other factors in the model. They take a broader view and also consider a specific empirical setting where measurement error is observable.

measurement error occurs when the empirical proxy does not accurately capture the underlying theoretical construct. For example, researchers commonly proxy for the construct of company size with the natural log of assets. However, there are more aspects of company size than just accounting assets (e.g., number of transactions, number of employees, differences between asset book value and fair value, etc.). This type of measurement error presents significant obstacles to accounting research as researchers frequently use rough quantitative measures to proxy for complex/nuanced constructs such as financial distress (e.g., Altman Z score), corporate governance strength (e.g., G-Index), financial reporting quality (e.g., abnormal accruals), or fraud risk (e.g., F-score).[22]

To illustrate the effect of control variable measurement error, we use the *CPA* and *Big4* settings from above. We assume that *Skill* and *ln(Assets)* capture the respective underlying constructs without error and that the estimates from Table 2 capture the "true" causal effects of *CPA* (Panel A) and *Big4* (Panel B) on *Earnings* and *ln(Fees),* respectively. In each setting, we estimate the regression 1,000 times while progressively adding noise to the control variables (*Skill* and *ln(Assets)*).[23] We capture and plot the coefficient estimates from each regression in Figure 5. With no noise, the estimates replicate the "true" effect from Table 2 column 2. As noise increases, however, the effect of *Skill* (*ln(Assets)*) attenuates to zero. More concerning, the estimated *CPA* (*Big4*) effect becomes overstated as the *Skill* (*ln(Assets)*) effect attenuates. That is, as the coefficient estimate on *Skill* (*ln(Assets)*) converges to 0, the estimate on *CPA* (*Big4*)

[22] In many cases, outliers and non-linearities can have effects similar to measurement error. That is, misspecification of a variable reduces its effectiveness as a control. These concepts, while related, are outside the scope of our discussion. See Leone, Minutti-Meza, and Wasley (2019) for discussion on outliers and Shipman et al. (2017) for a discussion of non-linearities.

[23] Recall that *Skill* is a random binary variable. To add noise, we progressively change the likelihood that our measured skill variable represents actual skill versus a random binary value. Over each iteration from 1-1,000 the noisy skill variable goes from 100% actual skill to 100% noise. For the Big 4 setting, we multiply *ln(Assets)* for each observation by a random number drawn from a normal distribution having a mean of 1 and a standard deviation of s. In the 1st regression s = 0.001 (very little noise), the 2nd regression s = 0.002, … , in the 1,000th regression s = 1.

converges to the uncontrolled effect in column 1 of Table 2. As noise in the control increases, it effectively becomes a random variable which is uncorrelated with *X* and *Y*, thus reintroducing OVB and biasing estimates on the coefficient for *X*.[24] Thus, the extent to which *Z* effectively addresses OVB is largely determined by how accurately and precisely *Z* captures the underlying construct. For highly correlated confounding constructs, researchers may consider using multiple operationalized measures (e.g., control for company size with assets, revenue, and equity simultaneously).

4.2 Variables that directly relate to only Y or X

4.2.1 Variables that relate only to Y (but not to X)

A variable may determine *Y*, but have no effect on *X*. This is common in randomized experiments (or natural experiments) where nothing, except chance, determines *X*. Even in non-experimental settings, some determinants of *Y* may not relate to *X*. While unbiased estimation does not require these "artifact variables" (Carlson and Wu 2012), including them as regressors improves estimate precision by reducing the unexplained variation in *Y*. We extend the CPA simulation to illustrate this concept. Unbiased estimation of the effects of *Skill* on *Earnings* requires no controls because we randomly assigned *Skill*. Suppose, though, that we have another variable, geographic cost of living (*Cost of Living*), that determines *Earnings* but does not relate to randomly assigned *Skill*.[25] We then add *Cost of Living* as a control to [2a] and present the results in Table 6, Panel A. Including *Cost of Living* reduces standard errors, increasing the test

---

[24] We demonstrate this issue in a regression framework, but the same issue applies to matching techniques. That is, the more noise in a matching variable, the lower the quality of the matches.
[25] *Cost of Living* is comprised of two equal components, (1) the 'noise' from parameter A8 and (2) a random value drawn from a normal distribution with a mean of 0 and a standard deviation of $10,000. This makes *Cost of Living* related to *Earnings*, but unrelated to any of the regressors.

statistic.[26]

4.2.2 Variables that relate to X (but not to Y)

A variable might instead determine *X* but have no direct relation with *Y* (conditional upon *X*). The effect of including such a variable as a control depends on the nature of the relation. In one case, the variable relates to *X* and relates to *Y* through *X.* This situation reflects the primary condition for an instrumental variable.[27] While potentially a good instrument for *X*, such a variable does not serve as a useful as a control. To illustrate, consider the random assignment of a CPA test prep course to CPA test takers. Consider further, that the CPA prep course does not affect the accountant's earnings potential in any way *except* by increasing the likelihood of obtaining a CPA. To simulate this, we make two modifications to the parameters A1-A8 above. First, we replace *Skill* with *CPA Prep* as the sole determinant of *CPA* (A2-A4). Second, we remove the direct relation between *CPA Prep* and *Earnings* (A6 – formerly *Skill* and *Earnings*) such that only *CPA* determines *Earnings.* In Table 6, Panel B, we regress *Earnings* on *CPA* in column 1 and then add *CPA Prep* as a control in column 2. Because no variables other than *CPA* determine *Earnings*, both columns produce unbiased estimates. While *CPA Prep* does not materially affect the estimate on *CPA* in column 2*,* it decreases the precision of the estimate on *CPA*. This occurs because *CPA Prep* strongly predicts *CPA* but does not increase the overall R^2 of the model relative to including *CPA* alone (no incremental explanatory power), which increases standard errors.

A second scenario occurs when *Z* relates to *X*, but not to *Y* either directly or through *X.* In

[26] In practice, it is rare to find completely orthogonal variables, so inclusion of predictors of *Y* nearly always affect the estimated coefficient on *X.* Consistent with Carlson and Wu (2012), we encourage researchers to consider the need for these *Z* variables and to present regressions both with and without *Z* to demonstrate that changes are largely driven by precision.
[27] See Lennox et al. (2012), Larcker and Rusticus (2010), and Tucker (2010) for discussion of these methods.

this case, Z predicts a component of X that is unrelated to Y. The component unrelated to Y is similar to measurement error in the sense that it is a measured part of X that does not exhibit the predicted relation with Y. Controlling for Z that relates to measurement error in X will partial out the measurement error in X, reducing the attenuation bias on X and yielding a more accurate estimate of the effect of X on Y (see Spector and Brannick (2011, p. 290-293) and Carlson and Wu (2012, p. 417-418)). To illustrate, we adjust the simulation on skill and earnings above. We replace *Skill* with a continuous measure (*Continuous Skill*) drawn from a normal distribution having a mean of 100 and a standard deviation of 10 (similar to intelligence quotient). *Earnings* has a fixed component of $25,000, increases by $500 for each unit of *Continuous Skill*, and has a random component drawn from a normal distribution with a mean of 0 and a standard deviation of $10,000. However, we do not measure *Continuous Skill* directly; rather, we measure it via a test *(Test Score)*. Similar to the previous example, half of the subjects have taken a practice test (*Practice Test*) which increases an individual's observed *Test Score* by 10 points. However, in this case, the practice test does not affect the underlying construct, *Continuous Skill* or by extension, *Earnings*. We present the results of this simulation in Table 6, Panel C. Absent a control for *Practice Test* in column 1, the coefficient on *Test Score* (401) attenuates relative to the true relation between *Continuous Skill* and *Earnings* of $500. However, in column 2, the coefficient no longer suffers from attenuation bias because, conditional upon *Practice Test*, we no longer measure *Continuous Skill* with error.[28] The negative coefficient on *Practice Test* in column 2 may seem surprising since *Practice Test* does not have a negative causal effect on *Earnings*. However, for two people with the same *Test Score*, an individual who had taken the

[28] In reality, a test will measure skill with error, so even if a practice test increased *Test Score* by a defined amount and we had data on this variable, the coefficient on *Test Score* would still be attenuated even with a control for *Practice Test*. For demonstration purposes, we did not add additional error to *Test Score*.

practice test actually has a 10 point lower *Continuous Skill* and therefore lower expected earnings (10 points × \$500 per point resulting in an expected -\$5,000 coefficient). *Test Score* in this case is a bad control for the effect of *Practice Test* on *Earnings* (which is zero by construction). This emphasizes an important point that control variable coefficient estimates rarely reflect true causal estimates and should be interpreted with caution.

4.3 Fixed Effects

We consider "fixed effects" to fall under the purview of a discussion on controls since they are simply a series of "dummy" controls. Fixed effects isolate within-group (e.g., company, industry, year) variation in the treatment and outcome. When using fixed effects, it is important to consider the source of within-group variation. In some cases, fixed effects improve causal interpretation. However, they can also isolate non-generalizable or endogenous variation.[29]

4.3.1 Fixed effects isolate non-generalizable variation

Suppose we are interested in the impact of an audit committee accounting expert (hereafter ACAE) on the occurrence of fraud. If unobservable but unchanging ("fixed") company factors such as culture relate to having an ACAE and the likelihood of fraud, we may consider including company fixed effects in our regression. However, using company fixed effects in this setting raises some important considerations. First, treatment effects may not be homogeneous. For example, Glaeser and Guay (2017) highlight that marginal compliers may have different treatment effects than average compliers. In our setting, companies with an ACAE for the entire period ("always takers") may experience the greatest benefit of an ACAE in terms fraud reducing corporate governance effects, while companies that experience changes in ACAE at some point during the sample period ("sometimes takers") may experience a lower benefit.

[29] See deHaan (2021) for additional discussion on the use of fixed effects in accounting research, including econometric concerns arising from improper fixed effects inclusion.

To illustrate, we simulate a setting where ACAE has a causal effect on fraud that differs for "always takers" versus "sometimes takers" using the following process and parameters:

B1) Create a panel dataset of 5,000 companies with 10 years of data each: 50,000 observations total
B2) Set 40 percent of companies as "always takers" (*AT*): P(*ACAE* | *AT*) = 1
B3) Set 20 percent of companies as "sometimes takers" (*ST*) changing to and retaining a financial expert at a random year in the sample: P(*ACAE* | *ST*) = 0.5 [30]
B4) Set the remaining companies as "never takers" (*NT*): P(*ACAE* | *NT*) = 0
B5) The rate of fraud is 2.5% for "always takers" P(*FRAUD* | *AT*) = 2.5%
B6) The rate of fraud is 5.0% for "sometimes takers" with experts P(*FRAUD* | *ST* ∩ *ACAE*) = 5.0%
B7) The rate of fraud is 7.5% for companies without experts P(*FRAUD* | *ACAE* = 0) = 7.5%

Given these parameters, the fraud rate is: 2.5% for "always takers," 5% for "sometimes takers" with an ACAE, and 7.5% for all company types with no ACAE. Here, "always takers" experience a greater effect of an ACAE (5% fraud reduction) than "sometimes takers" (2.5% fraud reduction). We present the descriptive statistics for this simulation in Table 7 Panel A and present estimates of the following two models in Table 7 Panel B:

$$Fraud_{it} = \beta_0 + \beta_1 ACAE_{it} + \varepsilon_{it} \quad [5a]$$

$$Fraud_{it} = \beta_0 + \beta_1 ACAE_{it} + Firm\,Fixed\,Effects + \varepsilon_{it} \quad [5b]$$

The estimate in column 1, -4.3%, reflects the difference in means between ACAE groups from Panel A and approximates the average effect of an ACAE on the likelihood of fraud for the entire sample.[31] In column 2, the effect declines to -1.7%.[32] This occurs because within-company variation in ACAE only occurs for "sometimes takers" and fixed effects isolate the effect for companies that experience a change in ACAE.

A related issue occurs if the underlying construct is sticky but the variable is noisily

[30] For simplicity, we do not simulate companies that change away from an ACAE. However, this design choice does not affect the inferences drawn from the simulation.
[31] ACAE reduces fraud by 5% for 80% of the sample and by 2.5% for 20% of the sample (sample average of 4.5%).
[32] This estimate deviates from the expected value of -0.025 since we only ran the simulation once. Repeating the simulation produces a range of estimates that converge to -0.025.

measured. In this case, cross-sectional variation in the variable may correlate (even strongly) with the construct, but within-group variation does not reflect "real" variation in the construct. In fact, group fixed effects may isolate the measurement error. As an example, consider evaluating the effect of geographic religiosity on fraud where the variable used to capture religiosity is derived from an annual survey and reflects the strength with which people in different U.S. states identify as religious. Intuition suggests that religiosity is fairly "sticky" and changes, if any, occur gradually, but survey sampling error may give the appearance of year-over-year changes in religiosity within a state. However, this variation largely reflects noise rather than actual changes in underlying religiosity. As such, state fixed effects (or company fixed effects) analyses will not yield reliable estimates of the effect of religiosity level on fraud.

4.3.2 Fixed effects isolate endogenous variation

In some cases, fixed effects can isolate endogenous within-group variation. Continuing the ACAE example above, suppose that companies tend to add an ACAE following an instance of fraud (possibly to address concerns about weak corporate governance). To reflect this condition, we stipulate that companies without an ACAE that experience fraud add (and retain for the remainder of the sample period) an ACAE following fraud 50 percent of the time. Using the new simulated data, we estimate models 5a and 5b from above and present the findings in Table 7, Panel C. In column 1 with no fixed effects, the effect of an ACAE on fraud is similar to the original simulation in Panel B, though the estimate in Panel C is closer to the ACAE effect for "sometimes takers" than Panel B because the new parameter increased the number of "sometimes takers" with an ACAE. However, the specification with company fixed effects in column 2 suggests a negative relation between ACAE and fraud that greatly exceeds the true effect for "sometimes takers." This occurs because within-company variation in ACAE occurs disproportionately for companies that have a fraud while having no ACAE. The companies that add an ACAE due to a prior fraud must have had no ACAE during the fraud and fraud occurs infrequently after the switch (this would be the case even if the company had not switched to an ACAE). As such, the negative coefficient captures the reverse causality of a fraud occurrence triggering the addition of an ACAE.[33] This illustrates that fixed effects can magnify an endogenous relation between variables.

4.4 Controls for Interactive Test Variables

Accounting studies frequently investigate whether another variable, *I*, moderates the relation between *X* and *Y* by including an interaction between *X* and *I*. If *X* correlates with *Z*, the interaction $X \times I$ also plausibly correlates with interactions, $Z \times I$. Thus, failure to include $Z \times I$ may lead to OVB if $Z \times I$ determines *Y*.[34] To demonstrate, we construct a simulation using financial data extracted from Compustat and Audit Analytics for the period 2005-2019. We construct a *Y* (*Outcome*) that is determined by *ln(Assets)* but with a relation that changes based on a random event (*Event*) that occurs for 50 percent of the population. We generate *Outcome* using the following function:

$$Outcome_{it} = 0.35ln(Assets)_{it} - 0.15ln(Assets)_{it} \times Event_{it} + \varepsilon_{it} \quad [6]$$

Where ε_{it} is a normally distributed error with a mean of 0 and a standard deviation of 0.50. To illustrate OVB in interactive settings, we select two other variables which have some

[33] When considering the source of variation (e.g., cross-sectional vs. within-company), it is worthwhile to consider how much variation in *X* occurs between groups. Here, it may be useful to regress *X* on the fixed effects. A high R^2 from this regression will indicate that most variation in *X* occurs between groups (not much variation within groups), while a low R^2 indicates more variation within groups. See Armstrong, Glaeser, Huang, and Taylor (2019) for an example.

[34] For a discussion of common misconceptions in the interpretation of coefficients on interactions, we refer the reader to Burks, Randolph, and Seida (2019). We focus on the appropriate control variables when the test variable is an interaction.

relation with *ln(Assets)*: return on assets (*ROA*) and internal control weaknesses (*Weak*).[35] We provide variable definitions and descriptive statistics in Table 8, Panel A and a correlation matrix in Table 8, Panel B. *ROA and Weak* are correlated with *ln(Assets)* and *ln(Assets)* × *Event* is correlated with each interaction*, Z* × *Event*, even though *Event* is uncorrelated with the controls. Thus, a model with one or more interactions, *Z* × *Event,* may suffer from OVB if the model does not include *ln(Assets)* × *Event* as a control. For example, we could propose the following regression to test whether *Event* moderates the relation between *ROA* and *Outcome*.

$$Outcome_{it} = \beta_0 + \beta_1 ln(Assets)_{it} + \beta_2 ROA_{it} \times Event_{it} + \beta_3 ROA_{it} + \beta_4 Weak_{it} + \beta_5 Event_{it} + \varepsilon_{it} \quad [7]$$

We estimate [7] in Table 8 Panel C, column 1. Column 2 is similar except that we interact *Weak* with *Event* (i.e., *Event* moderates the relation between *Weak* and *Outcome*). These specifications resemble those we frequently encounter in accounting research whereby the moderating variable is only interacted with *X* and not *Z*. Although *ROA* and *Weak* have no main or interactive effect on *Outcome* by construction (i.e., [6]), we observe significance on each interaction *and* significance on the related uninteracted variable when we exclude *ln(Assets)* × *Event* from the regression, even though we control for *ln(Assets)*. We could conclude, based on column 1, that *ROA* has a positive effect on *Outcome* when *Event* = 0 that is neutralized when the event occurs. However, after controlling for *ln(Assets)* × *Event* (columns 3 and 5), these relations disappear. *Weak* and *Weak* × *Event* exhibit similar tendencies.

We can test interactive effects by partitioning the sample on a discrete variable, *I*, and comparing coefficient estimates between samples. This approach is econometrically equivalent

[35] To limit the effect of outliers, we drop observations with assets less than one million, censor ROA at -100%, and winsorize continuous variables at the 1st and 99th percentiles.

to interacting all variables (*X, Z*, and any fixed effects) with the partitioning variable, *I*.[36] Often, this approach is simpler than presenting a large slate of interactions and easier for the reader to consume. We present partitioned regressions in columns 6 and 7. We note that only *ln(Assets)* significantly predicts *Outcome* in both samples, and that the difference, -0.15 (column 8), approximates the interactive effect. Further note that the coefficient estimates and test statistics in column 6 are identical to the main effects in column 5, and the coefficient differences (column 8) are identical to the relevant interactive effects in column 5.

Sample partitions or fully-interacted models are useful in most settings to address potentially omitted interaction variables. However, there are some settings where interacted controls may be unnecessary. As discussed throughout, theory should guide which controls the model includes.

5. Concluding Remarks and Suggestions for Future Research

We conclude by providing some suggestions and best practices for future studies. While the growing use of as-if random variation in treatment solves many of the issues we discuss, we expect researchers will continue to rely on statistical control and observational data to make causal inference. Because appropriate research design depends on the nature of the research question and underlying data, these suggestions are general in nature and not a panacea. We hope our suggestions are useful and help improve the quality of this type of research.

1) *Begin with a thought experiment using a simple correlation between Y and X and identify "alternative explanations" for the relation* – By starting with a simple correlation between *Y* and *X*, researchers can more readily identify alternative explanations, which will assist in the

[36] We intentionally selected an a randomly assigned *I* but note that studies often test interactive effects of endogenous variables from *Z*. While this is beyond the scope of our discussion, interacting *I* with all *Z* may not sufficiently address OVB when *I* is an endogenous variable.

selection of Z. This mindset will generally lead to the selection of "good" controls (confounders) and is unlikely to lead the researcher to include outcomes of X (mediators) or Y (colliders) as these variables cannot easily be motivated as "alternative explanations."

2) *Use causal diagrams to identify causal mechanisms* – Describing causal mechanisms and the direction of causal effects can facilitate identification of "good" and "bad" controls. We highly encourage researchers use these tools when designing empirical tests.

3) *Consider the timing of variable measurement* – "Good" controls capture constructs that are pre-determined at the time of measurement of the treatment (X) (Angrist and Pischke 2015). If a variable is measured after the treatment, it could be an outcome of X or Y. While all variables measured after X or Y are not necessarily "bad" controls, this exercise is a good starting point for the identification of "bad" controls.

4) *Interpret the model in light of the included controls* – Researchers should consider the feasibility of holding Z constant while varying X and/or Y. If this seems infeasible, then it indicates that Z is likely a "bad" control. Additionally, while mediators are generally "bad" controls if the researcher wants to investigate "full" causal effects, they may be appropriate if a researcher is focused on a "direct effect." In this case, researchers should discuss the estimates in light of included mediator(s). On the other hand, it is hard to envision scenarios when controlling for "outcomes of the outcome," or colliders, will yield informative estimates.

5) *Consider measurement error in control variables* – While there is no easy fix for measurement error, we recommend that researchers continue to (1) recognize the potential effects of measurement error, (2) pursue better measures of important constructs, (3) identify settings where measurement error may be less pervasive, and (4) control for measurement

error, where feasible.

6) *Interactions* – Consider the potential need for interactive controls as uninteracted controls may not fully address OVB for interacted test variables. Alternatively, sample partitions with tests of differences, in lieu of interactions, emulates a "fully interacted" model.

7) *Present models with and without certain controls* – Many variables contain aspects of "good" and "bad" control. In these instances, we suggest displaying results with and without the control and explaining why such variables meet the criteria of good and/or bad controls (Oster 2019). Stock and Watson (2011 p. 233) advocate for the definition of a base specification that includes a "core or base set of regressors [selected] using a combination of expert judgment, economic theory, and knowledge of how the data were collected…" Researchers can augment the base specification with suspect controls, which aid in the understanding of how various controls impact inferences. When controls alter inferences, researchers should rely on their own expertise and theory to understand the differences.[37]

8) *Utilize as-if random variation if possible* – As-if random variation does not require control variables for unbiased estimation. In these settings, controls that predict Y may improve estimate precision (see Section 4.2.1), but "bad controls" can introduce endogeneity, rendering an otherwise effective setting ineffective. If controls materially alter inferences in these settings, it is worth considering whether treatment variation is indeed "as-if random." In these settings, we recommend reporting results with and without controls (see previous suggestion) to demonstrate that results are not sensitive to specification choices.

[37] An expanded model including a Z with limited coverage may yield significantly different estimates on X than the reduced form. To assess whether sample attrition or OVB drives this difference, we suggest estimating the reduced model on the limited sample with Z coverage and comparing estimates to the expanded model including Z. If coefficients are similar, this suggests that differences in the treatment effect from the full and limited sample may relate to sample composition differences, rather than OVB in the reduced form.

References

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*: Princeton University Press.

Angrist, J. D., and J.-S. Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect*: Princeton University Press.

Armstrong, C. S., S. Glaeser, S. Huang, and D. J. Taylor. 2019. The Economics of Managerial Taxes and Corporate Risk-Taking. *The Accounting Review* 94 (1): 1-24.

Atinc, G., M. J. Simmering, and M. J. Kroll. 2012. Control Variable Use and Reporting in Macro and Micro Management Research. *Organizational Research Methods* 15 (1): 57-74.

Ayres, I. 2005. Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of "Included Variable" Bias. *Perspectives in Biology Medicine* 48 (1 Supplement): S68-S87.

Becker, T. E. 2005. Potential Problems in the Statistical Control of Variables in Organizational Research: A Qualitative Analysis with Recommendations. *Organizational Research Methods* 8 (3): 274-289.

Bertomeu, J., A. Beyer, and D. J. Taylor. 2016. From Casual to Causal Inference in Accounting Research: The Need for Theoretical Foundations. *Foundations and Trends in Accounting* 10 (2-4): 262-313.

Bloomfield, R., M.W. Nelson, and E. Soltes. 2016. Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accounting Research* 54 (2): 341-395.

Burks, J. J., D. W. Randolph, J. A. Seida. 2019. Modeling and Interpreting Regressions with Interactions. *Journal of Accounting Literature* 42. 61-79.

Carlson K. D., and J. Wu. 2012. The Illusion of Statistical Control: Control Variable Practice in Management Research. *Organizational Research Methods* 15 (3): 413-435.

deHaan, E. 2021. Using and Interpreting Fixed Effects Models. *Working Paper*

deHaan, E., A. Lawrence, and R. Litjens. 2021. Measurement Error in Google Ticker Search. *Working Paper*.

Francis, J., D, Nanda, and P, Olsson. 2008. Voluntary Disclosure, Earnings Quality, and Cost of Capital. *Journal of Accounting Research* 46 (1): 53-99.

Garfinkel, J. A. 2009. Measuring Investors' Opinion Divergence. *Journal of Accounting Research* 47 (5): 1317-1348.

Glaeser, S., and W. R. Guay. 2017. Identification and Generalizability in Accounting Research: A Discussion of Christensen, Floyd, Liu, and Maffett (2017). *Journal of Accounting and Economics* 64 (2-3): 305-312.

Gow, I. D., D. F. Larcker, and P. C. Reiss. 2016. Causal Inference in Accounting Research. *Journal of Accounting Research* 54 (2): 477-523.

Greene, W. H. 2012. *Econometric Analysis*. Pearson Education.

Guay, W., D. Samuels, and D. Taylor. 2016. Guiding Through the Fog: Financial Statement Complexity and Voluntary Disclosure. *Journal of Accounting and Economics* 62 (2-3): 234-269.

Ittner, C. D. 2014. Strengthening Causal Inferences in Positivist Field Studies. *Accounting, Organizations and Society* 39 (7): 545-549.

Jennings, J. N., J. M. Kim, J. A. Lee, and D. J. Taylor. 2021. Measurement Error and Bias in Causal Models in Accounting Research. *Working Paper.*

Larcker, D. F., and T. O. Rusticus. 2010. On the Use of Instrumental Variables in Accounting Research. *Journal of Accounting & Economics* 49 (3): 186-205.

Lawrence, A., M. Minutti-Meza, and P. Zhang. 2011. Can Big 4 versus Non-Big 4 Differences in Audit-Quality Proxies Be Attributed to Client Characteristics? *The Accounting Review* 86 (1): 259-286.

Lennox, C. S., J. R. Francis, and Z. Wang. 2012. Selection Models in Accounting Research. *The Accounting Review* 87 (2): 589-616.

Leone, A. J., M. Minutti-Meza, and C. E. Wasley. 2019. Influential Observations and Inference in Accounting Research. *The Accounting Review* 94 (6): 337-364.

Maxwell, S. E., and H. D. Delaney. 1990. *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Wadsworth Publishing.

Oster, E. 2019. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics* 37 (2): 187-204.

Pearl, J. 1995. Causal Diagrams for Empirical Research. *Biometrika* 82 (4): 669-688.

Shipman, J. E., Q. T. Swanquist, and R. L. Whited. 2017. Propensity Score Matching in Accounting Research. *The Accounting Review* 92 (1): 213-244.

Spector, P. E., and M. T. Brannick. 2011. Methodological Urban Legends: The Misuse of Statistical Control Variables. *Organizational Research Methods* 14 (2): 287-305.

Speklé, R. F., and S. K. Widener. 2018. Challenging Issues in Survey Research: Discussion and Suggestions. *Journal of Management Accounting Research* 30 (2): 3-21.

Stock, J. H., and M. W. Watson. 2011. *Introduction to Econometrics* (3rd edition). Addison-Wesley.

Tucker, J. W. 2010. Selection Bias and Econometric Remedies in Accounting and Finance Research. *Journal of Accounting Literature* 29: 31-57.

Westfall, J., and T. Yarkoni. 2016. Statistically Controlling for Confounding Constructs Is Harder Than You Think. *PLoS ONE* 11 (3): 1-22.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Wooldridge, J. M. 2013. *Introductory Econometrics: A Modern Approach.* Cengage Learning.

Figure 1: Example Causal Diagram

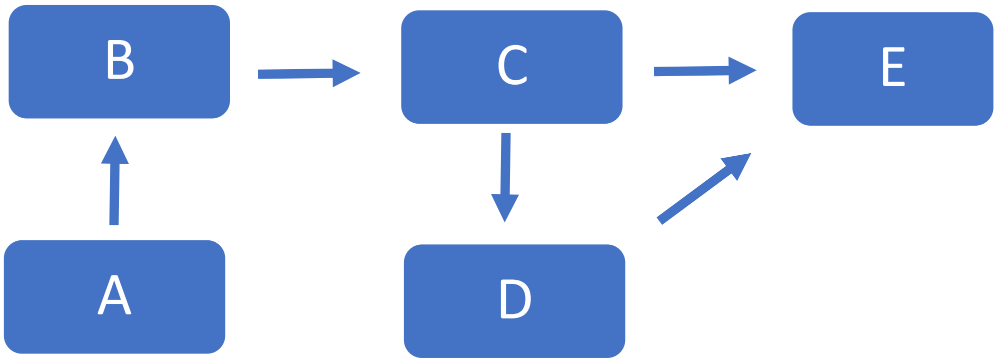

Figure 2a: Confounder Control: ***What is the effect of the CPA certification on earnings?***

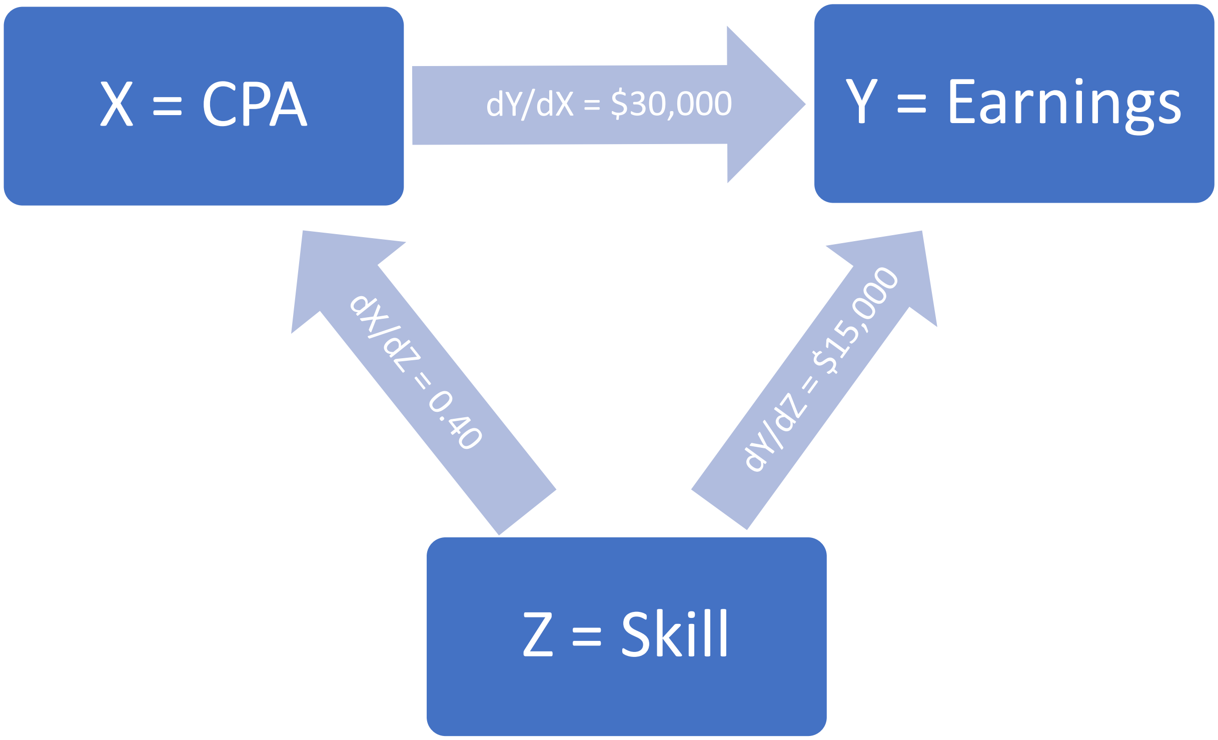

Figure 2b: Confounder Control: ***Do Big 4 auditors charge higher fees?***

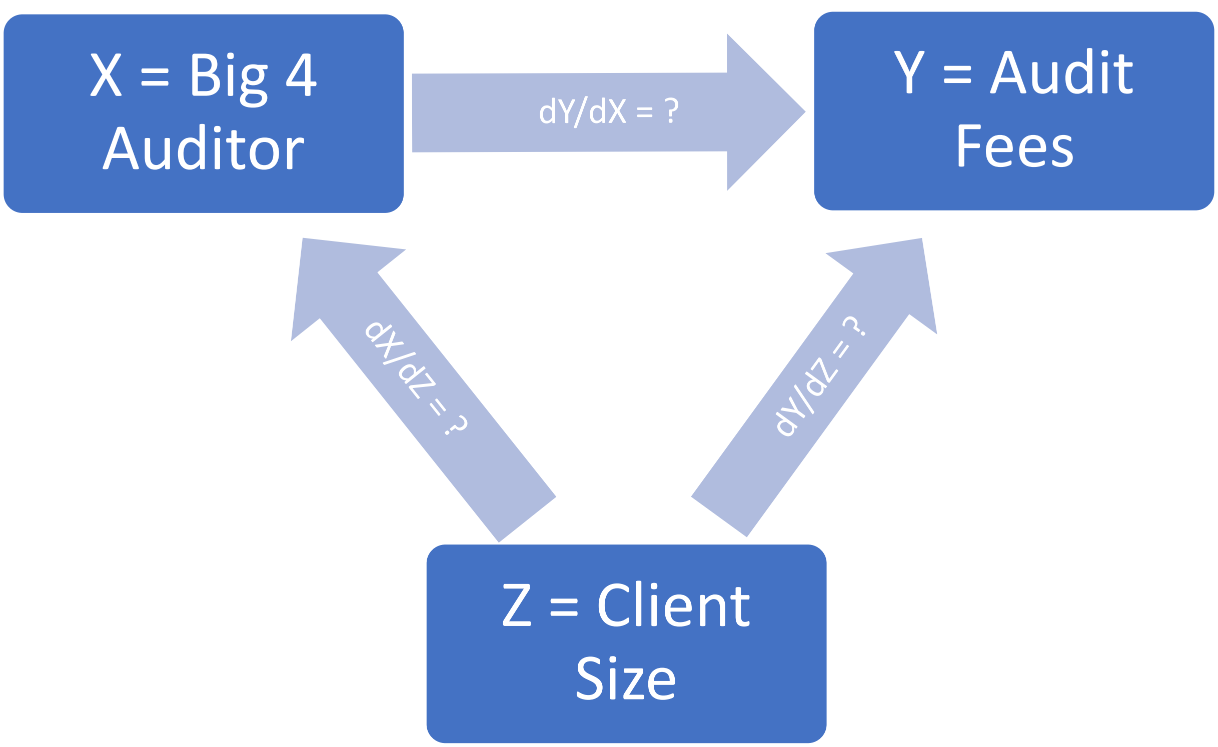

Figure 3a: Mediator Control: ***What is the total effect of accounting skill on earnings?***

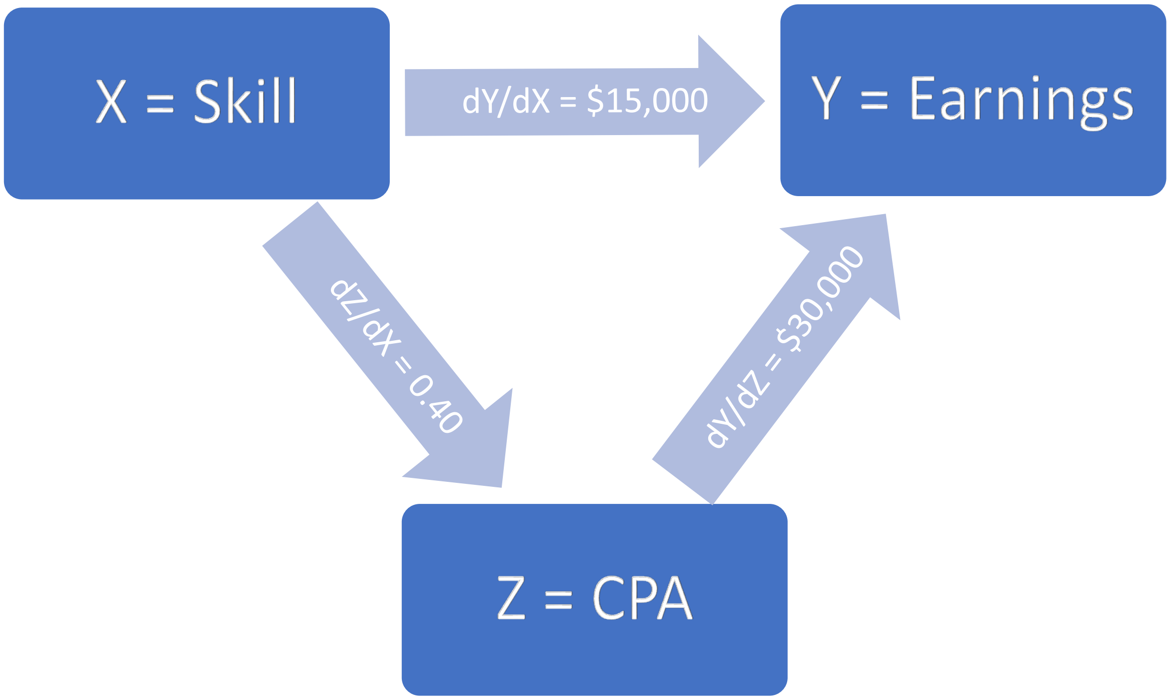

Figure 3b: Mediator Control: ***Do larger companies pay higher audit fees?***

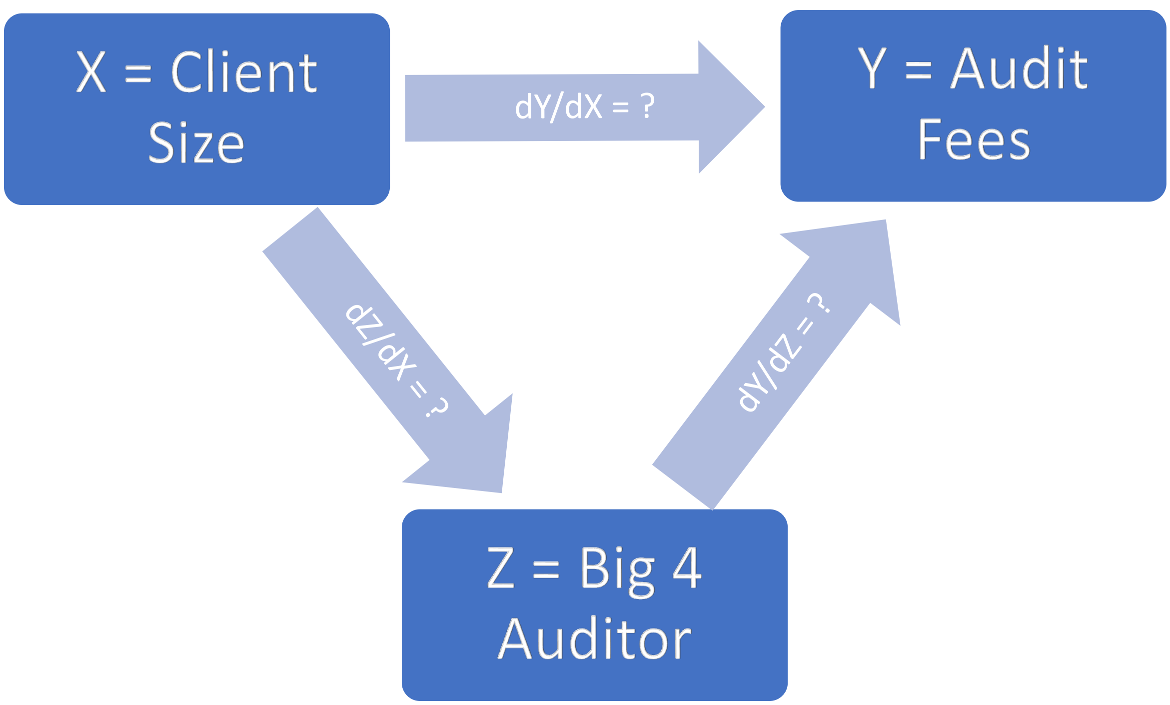

Figure 4a: Collider Control: ***Are higher skilled accountants more likely to be CPAs?***

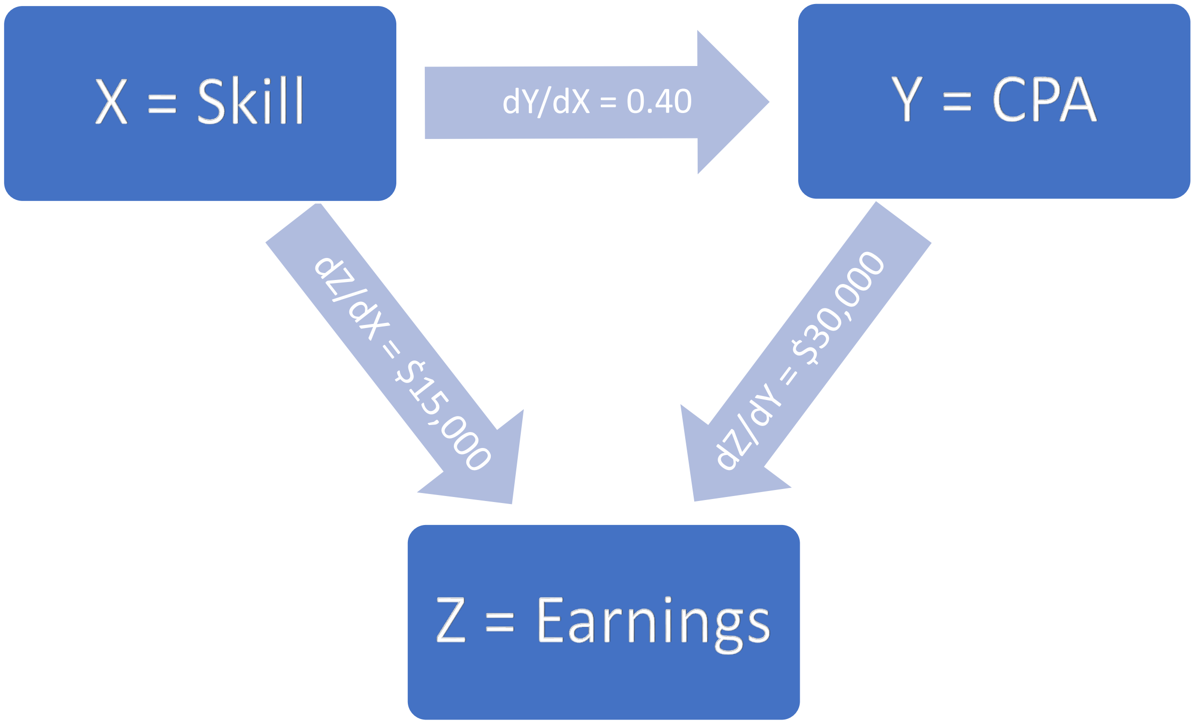

Figure 4b: Collider Control: ***Do larger clients tend to select Big 4 auditors?***

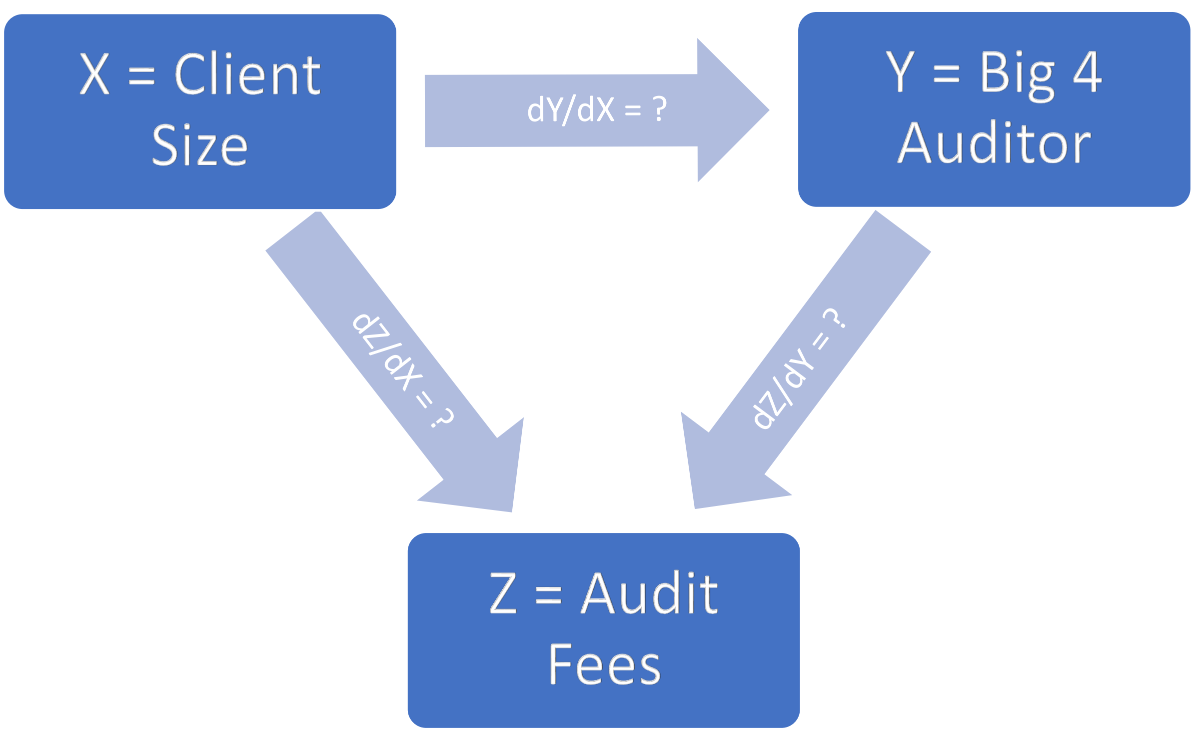

Figure 5: Graph of Coefficient Estimates from Table 2 with Noise Added to Control Variable

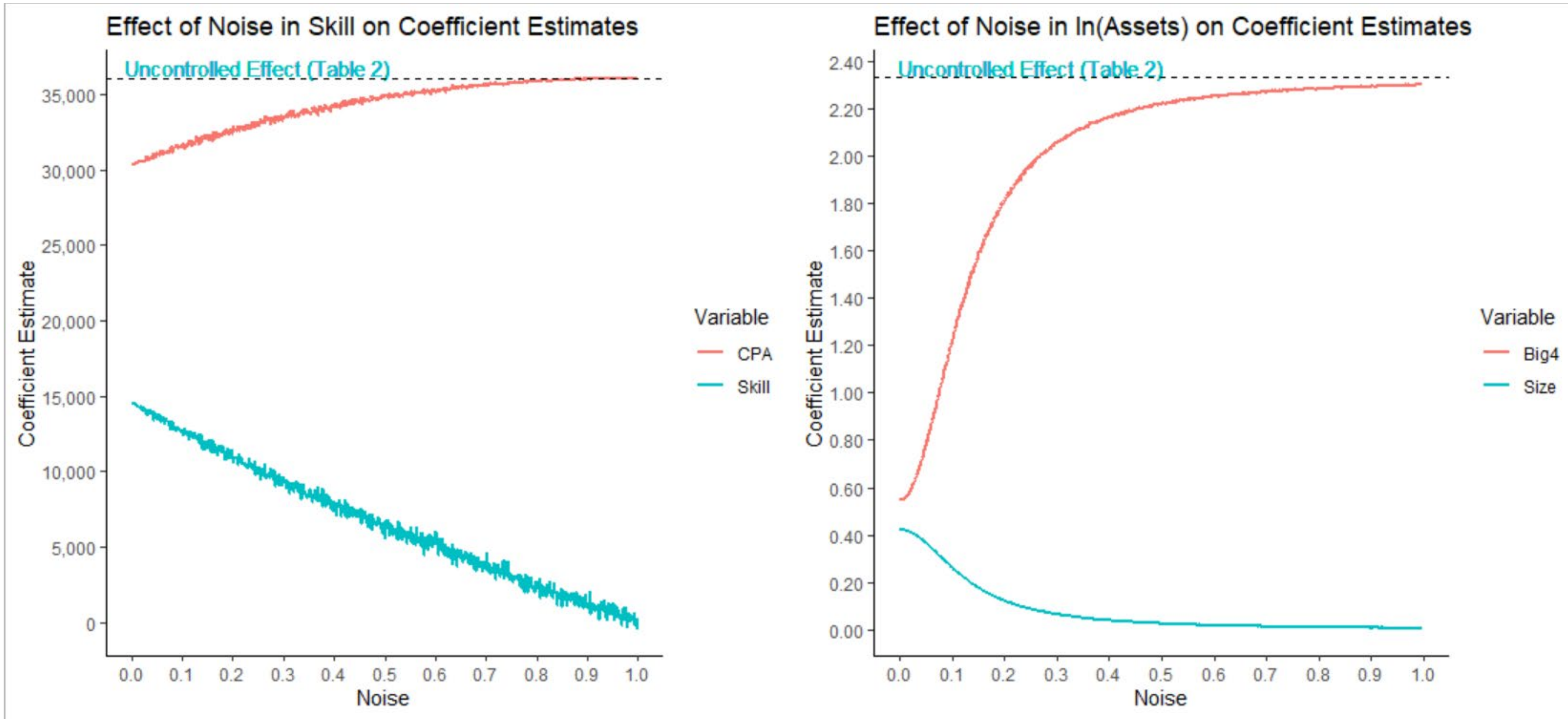

Table 1: Descriptive Statistics for Illustrations

Panel A: Descriptive statistics for simulated CPA data (simulation setting)

Descriptive statistics:

	Full Sample Mean
Skill	50.0%
CPA	50.1%
Average *Earnings*	$72,444

	***Skill* = 1**	***Skill* = 0**	**Diff**	**t-stat**	
CPA	69.8%	30.4%	39.4%	30.34	***
Earnings	$85,727	$59,161	$26,565	54.51	***

Panel B: Descriptive statistics for AA data with fees and assets in non-financial industries (auditor setting)

Attrition:	Observations
Unique company-years in AA with fees and assets data from 2003-2015	107,131
Less: Financial and utilities industries	(37,670)
Final AA Sample	69,461

Descriptive statistics:

	Observations	**Full Sample Mean**
Big4 = 0	24,770	36%
Big4 = 1	44,691	64%
Total	69,461	100%

	***Big4* = 1**	***Big4* = 0**	**Diff**	**t-stat**	
ln(Fees)	14.06	11.73	2.33	223.91	***
ln(Assets)	20.42	16.19	4.23	217.84	***

*, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).

Table 2: Illustration of Confounder Controls

Panel A: Simulation setting ([1a] and [1b]): What is the effect of the CPA certification on earnings?

VARIABLES	(1) *Earnings*	(2) *Earnings*
CPA	36,104.60***	30,347.94***
	(105.14)	(97.43)
Skill		14,595.96***
		(46.86)
Constant	54,362.89***	49,947.84***
	(223.71)	(223.57)
Observations	5,000	5,000
Adjusted R-squared	0.689	0.784

Panel B: Auditor setting ([1c] and [1d]): Do Big 4 auditors charge higher audit fees?

VARIABLES	(1) *ln(Fees)*	(2) *ln(Fees)*
Big4	2.33***	0.55***
	(88.17)	(31.17)
ln(Assets)		0.42***
		(113.81)
Constant	11.73***	4.91***
	(601.93)	(80.51)
Observations	69,461	69,461
Adjusted R-squared	0.419	0.778

All models estimated using OLS. Standard errors are clustered by company in Panel B. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).

Table 3: Illustration of Mediator Controls

Panel A: Simulation setting ([2a] and [2b]): What is total effect of accounting skill on earnings?

VARIABLES	(1) *Earnings*	(2) *Earnings*
Skill	26,565.19***	14,595.96***
	(54.51)	(46.86)
CPA		30,347.94***
		(97.43)
Constant	59,161.48***	49,947.84***
	(171.67)	(223.57)
Observations	5,000	5,000
Adjusted R-squared	0.373	0.784

Panel B: Auditor setting ([2c] and [2d]): Do larger companies pay higher audit fees?

VARIABLES	(1) *ln(Fees)*	(2) *ln(Fees)*
ln(Assets)	0.47***	0.42***
	(159.27)	(113.81)
Big4		0.55***
		(31.17)
Constant	4.27***	4.91***
	(78.81)	(80.51)
Observations	69,461	69,461
Adjusted R-squared	0.765	0.778

All models estimated using OLS. Standard errors are clustered by company in Panel B. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).

Table 4: Illustration of Collider Controls

Panel A: Simulation setting ([3a] and [3b]): Are higher skilled individuals more likely to be CPAs?

	(1)	(2)
VARIABLES	*CPA*	*CPA*
Skill	0.39***	-0.18***
	(30.34)	(-18.57)
Earnings[a]		0.02***
		(97.43)
Constant	0.30***	-0.97***
	(33.03)	(-68.67)
Observations	5,000	5,000
Adjusted R-squared	0.156	0.709

Panel B: Auditor setting ([3c] and [3d]): Do larger clients tend to select Big 4 auditors?

	(1)	(2)
VARIABLES	*Big4*	*Big4*
ln(Assets)	0.10***	0.05***
	(119.29)	(24.04)
ln(Fees)		0.11***
		(29.53)
Constant	-1.17***	-1.63***
	(-72.43)	(-73.26)
Observations	69,461	69,461
Adjusted R-squared	0.406	0.441

All models estimated using OLS. Standard errors are clustered by company in Panel B. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).
[a] value scaled by 1,000 for expositional purposes.

Table 5: Illustration of Same Construct Controls

VARIABLES	(1) *ln(Fees)*	(2) *ln(Fees)*
Big4	0.55***	0.17***
	(31.17)	(7.05)
ln(Assets)	0.42***	0.41***
	(113.81)	(106.21)
ln(Auditor Clients)		0.11***
		(19.32)
Constant	4.91***	4.79***
	(80.51)	(80.33)
Observations	5,000	69,461
Adjusted R-squared	0.778	0.782

All models estimated using OLS. Standard errors are clustered by company. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).

Table 6: Illustration of Controls Related to either *X* or *Y*

Panel A: Cost of living control: What is total effect of accounting skill on earnings?

	(1)	(2)
VARIABLES	*Earnings*	*Earnings*
Skill	26,565.19***	26,668.69***
	(54.51)	(60.37)
Cost of Living		1.02***
		(32.96)
Constant	59,161.48***	59,310.57***
	(171.67)	(189.85)
Observations	5,000	5,000
Adjusted R-squared	0.373	0.485

Panel B: CPA test prep control: What is the effect of the CPA certification on earnings?

	(1)	(2)
VARIABLES	*Earnings*	*Earnings*
CPA	30,188.58***	30,347.94***
	(105.46)	(97.43)
CPA Prep		-404.04
		(-1.30)
Constant	49,825.63***	49,947.84***
	(245.95)	(223.57)
Observations	5,000	5,000
Adjusted R-squared	0.690	0.690

Panel C: Practice tests and continuous skill: What is total effect of accounting skill on earnings?

VARIABLES	(1) *Earnings*	(2) *Earnings*
Test Score	401.05***	496.60***
	(30.45)	(34.69)
Practice Test		-4,861.69***
		(-15.27)
Constant	32,820.73***	25,208.55***
	(23.56)	(17.38)
Observations	5,000	5,000
Adjusted R-squared	0.156	0.194

All models estimated using OLS. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).

Table 7: Illustrations using Fixed Effects

Panel A: Descriptive statistics for simulated ACAE and fraud data

Descriptive statistics:

	Observations
ACAE	50.0%
Fraud	5.2%
Observations	50,000

	ACAE = 1	*ACAE* = 0	Diff	t-stat	
Fraud	7.40%	3.08%	4.32%	21.78	***

Panel B: Fixed effects using the simulated ACAE and fraud data

	(1)	(2)
VARIABLES	*Fraud*	*Fraud*
ACAE	-4.32***	-1.71***
	(-21.78)	(-3.27)
Fixed Effects	No	Yes
Observations	50,000	50,000
Adjusted R-squared	0.01	0.11

Panel C: Fixed effects using the simulated ACAE and fraud data with reverse causality

	(1)	(2)
VARIABLES	*Fraud*	*Fraud*
ACAE	-3.78***	-11.47***
	(-18.84)	(-27.87)
Fixed Effects	No	Yes
Observations	50,000	50,000
Adjusted R-squared	0.01	0.12

All models estimated using OLS. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests). Estimates in Panels A and B are multiplied by 100 for expositional purposes.

Table 8: Interaction Setting using Archival Data and a Simulated Outcome

Panel A: Descriptive statistics for archival data with simulated outcome

VARIABLES	Observations	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
Outcome	109,909	1.68	1.01	0.96	1.60	2.32
ln(Assets)	109,909	6.11	2.63	4.21	6.21	7.93
ROA	109,909	-0.08	0.28	-0.08	0.01	0.05
Weak	109,909	0.07	0.26	0.00	0.00	0.00
Event	109,909	0.50	0.50	0.00	1.00	1.00
ln(Assets) × Event	109,909	3.07	3.58	0.00	0.43	6.22
ROA × Event	109,909	-0.04	0.20	0.00	0.00	0.01
Weak × Event	109,909	0.04	0.19	0.00	0.00	0.00

Panel B: Correlation matrix

VARIABLES	*Outcome*	*ln(Assets)*	*ROA*	*Weak*	*Event*	*ln(Assets)× Event*	*ROA × Event*	*Weak × Event*
Outcome	**1.00**							
ln(Assets)	**0.71**	**1.00**						
ROA	**0.35**	**0.50**	**1.00**					
Weak	**-0.13**	**-0.18**	**-0.16**	**1.00**				
Event	**-0.45**	0.00	0.00	0.00	**1.00**			
ln(Assets) × Event	**-0.20**	**0.37**	**0.18**	**-0.06**	**0.85**	**1.00**		
ROA × Event	**0.27**	**0.34**	**0.69**	**-0.12**	**-0.20**	**0.09**	**1.00**	
Weak × Event	**-0.16**	**-0.13**	**-0.11**	**0.70**	**0.19**	**0.07**	**-0.20**	**1.00**

Bold indicates significance at the 0.01 level (using two-tailed tests).

Panel C: Omitted variable bias in interactive settings

						Partitioned Regression		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	*(8)*
	Outcome	*Outcome*	*Outcome*	*Outcome*	*Outcome*	*Outcome*	*Outcome*	*(7) - (6)*
VARIABLES						*Event = 0*	*Event = 1*	
ln(Assets)	0.27***	0.27***	0.35***	0.35***	0.35***	0.35***	0.20***	*-0.15****
	(354.22)	(354.64)	(369.56)	(392.68)	(368.03)	(368.03)	(213.27)	*(112.43)*
ROA × Event	-0.69***		0.00		0.00			
	(-55.22)		(0.32)		(0.23)			
Weak × Event		0.26***		-0.01	-0.01			
		(18.18)		(-1.07)	(-1.04)			
ln(Assets) × Event			-0.15***	-0.15***	-0.15***			
			(-113.37)	(-129.12)	(-112.43)			
ROA	0.34***	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	*0.00*
	(37.54)	(-0.65)	(-0.97)	(-1.03)	(-0.90)	(-0.90)	(-0.57)	*(0.23)*
Weak	-0.01	-0.14***	-0.01	-0.00	-0.00	-0.00	-0.02*	*-0.01*
	(-1.18)	(-14.47)	(-1.55)	(-0.31)	(-0.32)	(-0.32)	(-1.86)	*(1.04)*
Event	-0.97***	-0.93***	-0.00	-0.00	-0.00			
	(-203.59)	(-188.47)	(-0.40)	(-0.38)	(-0.20)			
Constant	0.49***	0.48***	0.01	0.01*	0.01	0.01	0.01	*0.00*
	(84.66)	(81.63)	(1.61)	(1.66)	(1.46)	(1.46)	(1.18)	*(0.04)*
Observations	109,909	109,909	109,909	109,909	109,909	54,769	55,140	
Adjusted R-squared	0.724	0.716	0.753	0.753	0.753	0.769	0.521	

ln(Assets) is defined as the natural log of total assets, *ROA* is defined as net income divided by total assets, and *Weak* is set equal to one if the company reports at least one 404(1) material weakness, and zero otherwise. All models estimated using OLS. Standard errors are clustered by company. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively (using two-tailed tests).