

Measurement Error, Fixed Effects, and False Positives in Accounting Research

Jared Jennings
Olin Business School
Washington University in St. Louis

Jung Min Kim
Kellogg School of Management
Northwestern University

Joshua Lee
Marriott School of Business
Brigham Young University

Daniel Taylor*
The Wharton School
University of Pennsylvania

March 2022

Abstract:

We show theoretically and empirically that measurement error can bias in favor of falsely rejecting a true null hypothesis (i.e., a “false positive”) and that regression models with high-dimensional fixed effects can exacerbate measurement error bias and increase the likelihood of false positives. We replicate inferences from prior work in a setting where we are able to directly observe the amount of measurement error, and show that the combination of measurement error and fixed effects materially inflates coefficients and distorts inferences. We provide researchers with a simple diagnostic tool to assess the possibility that the combination of measurement error and fixed effects might give rise to a false positive, and encourage researchers to triangulate inferences across multiple empirical proxies and multiple fixed effect structures.

Keywords: measurement error, fixed effects, causal models, accounting research

*Corresponding author: dtayl@wharton.upenn.edu. We thank Beth Blankespoor, Ed deHaan, Patricia Dechow (editor), Ian Gow, Wayne Landsman, Alexander Ljungqvist, Nathan Marshall, Brian Miller, Robbie Moon, Cathy Schrand, Sarah Zechman, Frank Zhou, Christina Zhu, two anonymous reviewers, and seminar participants at the JFR mini-method conference, the City University of Hong Kong, University of Georgia, University of Iowa, University of Maryland, Michigan State University, New York University, University of Washington and The Wharton School for helpful comments. We thank our schools for financial support. Headquarter data used in the paper are available on Josh Lee’s website.

1. Introduction

A growing literature in accounting seeks to draw causal inferences from non-experimental data. Perhaps not surprisingly, this trend toward causal inferences has led to an increasing focus on “econometric identification.” Our sense is that most conversations on econometric identification in accounting research—in academic papers, in the editorial and review process, and in seminars—tend to revolve around econometric assumptions related to exogeneity and correlated omitted variables but not measurement error.¹ While a large literature in accounting in the 1980s and 1990s focused on measurement error, our sense is that the contemporary literature underestimates the validity threat posed by measurement error, especially as it relates to causal inferences.² If anything, the threat posed by measurement error is even greater in studies seeking to draw causal inferences, where regression coefficients are often interpreted literally as estimates of economic magnitudes.

To understand the extent to which researchers consider measurement error a serious concern, we survey all papers published in *The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, and *Review of Accounting Studies* over the past decade. We assess the extent to which these papers employ quasi-experimental designs (e.g., difference-in-differences and staggered adoption designs), and the extent to which these papers discuss validity threats arising from correlated omitted variables and measurement error. Panel A of Figure 1 shows that in 2010, under 7% of papers used quasi-experimental designs. By 2019,

¹ Throughout the paper we use the term “measurement error” to refer to the gap between a theoretical construct and its measurement. For example, using log of total assets to measure the construct of firm size.

² As anecdotal evidence we point to the Gow et al. (2016) survey of causal inference in accounting research. In the survey, the phrase “measurement error” never appears; exogen- appears nine times; and endogen- appears seven times. Examples of the earlier literature include Christie (1987), Barth (1991), and Collins, Kothari, Shanken, and Sloan (1994). See Kothari (2001) for a review.

this rate more than quadruples to 33% of papers. Panel B of Figure 1 shows the percentage of papers using quasi-experimental designs that mention correlated omitted variables increases from 36.4% in 2010 to 72.5% in 2019, whereas the percentage that mention measurement error is 45.5% in 2010 and 20.3% in 2019. Interestingly, as interest in causal inferences has increased, concerns about measurement error appear to have declined. These findings appear to bear out our anecdotal experience as authors, reviewers, and editors—as researchers have become increasingly interested in econometric identification and drawing causal inferences, they have become relatively less concerned with measurement error. We aim to correct this oversight: we use a series of simulated and real-world examples to illustrate that measurement error is a credible threat to drawing causal inferences.

In this paper, we show theoretically and empirically that: (1) measurement error can lead researchers to spuriously estimate a causal effect when none exists—a false positive, (2) that the increasingly common practice of including high-dimensional fixed effects increases measurement error bias and the likelihood of false positives, and (3) that these effects are not merely theoretical—we show that they manifest in reported inferences in the literature.

We begin our analysis by providing a brief review of the theoretical effects of measurement error on inferences, and illustrate these effects using three distinct sets of simulations. We use these simulations to develop the intuition for how measurement error can bias in favor of rejecting a true null hypothesis and how fixed effects can exacerbate this bias. By using simulated data, we can be assured that we know the true data generating process (DGP) and that the variable of interest is exogenous.

In the first set of simulations, the DGP for the dependent variable (Y) mirrors the classic case of measurement error in the independent variable of interest (X). The first set of simulations

illustrates the conventional wisdom that measurement error biases against rejecting the null. In the second set of simulations, we allow the measurement error (in X) to be correlated with a noisy control variable (Z). This mirrors the empirical application discussed below—where erroneous information on the state of corporate headquarters from Compustat introduces measurement error that is correlated with included control variables (e.g., size, book-to-market, leverage, etc.). Our simulations show that when measurement error in the independent variable of interest is correlated with a control variable, measurement error can bias the coefficient on the variable of interest away from zero and lead to false positives. The intuition for this result is that when measurement error in the variable of interest is correlated with a control variable, the regression places weight on the variable of interest that would otherwise be placed on the control variable.

In our third set of simulations, we show that high-dimensional fixed effects can exacerbate measurement error bias and lead the researcher to estimate a causal effect when none exists.³ In a survey of empirical studies, deHaan (2020) notes that the justification for the inclusion of fixed effects is often vague, but that the conventional wisdom is that the inclusion of fixed effects—in and of itself—yields more robust and even causal inferences.⁴ Despite the increasing use of fixed effect in accounting research, the econometrics literature warns researchers that fixed effects are

³ The term “high-dimensional fixed effects” refers to the practice of including a large number of fixed effects in the regression specification in an effort to absorb as much variation as possible while still allowing for estimation of the coefficient of interest (e.g., Correia, 2015). This practice is also colloquially known as the “kitchen sink approach to fixed effects” or “saturating the model with fixed effects.” Supporting the growing popularity of this practice, deHaan (2020) reports that 96% of empirical papers published in 2019 in *The Accounting Review*, *Journal of Accounting and Economics*, and *Journal of Accounting Research* have a model that features a fixed effect structure, 84% have a model that features two fixed effect structures, and 29% have a model with three or more fixed effect structures.

⁴ Khan, Serafeim, and Yoon (2016, p. 1699) best summarize the conventional wisdom on fixed effects: “[t]he inclusion of both time and firm fixed effects in the panel regressions is a generalization of the difference-in-differences approach that allows a causal interpretation in a regression setting ... The fixed effects soak up unobserved firm-specific and economy-wide factors that could otherwise cloud identification.”

not always appropriate and can lead to spurious inferences, especially in the presence of measurement error (see discussion in Griliches and Hausman (1986)).

In our third simulation we assume that the DGP for the variable of interest includes a vector of firm fixed effects, and we progressively increase the amount of variation in the independent variable of interest that the fixed effects absorb. For example, if the variable of interest is stock returns, firm fixed effects will absorb very little variation (e.g., stock returns are relatively random), but if the variable of interest is an indicator variable for whether the company has a staggered board, firm fixed effects will absorb substantial variation (e.g., there is very little within-firm variation in board structure). For the purposes of our simulations, we consider absorption rates from 10% to 90%. For reference, in our empirical application, we find firm effects absorb 99% of the variation in the independent variable of interest.⁵

We find that the inclusion of firm fixed effects in the regression specification exacerbates measurement error bias and increases the likelihood of false positives (relative to a specification without firm fixed effects). Indeed, the more variation the fixed effect absorb, the greater the measurement error bias. The intuition for this result is that including fixed effects in the regression specification sucks out the “good variation” in the independent variables (i.e., variation not due to measurement error). As a result, measurement error comprises a greater percentage of the remaining (residual) variation in the independent variable. In essence, fixed effects “throw the baby out with the bath water” which makes the regression specifications more susceptible to the measurement error biases we documented earlier, and in turn, exacerbates the likelihood the researcher finds evidence of a causal effect when none exists.

⁵ See also Armstrong et al. (2021) and Donelson et al. (2022) who document similarly high absorption rates in the context of state income tax rates and universal demand laws respectively.

The preceding analysis suggests that the combination of measurement error and high levels of fixed absorption can generate false positives. Fortunately, the results also suggest a simple diagnostic tool researchers can use to assess the amount of variation in the independent variable of interest that is absorbed by a particular set of fixed effects: the R-squared from a regression of the independent variable of interest on the fixed effect structure. The R-squared from this regression is an estimate of the absorption rate. It indicates the fraction of variation in the independent variable that is removed, or absorbed, by the fixed effects when they are included as a control variable. While we cannot offer a definitive answer on what absorption rate is empirically acceptable, we suggest extreme caution in circumstances when fixed effects absorb more than 90% of the variation in the independent variable of interest.

It is possible that the combination of measurement error and fixed effects with high absorption levels *could* be driving many of the results in the literature, giving rise to a false impression of a causal effect when—in truth—no such effect exists (Ohlson, 2021). To shed light on this possibility, we next examine how measurement error and fixed effects affect inferences in actual research settings considered in the literature. Specifically, we examine a popular quasi-natural experimental setting where we can observe the measurement error in the treatment variable.

To briefly digress, there are two broad sources of measurement error: (i) a mismatch between theoretical construct and empirical proxy, and (ii) incorrect data, and. In some sense, the latter is a specific case of the former. A popular example of the former is the use of the number of independent directors on the board to measure corporate governance (see discussion in Larcker, Richardson, and Tuna, 2007). Popular examples of the latter include retroactive rounding of earnings-per-share (e.g., Payne and Thomas, 2003) and incorrect event dates (e.g., Berkman and Truong, 2009). While this distinction is not important for theoretical considerations of how

measurement error affects regression coefficients, it is of utmost importance in reliably assessing the effect of measurement error on inferences in the literature. For example, measurement error due to incorrect data can be ex post verified with absolute certainty, whereas a mismatch between theoretical construct and empirical proxy is entirely subjective and depends on the implicit theoretical assumptions one is willing to make (e.g., Bertomeu, Beyer, and Taylor, 2016).⁶ So as to avoid any subjective judgements regarding what constitutes measurement error, we focus our empirical analysis on measurement error that stems from incorrect data. Specifically, we focus on measurement error that results when researchers use incorrect data on the state of the firm's corporate headquarters.

A large literature exploits variation in state of corporate headquarters to estimate the causal effects of state regulations and taxes. Our survey in Appendix A reveals more than 100 papers in top journals across accounting and finance that use the state of corporate headquarters in their analysis (to define an independent variable), and that approximately 80% of these papers obtain data on the state of corporate headquarters from Compustat (see Appendix A for survey results). However, Compustat reports only the current location of the firm's headquarters and backfills all observations. For example, if a firm moves its headquarters from Texas to California in 2019, Compustat codes California as the state of headquarters in 2019 and *in all prior years* (e.g., Heider and Ljungqvist, 2015; Jennings, Lee, and Matsumoto, 2017). To measure the extent to which this backfilling introduces measurement error in the historical state of headquarters, we benchmark

⁶ To illustrate this distinction, consider the following passage from Loughran and McDonald (2014, p. 1646) which advocates for measuring 10-K readability using file size (in megabytes) over the Fog Index: "10-K file size is exceptionally easy to determine and is not prone to the substantial measurement errors of other textual procedures requiring parsing of the 10-K documents." While it may be the case that file size is easy to compute, and is measured without error (i.e., we know the number of megabytes with absolute certainty), the measurement error for which file size measures the theoretical construct—readability—could be substantial, and is difficult to assess without experimental data (e.g., Bonsall, Leone, Miller, and Rennekamp, 2017).

Compustat data against the true state of corporate headquarters listed on the firm's annual 10-K filing on EDGAR.⁷ We find error rates on the 2019 Compustat file routinely exceed 10% and approach 20% as one goes back further in time. The notion that more than 100 papers rely on Compustat as a source of corporate headquarters data when such data is prone to extensive errors suggests measurement error bias is a pervasive issue in the literature.

To illustrate the effect of measurement error bias on inferences in the literature, we follow prior research and estimate the “causal effect” of state non-compete laws on voluntary disclosure and earnings management. We begin by showing that measurement error in the non-compete index (attributable to using Compustat state of headquarters) is correlated with firm size, market-to-book ratio, return on assets, leverage, and other common control variables. Thus, our empirical setting matches that contemplated in our simulations—measurement error is correlated with included control variables.

Using the state of headquarters listed on Compustat, we find an index of state non-compete laws computed using the state of headquarters listed on Compustat are negatively related to voluntary disclosure (management forecasts, 8-K filings, and press releases) and positively related to earnings management (intentional restatements).⁸ The results imply a large, economically significant effect of state non-compete laws on these outcome variables. For example, a one standard deviation increase the non-compete index increases the probability of a restatement by 2.54% (where the unconditional mean is 2%). However, when the non-compete index is computed

⁷ This data is being made publicly available with the publication of this paper and is available from the authors upon request.

⁸ We use the Garmaise (2011) index of state non-compete laws updated by Ertimur, Rawson, Rogers, and Zechman (2018). For examples of studies using the Garmaise (2011) index as either a control variable or a treatment variable, see among others, Aobdia (2018), Glaeser (2018), Ertimur, Rawson, Rogers, and Zechman (2018), and Chen, Zhang, and Zhou (2018).

using the true state of headquarters, we find that it is unrelated to all four outcomes, and that the bias imparted by measurement error is substantial—overstating the effect by 4x or more.

Importantly, we find these results are an artifact of the inclusion of firm fixed effects. Using our suggested diagnostic test, we find that the inclusion of firm and year fixed effects absorb over 99% of the variation in the state non-compete index. Consistent with the results of our simulations, we show that this extreme level of absorption magnifies measurement error bias. When firm fixed effects are *excluded*, we find (1) that the coefficients on the two non-compete indices are similar, suggesting measurement error bias is minimal, and (2) no evidence of a relation between either non-compete index and any of the four outcome variables.

Collectively, we interpret our results as suggesting that measurement error can bias in favor of falsely rejecting a true null hypothesis (i.e., a “false positive”); that regression models with high-dimensional fixed effects are especially prone to false positives; and that these concerns distort inferences in the literature. We encourage researchers to report results using a variety of measures as their variable of interest and using a variety of fixed effect structures. Triangulating across multiple measures and multiple fixed effect structures reduces the possibility that the totality of results is spurious. At the same time, we encourage reviewers and editors to be open to conflicting results across specifications and to allow authors to explain and examine such conflicts rather than suppress them or operate under the (mistaken) belief that the specification with more fixed effects is more robust. Specifications with more fixed effects are not necessarily “better” or more “robust.” If anything, our results suggest the opposite. When fixed effects absorb more than 90% of the variation in the variable of interest, even minimal measurement error can produce a large bias and lead to a false positive. Conflicting results should be transparently presented and discussed, and not hidden from the reader.

2. Theory and Simulations

In this section we provide a brief review of the theoretical effect of measurement error on coefficient bias. The effect of measurement error is briefly discussed in standard econometric texts including Greene (2003) and Angrist and Pischke (2009). Angrist and Pischke (2009) limit their discussion of measurement error to classical measurement error (i.e., white noise) and the attendant attenuation bias (p. 83-84). Greene (2003) notes that when variables in the regression model are correlated, then measurement error in a single independent variable – even when the error is uncorrelated with all other variables – can contaminate the coefficients on all variables (p. 84-86). Other work in the statistics literature considers circumstances in which measurement error can bias in favor of rejecting a true null hypothesis. For example, Shadish et al. (2002) explains that when there are correlations among three or more variables in the regression model (Y , X , and Z), the effect of measurement error in X does not always attenuate the coefficient on X . Similarly, Westfall and Yarkoni (2016) show that measurement error in multiple explanatory variables can lead to false rejection of a true null hypothesis, even when sample sizes are large and reliability is moderate. Perhaps most importantly, and relevant for this paper, Loken and Gelman (2017) demonstrate that if researchers are “hunting” for a coefficient that is statistically different from zero, measurement error can increase the likelihood of finding results (see also Burgstahler, 1987).

In the remainder of this section, we provide a brief review of the theoretical effects of measurement error on inferences using three distinct sets of simulations. We use these simulations to develop the intuition for how measurement error can bias in favor of rejecting a true null hypothesis (i.e., a “false positive”) and how fixed effects can exacerbate this bias. Throughout all simulations we employ a balanced panel of 100 cross-sectional units, or “firms”, and 30 temporal units, or “years.” For each simulation, we construct the DGP, estimate the regression specification,

and repeat the process 1,000 times, tabulating the distribution of regression coefficients. The goal of these simulations is to simplify the discussion of measurement error for an accounting audience.

2.1. *Classic Measurement Error in the Independent Variable of Interest*

We simulate the classic case of a linear regression with a single independent variable. Specifically, we simulate the following DGP:

$$Y = \beta_1 X \tag{1}$$

where $\beta_1 = 1$ and $X \sim N(0,1)$. Throughout all of our simulations, we assume the researcher perfectly observes Y but observes a noisy measure of X , which we denote $\hat{X} = X + \varepsilon_x$, where $\varepsilon_x \sim N(0, \sigma^2)$. Here, σ represents the amount of measurement error in the empirical proxy for the construct of interest. Larger values of σ correspond to more measurement error. We consider three different values of measurement error: $\sigma = 0$, $\sigma = 0.1$, and $\sigma = 0.5$. For each level of measurement error, we estimate the linear regression model in (1) using \hat{X} as the independent variable. We repeat this process 1,000 times and report the average coefficient (average $\hat{\beta}_1$) for each level of measurement error in Table 1.

The first row of Table 1 presents the benchmark specification without measurement error, and confirms our estimates are unbiased $\hat{\beta}_1 = \beta_1 = 1$. This serves as a specification and sanity check. The second and third rows of Table 1 add progressively more measurement error. As measurement error increases, Table 1 shows that $\hat{\beta}_1$ attenuates toward zero. The intuition for this result is that measurement error in \hat{X} does not affect the covariance between Y and \hat{X} , but does inflate the variance of \hat{X} . For example, the formula for $\hat{\beta}_1$ is given by:

$$\hat{\beta}_1 = \frac{cov(Y, \hat{X})}{var(\hat{X})} = \frac{cov(Y, X)}{var(X) + var(\varepsilon_x)} = \frac{1}{1 + \sigma^2} \tag{2}$$

which shows that measurement error affects the regression coefficient only through its effect on the denominator. In this simple setting, measurement error attenuates the coefficient of interest toward the null hypothesis (of zero) and biases against rejecting the null hypothesis.

2.2. *Correlation with a Control*

Next, we allow the measurement error in the variable of interest to be correlated with a control variable. We consider this scenario because it matches how measurement error is often encountered in practice, and mirrors the specific empirical setting we consider in our subsequent empirical tests (a setting that is present in over 100 published papers in top accounting and finance journals, see Appendix A).

We assume the true DGP, which is unknown to the researcher, is as follows:

$$Y = \beta_1 X + \beta_2 Z \tag{3}$$

where $\beta_1 = 0$, and $\beta_2 = 1$. We assume the variable of interest is X , and the researcher observes X with error, $\hat{X} = X + \varepsilon_x$. Unlike the preceding analysis, we allow the measurement error ε_x to be correlated with Z , i.e., $\varepsilon_x = \rho Z + \mu$. Here, ρ represents the amount of measurement error in \hat{X} that is correlated with the control variable. We assume the researcher observes a noisy measure of the control, $\hat{Z} = Z + \varepsilon_z$, and that all variables are drawn from i.i.d. standard normal distributions.⁹ We estimate the linear regression model in (3) using the observed values of the variables, and progressively add correlated measurement error ($\rho = 0, 0.1, 0.5$). We repeat this process 1,000 times and report the average coefficients in Table 2.

⁹ Note that inferences would be stronger if the control variable was omitted from the regression equation. In this case, because the control is correlated with measurement error in the included variable of interest, the omission of the control would create classic omitted variable bias.

The first row of Table 2 represents the benchmark case where $\rho = 0$. Here, $\hat{\beta}_1$ is unbiased (i.e., $\beta_1 = \hat{\beta}_1 = 0$), but $\hat{\beta}_2$ is attenuated.¹⁰ However the second and third rows illustrate that as ρ increases, the bias in $\hat{\beta}_1$ increases, pushing the estimate away from zero. The intuition for this result is that correlated measurement error introduces a correlation between \hat{X} and Z . Since Z has a causal effect on Y (and is measured with noise), the regression puts non-zero weight on \hat{X} . In other words, as the correlation between Z and measurement error in \hat{X} increases, the regression will place increasing weight on \hat{X} . Thus, when measurement error in the variable of interest is correlated with a control variable, measurement error does *not* necessarily attenuate the coefficient.

This example serves to illustrate a setting where measurement error can lead to false positives—a setting that features a control variable correlated with measurement error in the variable of interest. This setting is commonly encountered in practice and mirrors the empirical setting we consider in our subsequent empirical tests: measurement error in the location of corporate headquarters (variable of interest) is correlated with firm attributes that are included as noisy controls.

2.3. *Measurement Error and Fixed Effects*

Our next simulation examines how fixed effects interact with measurement error. In particular, we repeat the simulation analysis in Section 2.2 after adding a firm-specific component to the variable of interest. Specifically, we make the additional assumption that the variable of interest includes a firm-specific component, $X_{i,t} = x_{i,t} + \sum_{i=1}^{100} \theta_i Firm_i$, where $Firm_i$ is an indicator variable that equals one for i -th firm and zero, otherwise, θ_i is the coefficient on the i -th fixed effect, and $x_{i,t} \sim N(0,1)$ and $\theta_i \sim N(0,1)$. As a result, the variable of interest observed by the

¹⁰ Recall that there is measurement error in \hat{Z} , which attenuates the coefficient on \hat{Z} per the classic attenuation bias described in Section 2.1.

researcher has three components: (1) a within-firm component ($x_{i,t}$), (2) a firm-specific component ($\sum_{i=1}^{100} \theta_i Firm_i$), and (3) a measurement error component (i.e., ε_x).

Table 3 shows results from repeating the simulation from Section 2.2 after separately including and excluding firm effects. Column 1 and 2 (3 and 4) of Table 3 present the results *excluding (including)* fixed effects. Three results are notable:

(i) The first row of Table 3 represents the benchmark. Results are identical to Table 2: $\hat{\beta}_1$ is unbiased and $\hat{\beta}_2$ is attenuated.

(ii) Similar to Table 2, the second and third row of Table 3 show that as progressive amounts of measurement error are added, the bias in $\hat{\beta}_1$ increases, pushing the estimate away from zero.

(iii) Columns (1) and (3) indicate that bias is larger when fixed effects are *included*. For example, when $\rho = 0.1$, the coefficient on $\hat{\beta}_1 = 0.016$ when firm fixed effects are excluded and $\hat{\beta}_1 = 0.025$ when they are included. Thus, fixed effects amplify the coefficient bias by ~50%.

Figure 2 generalizes the results in Table 2 to multiple different parameterizations of the simulation and plots the bias in $\hat{\beta}_1$ for progressively higher levels of measurement error. Figure 2 shows that at every level of measurement error, fixed effects exacerbate the measurement error bias.

The intuition for why fixed effects exacerbate measurement error bias is that including fixed effects in the regression specification sucks out the “good variation” in the variable of interest (i.e., the variation not due to measurement error). As a result, measurement error comprises a greater percentage of the remaining (residual) variation in the variable of interest. To see this, note that the variable of interest has three components $\hat{X} = x_{i,t} + \sum_{i=1}^{100} \theta_i Firm_i + \varepsilon_x$. When fixed effects are excluded from the regression, the proportion of \hat{X} driven by measurement error is:

$$\frac{\text{var}(\varepsilon_x)}{\text{var}(\hat{X})} = \frac{\text{var}(\varepsilon_x)}{\text{var}(x_{i,t} + \sum_{i=1}^{100} \theta_i \text{Firm}_i + \varepsilon_x)} = \frac{\text{var}(\varepsilon_x)}{2 + \text{var}(\varepsilon_x)} \quad (4)$$

In contrast, when fixed effects are included in the regression, and thus partialled out of \hat{X} , the proportion of the remaining variation driven by measurement error is:

$$\frac{\text{var}(\varepsilon_x)}{\text{var}(\hat{X} - \sum_{i=1}^{100} \theta_i \text{Firm}_i)} = \frac{\text{var}(\varepsilon_x)}{1 + \text{var}(\varepsilon_x)} \quad (5)$$

Note that (5) is larger than (4).. For example, when $\text{var}(\varepsilon_x) = 1$, and fixed effects are excluded, measurement error accounts for 33% of the variation in the variable of interest (one of three components). However, when fixed effects are included, measurement error accounts for 50% of the variation (one of two remaining components). Hence, the inclusion of fixed effects increases the relative amount of measurement error in the independent variable.

Collectively, we interpret these results as indicating that the common practice of including high-dimensional fixed effects can “throw the baby out with the bath water” which will make the regression specifications more susceptible to the measurement error biases we document—and exacerbate the likelihood of false positives. While the inclusion of fixed effects may initially seem sensible, they can nonetheless lead the researcher to spuriously estimate a causal effect when none exists. We echo the sentiments of Angrist and Pischke (2009): “At a minimum, therefore, it’s important to avoid overly strong claims when interpreting fixed effects estimates.”¹¹

2.3.1. Fixed Effect Absorption

¹¹ For those readers interested in a more detailed treatment of the various strengths and weaknesses of fixed effects see the following literature: (i) Angrist and Pischke (2009) discuss issues when both fixed effects and lag values of the dependent variable are included in the same model; (ii) Grieser and Hadlock (2019) discuss the strict exogeneity assumption of fixed effects; (iii) Berg et al. (2021) discusses how fixed effects exacerbate coefficient bias when seeking to estimate spillover effects, (iv) Whited et al. (2021) discuss fixed effects in the more general context of the “bad controls” problem discussed in Angrist and Pischke (2009), and (v) deHaan (2020) who reviews the use of fixed effects in accounting research

An interesting question that arises in this context is whether the structure of the fixed effects matter (i.e., number of observations per effect, number of effects, nested effects). That is, researchers might be trying to decide between including year effects, industry effects, firm effects, manager effects, or all four effects simultaneously (e.g., Armstrong et al. 2019). In our context, the structure of the fixed effects *per se* does not matter—what matters is the amount of variation that they absorb in the independent variable of interest (i.e., $\frac{\text{var}(\sum_{i=1}^{100} \theta_i \text{Firm}_i)}{\text{var}(X)}$). In our earlier simulation in Section 2.3 we implicitly assumed an absorption rate of 50% (i.e., 50% of the variation in X was driven by fixed effects). Empirically, however the absorption rate likely varies by setting. In some settings, fixed effects might absorb only 10% of the variation in the variable of interest, but in others they might account for upwards of 99% of the variation in the variable of interest (e.g., Armstrong et al., 2021; Donelson et al., 2022). Indeed, in Section 3, we find the common practice of including both firm and year effects absorbs upwards of 99.4% of the variation in the independent variable of interest, the non-compete index.

To examine how measurement error bias increases with the absorption rate, we make two changes to the simulation in Section 2.3. First, we assume $x_{i,t}$ is distributed $N(0, 2 - a)$. Second, we assume θ_i is distributed $N(0, a)$. As a result, the absorption rate is given by the expression $\frac{\text{var}(\sum_{i=1}^{100} \theta_i \text{Firm}_i)}{\text{var}(X)} = \frac{a}{2}$, where $0 < a < 2$. Thus, high levels of a indicate greater levels of absorption, with $\frac{a}{2}$ representing the fraction of variation absorbed.

Table 4 reports results from repeating the simulation in Section 2.3 for absorption rates of 10%, 50%, and 90% (which comports to $a = 0.2, 1$, and 1.8 , respectively). Moving across the three columns, we see that for every non-zero level of measurement error, the bias in $\hat{\beta}_1$ increases with the amount of variation absorbed by the fixed effects (recall that true $\beta_1 = 0$). Figure 3 plots the

bias, and illustrates how, *ceteris paribus*, greater fixed effect absorption implies greater measurement error bias. This shows that measurement error bias increases as fixed effects absorb more variation in the independent variable.

The intuition for this result generalizes the intuition from Section 2.2. As the absorption rate increases, fixed effects remove increasing amounts of variation from the variable of interest, leaving proportionately more measurement error. To see this, note that we can rewrite the expression for the amount of measurement error in the variable of interest in the presence of fixed effects (i.e., eqn (5)) as:

$$\frac{\text{var}(\varepsilon_X)}{\text{var}(\hat{X} - \sum_{i=1}^{100} \theta_i \text{Firm}_i)} = \frac{\text{var}(\varepsilon_X)}{2 - a + \text{var}(\varepsilon_X)}. \quad (6)$$

As a asymptotically approaches zero (i.e., fixed effects absorb 0% of the variation in X), the expression converges to that given by equation (4). Fixed effects will not exacerbate measurement error bias. However, as a asymptotically approaches two (i.e., fixed effects absorb 100% of the variation in X) the expression converges to 1. In this circumstance, measurement error will (asymptotically) account for 100% of the remaining variation in \hat{X} leading to extreme levels of bias.

2.3.2. False Positives

Thus far, our analysis has focused on bias in the estimated coefficient. Next, we examine how the bias would affect inferences regarding statistical significance. In our setting, measurement error bias should increase the rate of false positives—where a “false positive” is a researcher rejecting the null hypothesis that the coefficient is zero, when the null is in fact true. In the absence of bias, the rate of false positives is determined by the level of statistical significance used in the tests. For example, at a 5% two-tailed level of statistical significance, only 5% of results should be false positives.

To examine how measurement error bias affects the rate of false positives, we tabulate the fraction of times in the 1,000 iterations of the simulation in Table 4, that we reject the (true) null hypothesis that $\beta_1 = 0$. These fractions represent the percentage of the simulations where the bias in $\hat{\beta}_1$ was so large that it falls outside the 95% confidence interval. Because the bias is unbounded, but false positive rates are bounded at 100%, above a threshold level of bias (i.e., a certain level of measurement error and absorption) the false positive rate will plateau at 100%.

Table 5 presents false positive rates, using a two-tailed 5% level of statistical significance. The first row of Table 5 presents the benchmark and confirms our tests are correctly specified. Table 5 shows that for every non-zero level of measurement error ($\rho > 0$), either an increase in the absorption rate or an increase in measurement error increases the rate of false positives. The results are striking. At a modest level of measurement error ($\rho = 0.1$), measurement error bias increases the false positive rate from 5% to 60%, and the inclusion of fixed effects that absorb 90% of the variation in the variable of interest further magnifies the false positive rate to 94%. Figure 4 graphically illustrates these results for various levels of measurement error and fixed effect absorption. We emphasize: the results suggest that with extreme levels of fixed effect absorption and extreme levels of measurement error, researchers are all but guaranteed to falsely reject a true null hypothesis (i.e., false rejection rates = 100%)!

These findings are concerning and suggest the possibility that many of the results in the literature could be spurious. The combination of measurement error and fixed effects with high absorption levels *could* be driving many of the results in the literature, giving rise to a false impression of a causal effect when—in truth—no such effect exists (Ohlson, 2021).¹² In the next

¹² See Donelson et al. (2022) for an application of this point in the context of universal demand laws.

section, we conclusively demonstrate this concern is real. We replicate prior work and show that measurement error in the state of corporate headquarters combined with high-dimensional fixed effects that absorb over 99% of the variation in the variable of interest, gives the (false) impression that state non-compete laws cause voluntary disclosure and earnings management.

2.3.3. Diagnostic Tests

The preceding analysis suggests that the combination of measurement error and high levels of fixed absorption can generate false positives. Fortunately, the results also suggest a simple diagnostic tool researchers can use to assess the amount of variation in the independent variable of interest that is absorbed by a particular set of fixed effects: the R-squared from a regression of the independent variable of interest on the fixed effect structure. The R-squared from this regression is an estimate of the absorption rate. It indicates the fraction of variation in the independent variable that is removed, or absorbed, by the fixed effects when they are included as a control variable. While we cannot offer a definitive answer on what absorption rate is empirically acceptable, we suggest extreme caution in circumstances when fixed effects absorb more than 90% of the variation in the independent variable of interest.

More to the point, we encourage researchers to report results across a variety of fixed effect structures, including specifications that exclude fixed effects altogether. The common practice of adding more and more dimensions of fixed effects is not a panacea and can actually generate false positives. Thus, more fixed effects are not necessarily “better”, and we caution against the notion that more dimensions of fixed effects means the regression specification is somehow more robust. If anything, our results suggest the opposite. When fixed effects absorb more than 90% of the variation in the variable of interest, even minimal measurement error can produce a large bias and lead to a false positive.

3. Empirical Analysis

In this section, we provide a practical illustration of how measurement error can affect inferences in the literature. To do so, we examine a popular quasi-natural experimental setting where we know the measurement error in the treatment variable. Specifically, we use data on the state of corporate headquarters from Compustat to estimate the “causal effect” of state non-compete laws on voluntary disclosure and earnings management. We then benchmark these results against those obtained from using the true state of corporate headquarters listed on the firm’s annual 10-K filing.

3.1 Data on the Location of Corporate Headquarters

A large literature exploits variation in state of corporate headquarters to estimate the causal effects of state regulations and taxes. Appendix A catalogs more than 100+ papers in top journals across accounting and finance that use the state of corporate headquarters in their analysis. Of these papers, approximately 80% obtain data on the state of corporate headquarters from Compustat. However, Compustat backfills all observations with the current location of the firm’s headquarters. For example, if a firm moves its headquarters from Texas to California in 2019, Compustat codes California as the state of headquarters in 2019 and *in all prior years* (e.g., Heider and Ljungqvist, 2015; Jennings, Lee, and Matsumoto, 2017). Thus, this setting provides a unique opportunity to study the effect of measurement error on inferences in the literature, because we can directly observe the firm’s true state of headquarters listed on the front page of their annual 10-K filing. Thus, we can compare inferences based on the Compustat headquarter location data, which contains error, to inferences based on the true headquarter location listed on the annual 10-K filing.

The sample used in our analysis spans 1996 to 2013, the period for which we have data on state non-compete laws. To be included in the sample, we require the firm appears on CRSP/Compustat and has non-missing market value of equity, total assets, total liabilities, long-term debt, and income before extraordinary items. The resulting sample consists of 73,178 firm-years. Column (1) of Table 6 shows the number of firms each year in the sample. Column (2) reports the fraction of observations where the state of headquarters listed on the 2019 Compustat file differs from that on the firm's 10-K filing.

Table 6 shows that the error rates in Compustat headquarter data routinely exceed 10% per year. In addition, consistent with significant backfilling, Table 6 shows that the error rates monotonically decline from 18.9% in 1996 to 4.8% in 2013. The notion that more than 80 papers rely on Compustat as a source of corporate headquarters data when such data are prone to extensive errors suggests measurement error is a pervasive issue in the literature. Beyond Compustat and the 10-K, Appendix A suggests there are two other sources of data on headquarter location in the literature: Heider and Ljungqvist (2015) and the SEC header file. Because these data sources are used relatively infrequently (<5% of papers), and occasionally misapplied, we discuss them and their (mis)application in Appendix B.

3.2. *State Non-Compete Laws*

Having documented substantial error in Compustat headquarter data, we next examine the effect of this error on inferences as it relates to state non-compete laws. Following prior literature, we use the state-level index of non-compete laws developed in Garmaise (2011) and extended in Ertimur, Rawson, Rogers, and Zechman (2018). This index ranges from 0 to 9, with 0 (9) representing low (high) enforceability and varies over time and across states, which facilitates the sort of staggered adoption design commonly seen in quasi-experimental settings. As is customary

in the literature, we assign the value of the index to firm-years based on their state of headquarters. If the state of headquarters is measured with error, then the corresponding value of the non-compete index will likely also be measured with error. We calculate two non-compete indices, *NonCompete*[^] is the non-compete index calculated using the Compustat headquarters location, and *NonCompete*^{*} is the non-compete index calculated using the true headquarters location from the firm's 10-K.

We estimate the relation between these two non-compete indices and measures of voluntary disclosure and earnings management. We measure voluntary disclosure using the number of management forecasts issued during the year, *Forecast*, the number of 8Ks filed during the year, *8KFiling*, and the number of firm-initiated press releases during the year, *PressRelease*. We measure earnings management using an indicator variable equal to one if financial results for that year were subsequently restated as a result of fraud, misrepresentation, or an SEC investigation, *Restate* (e.g., Hennes, Leone, and Miller, 2008; Armstrong, Larcker, Ormazabal, and Taylor, 2013).¹³ Data on management forecasts (8K filings) comes from IBES (EDGAR) and spans 1996-2013, data on press releases comes from RavenPack and spans 2004-2013, and data on restatements comes from Audit Analytics and spans 2000-2013.

Table 7 provides descriptive statistics for the variables used in our analysis. We find that on average firms tend to have three management forecasts per year (mean *Forecast* = 2.97), eight 8-K filings during the year (mean *8KFiling* = 7.59), and nineteen press releases during the year

¹³ We selected these measures from among a variety of others, because they are simple, standard, and entail fewer “researcher degrees of freedom” (Loken and Gelman, 2017). For example, there are a myriad of accruals models we could choose from, but it is well known that inferences in the literature are often sensitive to the choice of accrual model (e.g., Armstrong et al., 2021).

(mean *PressRelease* = 18.52). We also find that 2% of firms restate their financials due to misrepresentation (mean *Restate* = 0.02).

In Table 8, we compute the measurement error in the non-compete index as a result of using the wrong state of headquarters ($Error = NonCompete^{\wedge} - NonCompete^*$), and examine its correlation with common control variables. Correlation between measurement error in the variable of interest, and control variables is the scenario we simulated in Sections 2.2. and 2.3. Table 8 indicates that the measurement error in our variable of interest is negatively correlated with *Size*, *Roa*, and *NumEst* and positively correlated with *Mtb*. All variables are defined in Table 7.

To estimate the effect of measurement error on inferences, we formally test for a relation between state non-compete laws and our four outcome variables by estimating the following regression specification:

$$Outcome_{i,t} = \beta NonCompete_{i,t} + \gamma Controls_{i,t} + \phi Firm_i + \omega Year_t + \varepsilon_{i,t} \quad (7)$$

where *Outcome* is either *Forecast*, *8KFiling*, *PressRelease*, or *Restate*; *NonCompete* is either *NonCompete[^]* or *NonCompete^{*}*; *Controls* is a vector of control variables commonly used in prior research, including firm size (*Size*), the market-to-book ratio (*Mtb*), leverage (*Leverage*), operating performance (*Roa*), and analyst and institutional investor coverage (*NumEst* and *NumInst*, respectively); *Firm* is a vector of firm fixed effects; and *Year* is a vector of year fixed effects. Based on prior research that non-compete laws are associated with lower levels of voluntary disclosure and more earnings management (e.g., Aobdia, 2018; Chen, Zhang, Zhou, 2018) we predict $\beta < 0$ when the dependent variable is either *Forecast*, *8KFiling*, or *PressRelease*, and $\beta >$

0 when the dependent variable is *Restate*. Throughout our analysis we estimate linear specifications and cluster standard errors by firm.¹⁴

Table 9 presents results from estimating equation (7). Consistent with the predictions of prior work, columns (1), (3), (5), and (7) show that the non-compete index computed using Compustat headquarter data is negatively related to management forecasts, 8-K filings, and press releases (*t*-stats of -2.82 , -2.19 , -2.02) and positively related to restatements (*t*-stat of 2.02). In addition, the coefficients imply an economically significant effect of non-compete laws on these outcome variables. For example, a one standard deviation increase in *NonCompete*[^] (std. dev of 2.31) increases the probability of a restatement by 2.54% , where the unconditional probability of restatement is 2% . In stark contrast to these results, columns (2), (4), (6), and (8) show that the non-compete index is unrelated to all four outcomes when it is calculated using the true state of headquarters (*t*-stats of -0.44 , -0.22 , 0.03 , and 1.01). For each dependent variable, we calculate the magnitude of the bias as the absolute value of the difference between the coefficients on *NonCompete*[^] and *NonCompete*^{*}. In this setting, the magnitude of the bias imparted by measurement error is substantial—over 4x the true effect.¹⁵

These findings illustrate that had the researcher used the state of headquarters listed on Compustat, they would have documented a significant “causal effect” of non-compete laws on management forecasts, 8-K filings, press releases, and restatements—potentially leading to a publication in a top journal. Alternatively, had the researcher used the true state of headquarters,

¹⁴ We do not use the logistics/probit regression due to the “incidental parameters problem” associated with fixed effects. Specifically, Arellano and Hahn (2007) suggest that nonlinear maximum likelihood models can be inconsistent and biased when fixed effects are included. Avoiding maximum likelihood models allows us to pin the bias introduced by the fixed effects to measurement error rather than the “incidental parameters problem.” Using linear models with discrete outcome variables is common practice in accounting (e.g., Guay et al., 2016).

¹⁵ For *Restate*, $|0.011 - 0.002| / 0.002 = 4.5$.

they would have found no evidence of an effect and the study would likely not have been published. This example nicely illustrates how publication incentives produce a subtle and unintentional bias in the literature toward noisy measures. See Loken and Gelman (2017) for a greater development and discussion of this point.

3.3. *State Non-Compete Laws and Absorption Rates*

In estimating equation (7), we followed standard practice in the literature and included firm and year fixed effects (e.g., deHaan, 2020). However, the simulation evidence in Section 2.4 suggests firm fixed effects can exacerbate measurement error bias and lead researchers to spuriously estimate a causal effect when none exists. This opens the door to the possibility that it is the combination of both (i) measurement error and (ii) firm effects that caused the “false positive” in Table 9. To investigate this possibility, we repeat the analysis in Table 9 after substituting a variety of fixed effect structures. We consider four structures: year effects, industry and year effects, industry×year effects, and firm and year effects. For each structure, we also compute and report the fixed effect absorption rate, i.e., the R-squared from a regression of $NonCompete^i$ on the specific fixed effect structure. We report the results in Table 10.

In Panel A of Table 10, we report the results using the four fixed effects structures with *Forecast* as the dependent variable. The column headings of Panel A indicate the absorption rate for year effects is 0.05%; industry and year fixed effects is 3.62%; industry×year fixed effects is 3.72%; and firm and year effects is 99.49%. The notion that firm and year effects absorb 99.49% of the variation in the variable of interest is striking. This is an even higher level of absorption than contemplated in our simulations.

Notably Panel A shows that it is only when firm and year fixed effects are included that the coefficient on $NonCompete^i$ becomes significant at the 1% level (Column (7)). There is no

evidence of a relation between *NonCompete*[^] and the dependent variable when other fixed effect structures are used. Panel B through D show similar results for *8KFiling*, *PressRelease*, and *Restate* as the dependent variables. Across all specifications, we find no evidence of a relation between mismeasured state non-compete laws and any of our outcome variables when only year fixed effects are included, industry and year fixed effect are included, or industry×year fixed effects are included. It is only when the absorption rate significantly increases (i.e., when firm and year fixed effects are included) that we find a relation between mismeasured non-compete laws and all four outcome variables. Thus, the evidence in Table 10 shows it is the *combination* of measurement error and high-levels of fixed absorption that cause the “false positives” in this setting.

4. Conclusion

We show theoretically and empirically that measurement error can bias in favor of falsely rejecting a true null hypothesis (i.e., a “false positive”), and that regression models with high-dimensional fixed effects are especially prone to false positives. Collectively, we interpret the results as suggesting that (i) measurement error does not always bias against finding results and can lead researchers to spuriously estimate a causal effect when none exists, (ii) that the inclusion of high-dimensional fixed effects can exacerbate measurement error bias and increase the likelihood of a false positive, and (iii) that measurement error pervades the accounting literature and in practice can produce the appearance of large causal effects when none exist (Ohlson, 2021).

We provide researchers with a simple diagnostic tool to assess the possibility that the combination of measurement error and high-dimensional fixed effects might give rise to a false positive: the R-squared from a regression of the independent variable of interest on the fixed effect

structure. This statistic measures the fraction of variation in the independent variable that is removed, or absorbed, by the fixed effects when they are included as control variables. While we cannot offer a definitive answer on what absorption rate is empirically acceptable, we suggest extreme caution in circumstances when the variable of interest is measured with error, and fixed effects absorb more than 90% of the variation in the independent variable of interest.

Finally, we encourage researchers to report results using a variety of measures as their variable of interest and using a variety of fixed effect structures. Triangulating across multiple measures and multiple fixed effect structures reduce the possibility that the totality of results is spurious. At the same time, we encourage reviewers and editors to be open to conflicting results across specifications and to allow authors to explain and examine such conflicts rather than suppress them or operate under the (mistaken) belief that the specification with more fixed effects is more robust. Specifications with more fixed effects are not necessarily “better” or more “robust.” If anything, our results suggest the opposite. When fixed effects absorb more than 90% of the variation in the variable of interest, even minimal measurement error can produce a large bias and lead to a false positive.

References

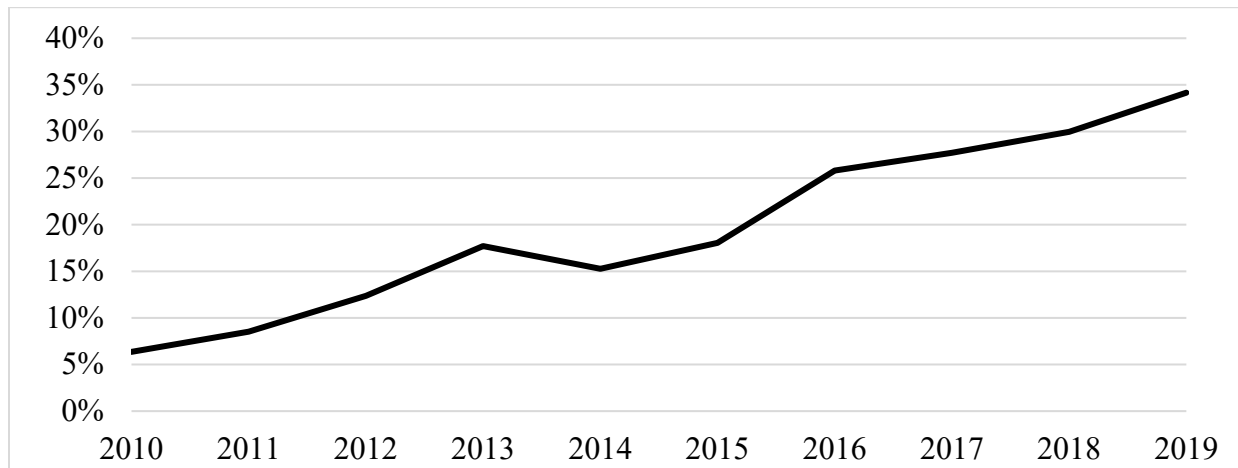
- Angrist, J. D., & Pischke, J. 2009. Mostly harmless econometrics : An empiricist's companion. Princeton: Princeton University Press.
- Aobdia, D. 2018. Employee mobility, noncompete agreements, product-market competition, and company disclosure. *Review of Accounting Studies*, 23(1), 296-346.
- Arellano, M. & Hahn, J., 2007. Understanding bias in nonlinear panel models: some recent developments. *Econometric Society Monographs*, 43, 381.
- Armstrong, C. S., Glaeser, S., Huang, S. & Taylor, D.J. 20,9. The economics of managerial taxes and corporate risk-taking. *The Accounting Review*, 94(1), 1-24.
- Armstrong, C. S., Larcker, D. F., Ormazabal, G., & Taylor, D. J. 2013. The relation between equity incentives and misreporting: The role of risk-taking incentives. *Journal of Financial Economics*, 109(2), 327-350.
- Armstrong, C. S., Kepler, J. D., Taylor, D. J., & Samuels, D. 2021. The evolution of empirical methods in accounting research and the growth of quasi-experiments. Working Paper.
- Barth, M. E. 1991. Relative measurement errors among alternative pension asset and liability measures. *The Accounting Review*, 66(3), 433-463.
- Berg, R., Reisinger, M., & Streitz, D., 2021. Spillover effects in empirical corporate finance. *Journal of Financial Economics*, forthcoming.
- Berkman, H., & Truong, C. 2009. Event day 0? After-hours earnings announcements. *Journal of Accounting Research*, 47(1), 71-103.
- Bertomeu, J., Beyer, A., & Taylor, D. J. 2016. From casual to causal inference in accounting research: The need for theoretical foundations. *Foundations and Trends (R) in Accounting*, 10(2-4), 262-313.
- Bonsall IV, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. 2017. A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2-3), 329-357.
- Burgstahler, D. 1987. Inference from empirical research. *The Accounting Review*, 62(1), 203-214.
- Chen, T. Y., Zhang, G., & Zhou, Y. 2018. Enforceability of non-compete covenants, discretionary investments, and financial reporting practices: Evidence from a natural experiment. *Journal of Accounting and Economics*, 65(1), 41-60.
- Christie, A. A. 1987. On cross-sectional analysis in accounting research. *Journal of Accounting and Economics*, 9(3), 231-258.
- Collins, D.W., Kothari, S.P., Shanken, J. & Sloan, R.G. 1994. Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association. *Journal of Accounting and Economics*, 18(3), 289-324.
- Correia, S. 2015. Singletons, cluster-robust standard errors and fixed effects: A bad mix. Working Paper.
- deHaan, E. 2020. Practical guidance on using and interpreting fixed effects models. Working Paper.
- Donelson, D. C., Kettell, L., McInnis, J., & Toynbee, S. 2022. The need to validate exogenous shocks: Shareholder derivative litigation, universal demand laws and firm behavior. *Journal of Accounting and Economics*, 73(1), 101427.

- Ertimur, Y., Rawson, C., Rogers, J.L. & Zechman, S.L. 2018. Bridging the gap: Evidence from externally hired CEOs. *Journal of Accounting Research*, 56(2), 521-579.
- Garmaise, M.J., 2011. Ties that truly bind: Noncompetition agreements, executive compensation, and firm investment. *The Journal of Law, Economics, and Organization*, 27(2), 376-425.
- Glaeser, S. 2018. The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. *Journal of Accounting and Economics*, 66(1), 163-193.
- Gow, I.D., Larcker, D.F. & Reiss, P.C. 2016. Causal inference in accounting research. *Journal of Accounting Research*, 54(2), 477-523.
- Greene, W. H. 2003. *Econometric analysis*. New Jersey: Pearson Education, Inc.
- Grieser, W. & Hadlock, C., 2019. Panel-data estimation in finance: testable assumptions and parameter (in) consistency. *Journal of Financial and Quantitative Analysis*, 54(1), 1-29.
- Griliches, Z., & Hausman, J. A. 1986. Errors in variables in panel data. *Journal of Econometrics*, 31(1), 93-118.
- Guay, W., Samuels, D. & Taylor, D., 2016. Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics*, 62(2-3), 234-269.
- Hanlon, M., Verdi, R., & Yost, B. 2020. CEO tax effects on acquisition structure and value. *The Accounting Review*, forthcoming.
- Heider, F., & Ljungqvist, A. 2015. As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes. *Journal of Financial Economics*, 118(3), 684-712.
- Hennes, K. M., Leone, A. J., & Miller, B. P. 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review*, 83(6), 1487-1519.
- Jennings, J., Lee, J., & Matsumoto, D. A. 2017. The effect of industry co-location on analysts' information acquisition costs. *The Accounting Review*, 92(6), 103-127.
- Khan, M., Serafeim, G., & Yoon, A. 2016. Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6), 1697-1724.
- Kothari, S.P. 2001. Capital markets research in accounting. *Journal of Accounting and Economics*, 31(1-3), 105-231.
- Lai, S., Li, Z., & Yang, Y. 2020. East, West, home's best: Do local CEOs behave less myopically. *The Accounting Review*, 95(2), 227-255.
- Larcker, D. F., Richardson, S. A., & Tuna, I. R. 2007. Corporate governance, accounting outcomes, and organizational performance. *The Accounting Review*, 82(4), 963-1008.
- Loken, E., & Gelman, A. 2017. Measurement error and the replication crisis. *Science*, 355(6325), 584-585.
- Loughran, T., & McDonald, B. 2014. Measuring readability in financial disclosures. *The Journal of Finance*, 69(4), 1643-1671.
- Ohlson, J. A. 2021. Researchers' data analysis choices: an excess of false positives?. *Review of Accounting Studies*.
- Payne, J. L., & Thomas, W. B. 2003. The implications of using stock-split adjusted I/B/E/S data in empirical research. *The Accounting Review*, 78(4), 1049-1067.
- Shadish, W., Cook, T. D., & Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

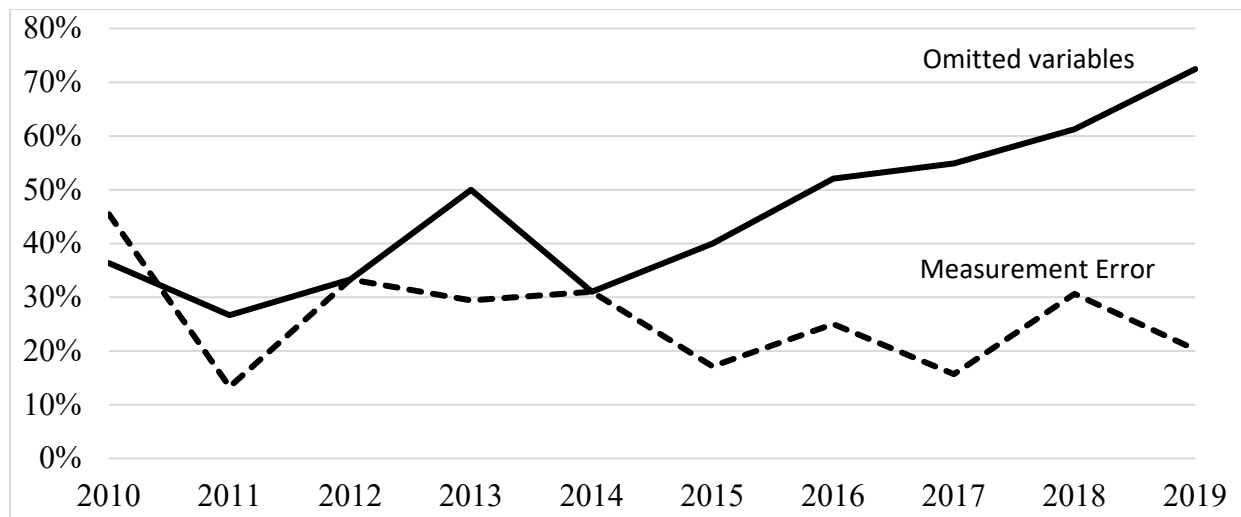
- Westfall, J., & Yarkoni, T. 2016. Statistically controlling for confounding constructs is harder than you think. *PloS ONE*, 11(3), 1-22.
- Whited, R., Swanquist, Q., Shipman, J. & Moon, J., 2021. Out of control: The (over) use of controls in accounting research. *The Accounting Review*, forthcoming.

Figure 1. Discussions of Measurement Error over Time

Panel A. Discussions of Quasi-Natural Experimental Designs in Published Papers

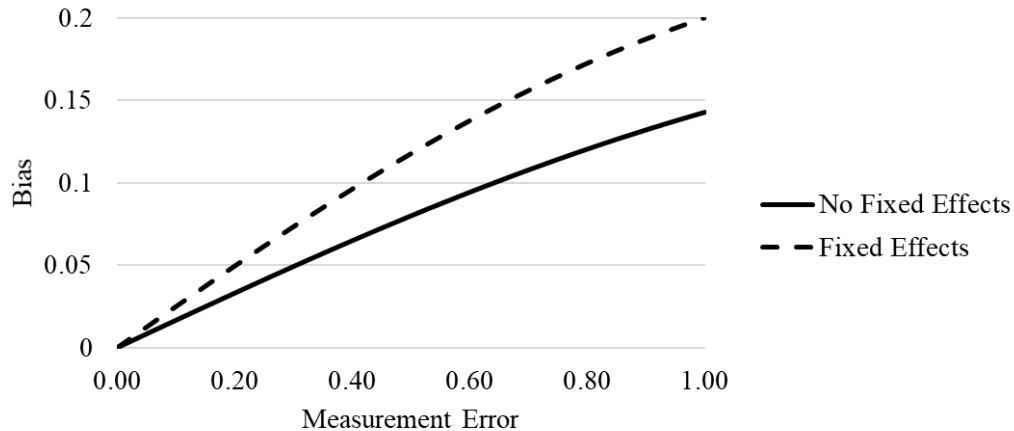


Panel B. Percentage of Papers with Quasi-Natural Experimental Design that Discuss Omitted Variables and Measurement Error



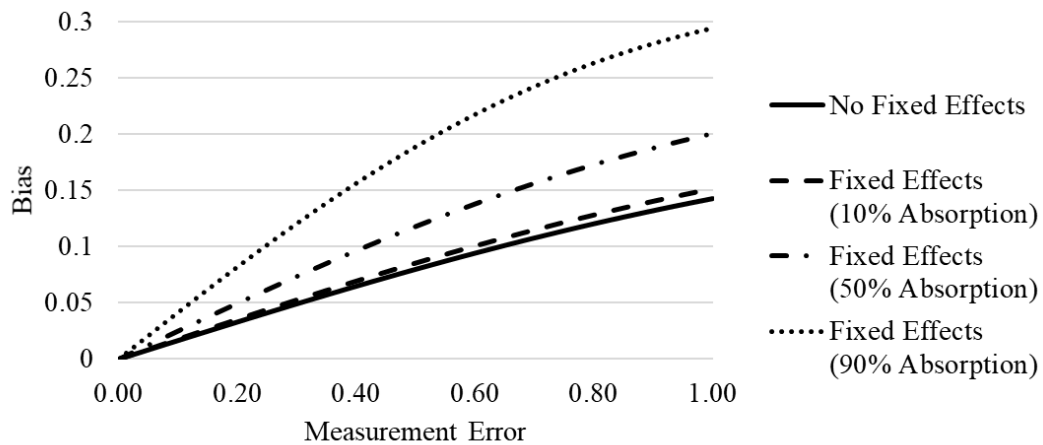
Panel A reports the percentage of papers published in the *Journal of Accounting and Economics*, *Journal of Accounting Research*, *The Accounting Review*, and *Review of Accounting Studies* that mention keywords related to quasi-natural experimental designs (i.e., “natural experiment”, “difference in differences”, “staggered adoption”). Within the set of papers that mention quasi-natural experimental designs keywords, Panel B reports the percentage of papers that also mention keywords related to omitted variables (i.e., “parallel trends”, “correlated omitted”, “omitted correlated”, “omitted variable”, and “omitted factor”) and measurement error (i.e., “measurement error”).

Figure 2. Measurement Error and Fixed Effects



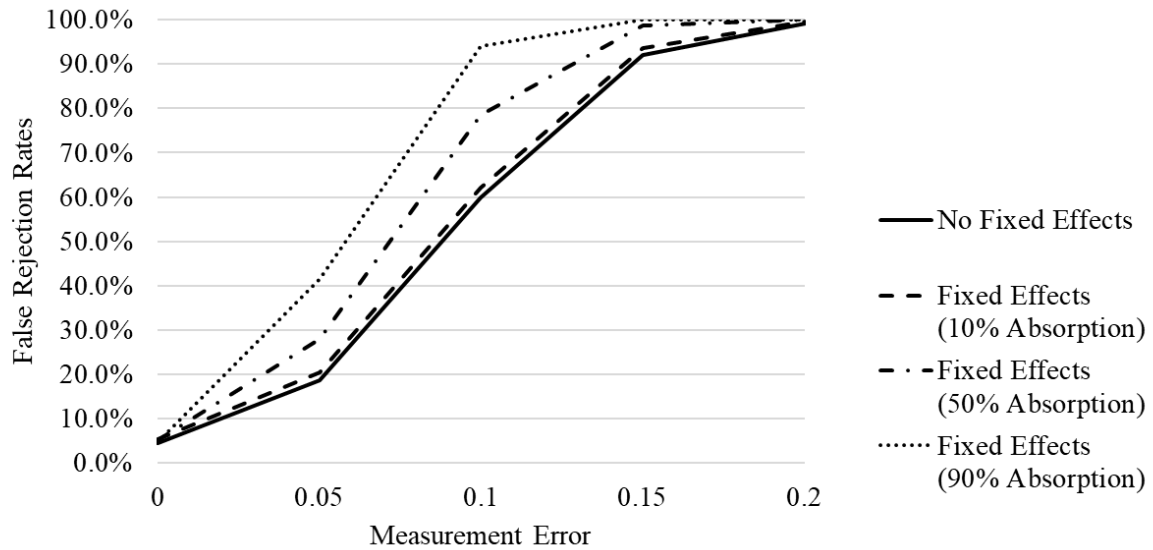
This figure plots the bias in the coefficient of interest ($\hat{\beta}_1$) for various levels of measurement error when firm fixed effects are excluded (included). Simulations follow Table 3.

Figure 3. Measurement Error and Fixed Effects Absorption Rates



This figure plots the bias in the coefficient of interest ($\hat{\beta}_1$) for various levels of measurement error when firm fixed effects are included, and absorb increasing amounts of variation in the variable of interest. Simulations follow Table 4.

Figure 4. Measurement Error, Fixed Effects, and False Positives



This figure plots the false rejection rate of a true null that $\hat{\beta}_1 = 0$ for various levels of measurement error when firm fixed effects are included, and absorb increasing amounts of variation in the variable of interest. Simulations follow Table 4.

Table 1. Classic Measurement Error

Measurement Error	Average $\hat{\beta}_1$
0.0	1.000
0.1	0.990
0.5	0.800

This table presents results from our first simulation, where measurement error exists in the independent variable. We assume the data generating process is $Y = \beta_1 X$, where $\beta_1 = 1$. We assume the researcher perfectly observes Y but observes a noisy measure of X , which we denote $\hat{X} = X + \varepsilon$. We assume variables X and ε are i.i.d., where $X \sim N(0,1)$ and $\varepsilon \sim N(0, \sigma^2)$. We estimate the coefficient $\hat{\beta}_1$, from a regression of Y on \hat{X} for each of three levels of measurement error ($\sigma = 0, 0.1$, or 0.5). For each level of measurement error, we simulate the distribution of coefficients using a balanced panel of 100 firms and 30 years. We repeat the simulation 1000 times, and report the mean values of the resulting distribution of 1000 $\hat{\beta}$ s for each level of measurement error.

Table 2. Correlated Measurement Error

Measurement Error	Average $\hat{\beta}_1$	Average $\hat{\beta}_2$
0.0	0.000	0.500
0.1	0.025	0.499
0.5	0.118	0.471

This table presents results from our second simulation, where measurement error exists in the independent variable of interest and the measurement error is correlated with an included control variable. We assume the data generating process is $Y = \beta_1 X + \beta_2 Z$, where $\beta_1 = 0$, $\beta_2 = 1$, and the coefficient of interest to the researcher is β_1 . We assume the researcher perfectly observes Y ; the researcher observes a noisy measure of the variable of interest ($\hat{X} = X + \varepsilon_x$); the researcher observes a noisy measure of the control variable ($\hat{Z} = Z + \varepsilon_z$); and ε_x is correlated with the control variable (i.e., $\varepsilon_x = \rho Z + \mu$), where all variables are drawn from i.i.d. standard normal distributions. We estimate the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ from a regression of Y on \hat{X} and \hat{Z} , varying ρ from 0.0 to 0.5. For each specification, we simulate the distribution of coefficients using a balanced panel of 100 firms and 30 years. We repeat the simulation 1000 times, and report the mean values of the resulting distribution of $\hat{\beta}$ s for each specification.

Table 3. Measurement Error and Fixed Effects

	Fixed Effects <i>Excluded</i>		Fixed Effects <i>Included</i>	
	(1)	(2)	(3)	(4)
	Average	Average	Average	Average
Measurement Error	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.0	0.000	0.500	0.000	0.500
0.1	0.016	0.499	0.025	0.499
0.5	0.080	0.480	0.117	0.471

This table presents simulation results from repeating the analysis in Table 2, assuming the variable of interest entails a firm-specific component. Specifically, the variable of interest is given by, $X_{i,t} = x_{i,t} + \sum_{i=1}^{100} \theta_i Firm_i$, where i indexes firms, t indexes years, and $Firm_i$ is the i -th firm fixed effect, and θ_i is the coefficient on the i -th fixed effect. As before, we assume the true data generating process unknown to the researcher, is given by $Y_{i,t} = \beta_1 X_{i,t} + \beta_2 Z_{i,t}$, where $\beta_1 = 0$, $\beta_2 = 1$, and the coefficient of interest to the researcher is β_1 . We assume the researcher perfectly observes Y ; the researcher observes a noisy measure of the variable of interest ($\widehat{X}_{i,t} = X_{i,t} + \varepsilon_x$); the researcher observes a noisy measure of the control variable ($\widehat{Z} = Z + \varepsilon_z$); and the measurement error in the variable of interest is correlated with the control variable (i.e., $\varepsilon_x = \rho Z + \mu$), where all variables are drawn from i.i.d. standard normal distributions. We estimate the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ from a regression of Y on \widehat{X} and \widehat{Z} , varying ρ from 0.0 to 0.5 with and without fixed effects. For each specification, we simulate the distribution of coefficients using a balanced panel of 100 firms and 30 years. We repeat the simulation 1000 times, and report the mean values of the resulting distribution of $\hat{\beta}$ s for each specification.

Table 4. Measurement Error and Fixed Effect Absorption Rates

	Fixed Effects <i>Included</i> Absorption = 10%	Fixed Effects <i>Included</i> Absorption = 50%	Fixed Effects <i>Included</i> Absorption = 90%
Measurement Error	Average $\hat{\beta}_1$	Average $\hat{\beta}_1$	Average $\hat{\beta}_1$
0.0	0.000	0.000	0.000
0.1	0.018	0.025	0.041
0.5	0.085	0.117	0.188

This table presents simulation results from repeating the analysis in Table 3, except that we alter the amount of variation in the variable of interest ($X_{i,t}$) that is absorbed by firm-fixed effects. In particular, we assume the variable of interest includes a firm-specific component, $X_{i,t} = x_{i,t} + \sum_{i=1}^{100} \theta_i Firm_i$, where $x_{i,t} \sim N(0, 2 - a)$, $Firm_i$ is the i -th firm fixed effect, and θ_i is the coefficient on the i -th fixed effect ($\theta_i \sim N(0, a)$). We report simulations for the circumstance where the fixed effect structure absorbs 10%, 50%, and 90% of the variation in the independent variable of interest (i.e., $a = 0.2, 1$, and 1.8 respectively). All variables are as defined in Table 3. For each specification, we simulate the distribution of coefficients using a balanced panel of 100 firms and 30 years. We repeat the simulation 1000 times, and report the mean values of the resulting distribution of $\hat{\beta}$ s for each specification.

Table 5. Measurement Error, Fixed Effects, and False Positives

Rejection rates for two-tailed test of a true null hypothesis, $\hat{\beta}_1 = 0$, at the 5% significance level				
	(1)	(2)	(3)	(4)
	Fixed Effects <i>Excluded</i>	Fixed Effects <i>Included</i>	Fixed Effects <i>Included</i>	Fixed Effects <i>Included</i>
Measurement Error	Absorption =	Absorption =	Absorption =	Absorption =
	10%	50%	90%	
0.0	5%	5%	5%	5%
0.1	60%	62%	79%	94%
0.5	100%	100%	100%	100%

This table presents rejection rates for two-tailed test of a true null hypothesis, $\hat{\beta}_1 = 0$, at the 5% significance level from the simulation analysis in Table 4. We report simulations for the circumstance where the fixed effect structure absorbs 10%, 50%, and 90% of the variation in the independent variable of interest. For each specification, we simulate the distribution of coefficients using a balanced panel of 100 firms and 30 years. We repeat the simulation 1000 times, and report the percentage of times the simulation yields a significant $\hat{\beta}_1$ at a 5% level for each specification.

**Table 6. Measurement Error in the Literature:
Error Rates in Compustat Headquarter Data**

Percentage of observations where the state of headquarters listed on the respective data source differs from the 10-K filing		
Year	(1) <i># of Firms</i>	(2) <i>2019 Compustat</i>
1996	5,070	18.9%
1997	5,233	18.0%
1998	5,092	17.4%
1999	4,357	17.0%
2000	4,236	15.5%
2001	4,373	14.8%
2002	4,130	14.2%
2003	3,852	13.2%
2004	4,174	11.9%
2005	4,111	11.3%
2006	4,023	10.6%
2007	3,924	9.9%
2008	3,785	9.2%
2009	3,532	8.0%
2010	3,389	7.1%
2011	3,327	6.5%
2012	3,290	6.0%
2013	3,280	4.8%
Total	73,178	12.5%

This table presents the error rates for the state of corporate headquarters listed on Compustat. For each year, we calculate the percentage of observations where the state of corporate headquarters on Compustat differs from that on the front page of the firm's annual 10-K filing. Column (1) shows the number of firms each year. Column (2) reports the fraction of firms where the listed state of headquarters on the 2019 Compustat file is incorrect.

**Table 7. Measurement Error in the Literature:
Descriptive Statistics**

Variable	N	Years	Mean	Std	Q1	Median	Q3
Non-Compete Index Variables							
<i>NonCompete</i> [^]	73,178	1996-2013	3.82	2.31	3.00	4.00	5.00
<i>NonCompete</i> [*]	73,178	1996-2013	3.77	2.33	3.00	4.00	5.00
Disclosure Variables							
<i>Forecast</i>	73,178	1996-2013	2.97	6.15	0.00	0.00	3.00
<i>8KFiling</i>	73,178	1996-2013	7.59	7.60	1.00	6.00	12.00
<i>PressRelease</i>	36,835	2004-2013	18.52	26.97	5.00	14.00	24.00
<i>Restate</i>	53,426	2000-2013	0.02	0.14	0.00	0.00	0.00
Control Variables							
<i>Size</i>	73,178	1996-2013	5.58	2.08	4.06	5.55	6.99
<i>Mtb</i>	73,178	1996-2013	1.99	1.78	1.04	1.37	2.16
<i>Leverage</i>	73,178	1996-2013	0.61	1.61	0.00	0.20	0.73
<i>Roa</i>	73,178	1996-2013	-0.05	0.28	-0.04	0.02	0.07
<i>NumEst</i>	73,178	1996-2013	4.71	6.02	0.00	2.00	7.00
<i>NumInst</i>	73,178	1996-2013	68.43	118.79	0.00	16.00	90.00

This table presents summary statistics for the variables used in our analysis. *NonCompete*[^] is the index of state non-compete laws for the state of headquarters listed on the 2019 Compustat file. *NonCompete*^{*} is the index of state non-compete laws for the state of headquarters listed on the firm's annual 10-K filing each year. The index of state non-compete laws comes from Garmaise (2011), as updated by Ertimur, Rawlson, Rogers, and Zechman (2018). *Forecast* is the number of management forecast issued during the fiscal year. *8KFiling* is the number of 8Ks filed during the fiscal year. *PressRelease* is the number of firm-initiated press releases issued during the fiscal year. *Restate* is an indicator variable that equals one if financial results for that year were subsequently restated as a result of fraud, misrepresentation, or an SEC investigation, and zero otherwise. *Size* is the natural logarithm of market value of equity. *Mtb* is the market value of equity plus book value of liabilities divided by book value of assets. *Leverage* is long term debt, divided by book value of equity. *Roa* is income before extraordinary items divided by total assets. *NumEst* is number of analysts with one-year ahead earnings forecasts as of the fiscal year-end. *NumInst* is the number of institutional investors as of the fiscal year-end.

**Table 8. Measurement Error in the Literature:
Measurement Error Correlations**

Control Variables	Pearson Correlation <i>Error</i> (1)	Spearman Correlation <i>Error</i> (2)
<i>Size</i>	−0.015^{***} [<0.01]	−0.015^{***} [<0.01]
<i>Mtb</i>	0.002 [0.55]	0.012^{***} [<0.01]
<i>Roa</i>	−0.011^{***} [<0.01]	−0.009^{**} [0.02]
<i>Leverage</i>	−0.003 [0.41]	−0.014^{***} [<0.01]
<i>NumEst</i>	−0.010^{***} [<0.01]	−0.004 [0.29]
<i>NumInst</i>	−0.002 [0.60]	−0.001 [0.85]

This table presents correlations between the measurement error ($=NonCompete^{\wedge}-NonCompete^*$) and the set of control variables. two-tailed p -values appear in brackets. *, **, and *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

**Table 9. Measurement Error in the Literature:
Non-Compete Agreements and Disclosure**

	Dependent variable							
	<i>Forecast</i>		<i>8KFiling</i>		<i>PressRelease</i>		<i>Restate</i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable of Interest Measured With Error								
<i>NonCompete</i> [^]	-0.367*** (-2.82)		-0.251** (-2.19)		-1.452** (-2.02)		0.011** (2.02)	
Variable of Interest Measured Without Error								
<i>NonCompete</i> [*]		-0.021 (-0.44)		-0.012 (-0.22)		0.007 (0.03)		0.002 (1.01)
Controls								
<i>Size</i>	0.527*** (9.62)	0.528*** (9.63)	0.478*** (8.98)	0.479*** (8.99)	1.380*** (5.54)	1.387*** (5.55)	0.006*** (3.15)	0.006*** (3.14)
<i>Mtb</i>	-0.352*** (-13.55)	-0.351*** (-13.53)	-0.221*** (-8.70)	-0.221*** (-8.69)	-0.540*** (-5.86)	-0.540*** (-5.85)	-0.001 (-1.42)	-0.001 (-1.43)
<i>Leverage</i>	0.054*** (2.74)	0.054*** (2.74)	0.094*** (4.40)	0.094*** (4.40)	0.030 (0.52)	0.030 (0.51)	0.000 (0.07)	0.000 (0.07)
<i>Roa</i>	0.139 (1.14)	0.138 (1.13)	-1.509*** (-10.95)	-1.510*** (-10.99)	-1.811*** (-4.50)	-1.815*** (-4.51)	-0.008* (-1.76)	-0.008* (-1.78)
<i>NumEst</i>	0.148*** (9.14)	0.148*** (9.13)	-0.054*** (-3.29)	-0.054*** (-3.29)	0.281*** (5.27)	0.281*** (5.28)	0.001* (1.89)	0.001* (1.89)
<i>NumInst</i>	0.007*** (6.25)	0.007*** (6.26)	0.004*** (3.13)	0.004*** (3.14)	0.016** (1.96)	0.017** (1.97)	-0.000 (-1.08)	-0.000 (-1.10)
Firm Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	73,178	73,178	73,178	73,178	36,835	36,835	53,426	53,426
Adjusted R ²	0.574	0.573	0.685	0.685	0.865	0.865	0.314	0.314
Within R ²	0.039	0.038	0.010	0.010	0.020	0.020	0.002	0.002

This table presents results from estimating the relation between our four outcome variables and the non-compete index, calculated separately for the state of headquarters listed on the 2019 Compustat file (*NonCompete*[^]) and the state of headquarters listed on the firm's annual 10-K filing (*NonCompete*^{*}). All specifications include firm and year fixed effects. All variables are defined in Table 7. Standard errors are clustered by firm, *t*-statistics appear in parentheses, and *p*-values appear in brackets. *, **, and *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

**Table 10. Measurement Error in the Literature:
The Role of High Dimensional Fixed Effects**

<i>Panel A. Management Forecasts</i>								
Dependent variable: <i>Forecast</i>							Table 9 Results	
	Absorbed Variation: 0.05% <i>Year Effects</i>		Absorbed Variation: 3.62% <i>Industry and Year Effects</i>		Absorbed Variation: 3.72% <i>Industry × Year Effects</i>		Absorbed Variation: 99.49% <i>Firm and Year Effects</i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable of Interest Measured With Error								
<i>NonCompete</i> [^]	−0.033		−0.009		−0.002		−0.367***	
	(−1.49)		(−0.41)		(−0.11)		(−2.82)	
Variable of Interest Measured Without Error								
<i>NonCompete</i> [*]		−0.030		0.000		0.008		−0.021
		(−1.34)		(0.01)		(0.39)		(−0.44)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Year	Year	Industry, Year	Industry, Year	Industry ×Year	Industry ×Year	Firm,Year	Firm,Year
N obs	73,178	73,178	73,178	73,178	73,178	73,178	73,178	73,178
Adjusted R ²	0.251	0.251	0.306	0.306	0.333	0.333	0.574	0.573
Within R ²	0.130	0.130	0.137	0.137	0.132	0.132	0.039	0.038

**Table 10. Measurement Error in the Literature:
The Role of High Dimensional Fixed Effects (cont'd)**

<i>Panel B. 8-K Filings</i>								
Dependent variable: <i>8KFiling</i>							Table 9 Results	
	Absorbed Variation: 0.05% <i>Year Effects</i>		Absorbed Variation: 3.62% <i>Industry and Year Effects</i>		Absorbed Variation: 3.72% <i>Industry × Year Effects</i>		Absorbed Variation: 99.49% <i>Firm and Year Effects</i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable of Interest Measured With Error								
<i>NonCompete</i> [^]	−0.005		0.003		0.001		−0.251^{**}	
	(−0.28)		(0.16)		(0.07)		(−2.19)	
Variable of Interest Measured Without Error								
<i>NonCompete</i> [*]		−0.005		0.000		−0.001		−0.012
		(−0.31)		(0.01)		(−0.06)		(−0.22)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Year	Year	Industry, Year	Industry, Year	Industry ×Year	Industry ×Year	Firm,Year	Firm,Year
N obs	73,178	73,178	73,178	73,178	73,178	73,178	73,178	73,178
Adjusted R ²	0.479	0.479	0.486	0.486	0.489	0.489	0.685	0.685
Within R ²	0.083	0.083	0.078	0.078	0.078	0.078	0.010	0.010

**Table 10. Measurement Error in the Literature:
The Role of High Dimensional Fixed Effects (cont'd)**

<i>Panel C. Press Releases</i>								
Dependent variable: <i>PressRelease</i>							Table 9 Results	
	Absorbed Variation: 0.00% <i>Year Effects</i>		Absorbed Variation: 3.86% <i>Industry and Year Effects</i>		Absorbed Variation: 3.89% <i>Industry × Year Effects</i>		Absorbed Variation: 99.78% <i>Firm and Year Effects</i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable of Interest Measured With Error								
<i>NonCompete</i> [^]	−0.142		−0.093		−0.094		−1.452^{**}	
	(−1.25)		(−0.81)		(−0.81)		(−2.02)	
Variable of Interest Measured Without Error								
<i>NonCompete</i> [*]		−0.182		−0.127		−0.127		0.007
		(−1.60)		(−1.10)		(−1.10)		(0.03)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Year	Year	Industry, Year	Industry, Year	Industry ×Year	Industry ×Year	Firm,Year	Firm,Year
N obs	36,835	36,835	36,835	36,835	36,835	36,835	36,835	36,835
Adjusted R ²	0.200	0.200	0.209	0.209	0.208	0.208	0.865	0.865
Within R ²	0.185	0.185	0.190	0.190	0.190	0.190	0.020	0.020

**Table 10. Measurement Error in the Literature:
The Role of High Dimensional Fixed Effects (cont'd)**

<i>Panel D. Restatements</i>								
Dependent variable: <i>Restate</i>							Table 9 Results	
	Absorbed Variation: 0.05% <i>Year Effects</i>		Absorbed Variation: 3.96% <i>Industry and Year Effects</i>		Absorbed Variation: 4.01% <i>Industry × Year Effects</i>		Absorbed Variation: 99.60% <i>Firm and Year Effects</i>	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable of Interest Measured With Error <i>NonCompete</i> [^]	0.001 (1.05)		0.001 (1.26)		0.001 (1.24)		0.011 ^{**} (2.02)	
Variable of Interest Measured Without Error <i>NonCompete</i> [*]		0.001 (1.22)		0.001 (1.47)		0.001 (1.45)		0.002 (1.01)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Year	Year	Industry, Year	Industry, Year	Industry ×Year	Industry ×Year	Firm,Year	Firm,Year
N obs	53,426	53,426	53,426	53,426	53,426	53,426	53,426	53,426
Adjusted R ²	0.006	0.006	0.007	0.007	0.008	0.008	0.314	0.314
Within R ²	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

This table presents results from repeating the analysis in Table 9 using different fixed effects structures. ‘Absorbed Variation’ reports *R*-squared from a regression of *NonCompete*[^] on the set of specified fixed effects. Columns (1) and (2) present results with only year fixed effects. Columns (3) and (4) presents results with industry and year fixed effects. Columns (5) and (6) present results with industry and year interacted fixed effects. Columns (7) and (8) present results from Table 9 with firm and year fixed effects. All specifications include control variables from Table 9. All variables are defined in Table 7. Standard errors are clustered by firm. *, **, and *** denote statistical significance at 10%, 5%, and 1% levels, respectively.

Appendix A. Data on Corporate Headquarter Locations

Compustat (82 of 103)		
Ali et al. (2019)	Davidson et al. (2015)	Karpoff et al. (2017)
Aobdia (2015)	Davis et al. (2016)	Kedia and Rajgopal (2011)
Aobdia (2018)	Dessaint and Matray (2017)	Kedia et al. (2015)
Armstrong et al. (2012)	Di Giuli and Kostovetsky (2014)	Kini et al. (2016)
Ayers et al. (2011)	Dougal et al. (2015)	Klasa et al. (2018)
Balakrishnan et al. (2014)	Doyle and Magilke (2009)	Kostovetsky (2015)
Baloria and Heese (2018)	Duchin and Sosyura (2013)	Kraft et al. (2018)
Barzuza and Smith (2014)	Duchin and Sosyura (2012)	Kumar et al. (2011)
Belo et al. (2013)	Engelberg and Parsons (2011)	Lerner (2006)
Bernile et al. (2018)	Engelberg and Parsons (2016)	Lou (2014)
Bharath et al. (2011)	Engelberg et al. (2013)	Ma (2017)
Brown and Knechel (2016)	Fee et al. (2013)	McGuire et al. (2012)
Brown et al. (2015)	Francis et al. (2005)	Morsfield and Tan (2006)
Brown et al. (2008)	Garcia and Norli (2012)	Nikolaev (2018)
Bushee et al. (2018)	Gow et al. (2018)	Ovtchinnikov and Pantaleoni (2012)
Bushman and Wittenberg-Moerman (2012)	Hasan et al. (2014)	Parsons et al. (2018)
Bushman et al. (2018)	Hauser (2018)	Reppenhagen (2010)
Cadman and Sunder (2014)	He (2015)	Ross (2010)
Cain et al. (2017)	Hilary and Hui (2009)	Schwert (2018)
Call et al. (2016)	Hillert et al. (2014)	Seasholes and Zhu (2010)
Carter et al. (2019)	Hochberg and Lindsey (2010)	Shahrur (2005)
Carvalho (2018)	Hoi et al. (2013)	Shevlin et al. (2017)
Chang et al. (2015)	Hollander and Verrist (2016)	Shive (2012)
Chang et al. (2019)	Houston et al. (2014)	Shue (2013)
Chi and Shanthikumar (2017)	Ivkovic and Weisbenner (2005)	Stuart and Yim (2010)
Cornaggia et al. (2015)	Ivkovic and Weisbenner (2007)	Vashishtha (2014)
Cvijanovic (2014)	Jha and Chen (2015)	Wang and Xia (2014)
Dai et al. (2015)		
10-K (10 of 103)	SEC Header (5 of 103)	HL (6 of 103)
Alexander et al. (2013)	Bourveau et al. (2018)	Asker et al. (2015)
Armstrong et al. (2019)	Bushee and Miller (2012)	Farre-Mensa and Ljungqvist (2016)
Call et al. (2017)	Dyreng et al. (2013)	Glaeser (2018)
Hasan et al. (2017)	Houston et al. (2019)	Heider and Ljungqvist (2015)
Hoi et al. (2019)	Li et al. (2018)	Ljungqvist et al. (2017)
Huang et al. (2019)		Yost (2018)
Jennings et al. (2017)		
Kubick et al. (2017)		
Law and Mills (2015)		
Mukherjee et al. (2017)		

This table summarizes the data sources for headquarter location used in prior accounting and finance studies published in top 4 accounting journals (*The Accounting Review*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Review of Accounting Studies*) and top 3 finance journals (*Journal of Finance*, *Journal of Financial Economics*, *The Review of Financial Studies*) from 2005-2019. We consider four different data sources: *Compustat*—the Compustat quarterly or annual files; *10-K*—the state of headquarters listed on the front page of the firm’s 10-K; *SEC Header*—the SEC header file; and *HL*—Compustat after correcting using the Heider and Ljungqvist (2015) corrections.

References for Appendix A

- Alexander, C. R., Bauguess, S. W., Bernile, G., Lee, Y. H. A., & Marietta-Westberg, J. 2013. Economic effects of SOX Section 404 compliance: A corporate insider perspective. *Journal of Accounting and Economics*, 56(2-3), 267-290.
- Ali, A., Li, N. & Zhang, W. 2019. Restrictions on managers' outside employment opportunities and asymmetric disclosure of bad versus good news. *The Accounting Review*, 94(5), 1-25.
- Aobdia, D. 2015. Proprietary information spillovers and supplier choice: Evidence from auditors. *Review of Accounting Studies*, 20(4), 1504-1539.
- Aobdia, D. 2018. Employee mobility, noncompete agreements, product-market competition, and company disclosure. *Review of Accounting Studies*, 23(1), 296-346.
- Armstrong, C. S., Balakrishnan, K., & Cohen, D. 2012. Corporate governance and the information environment: Evidence from state antitakeover laws. *Journal of Accounting and Economics*, 53(1-2), 185-204.
- Armstrong, C.S., Glaeser, S., Huang, S. & Taylor, D.J. 2019. The economics of managerial taxes and corporate risk-taking. *The Accounting Review*, 94(1), 1-24.
- Asker, J., Farre-Mensa, J., & Ljungqvist, A. 2015. Corporate investment and stock market listing: A puzzle?. *The Review of Financial Studies*, 28(2), 342-390.
- Ayers, B.C., Ramalingegowda, S. & Yeung, P.E. 2011. Hometown advantage: The effects of monitoring institution location on financial reporting discretion. *Journal of Accounting and Economics*, 52(1), 41-61.
- Balakrishnan, K., Core, J. E., & Verdi, R. S. 2014. The relation between reporting quality and financing and investment: Evidence from changes in financing capacity. *Journal of Accounting Research*, 52(1), 1-36.
- Baloria, V. P., & Heese, J. 2018. The effects of media slant on firm behavior. *Journal of Financial Economics*, 129(1), 184-202.
- Barzuza, M., & Smith, D. C. 2014. What happens in Nevada? Self-selecting into lax law. *The Review of Financial Studies*, 27(12), 3593-3627.
- Belo, F., Gala, V. D., & Li, J. 2013. Government spending, political cycles, and the cross section of stock returns. *Journal of Financial Economics*, 107(2), 305-324.
- Bernile, G., Bhagwat, V., & Yonker, S. 2018. Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, 127(3), 588-612.
- Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. 2011. Lending relationships and loan contract terms. *The Review of Financial Studies*, 24(4), 1141-1203.
- Bourveau, T., Lou, Y., & Wang, R. 2018. Shareholder litigation and corporate disclosure: Evidence from derivative lawsuits. *Journal of Accounting Research*, 56(3), 797-842.
- Brown, J. R., Ivković, Z., Smith, P. A., & Weisbenner, S. 2008. Neighbors matter: Causal community effects and stock market participation. *The Journal of Finance*, 63(3), 1509-1531.
- Brown, N. C., Stice, H., & White, R. M. 2015. Mobile communication and local information flow: Evidence from distracted driving laws. *Journal of Accounting Research*, 53(2), 275-329.
- Brown, S. V., & Knechel, W. R. 2016. Auditor–client compatibility and audit firm selection. *Journal of Accounting Research*, 54(3), 725-775.
- Bushee, B. J., & Miller, G. S. 2012. Investor relations, firm visibility, and investor following. *The Accounting Review*, 87(3), 867-897.
- Bushee, B. J., Gerakos, J., & Lee, L. F. 2018. Corporate jets and private meetings with investors. *Journal of Accounting and Economics*, 65(2-3), 358-379.
- Bushman, R. M., & Wittenberg-Moerman, R. 2012. The role of bank reputation in “certifying” future performance implications of borrowers’ accounting numbers. *Journal of Accounting Research*, 50(4), 883-930.
- Bushman, R. M., Davidson, R. H., Dey, A., & Smith, A. 2018. Bank CEO materialism: Risk controls, culture and tail risk. *Journal of Accounting and Economics*, 65(1), 191-220.
- Cadman, B., & Sunder, J. 2014. Investor horizon and CEO horizon incentives. *The Accounting Review*, 89(4), 1299-1328.
- Cain, M. D., McKeon, S. B., & Solomon, S. D. 2017. Do takeover laws matter? Evidence from five decades of hostile takeovers. *Journal of Financial Economics*, 124(3), 464-485.
- Call, A. C., Kedia, S., & Rajgopal, S. 2016. Rank and file employees and the discovery of misreporting: The role of stock options. *Journal of Accounting and Economics*, 62(2-3), 277-300.

- Call, A. C., Campbell, J. L., Dhaliwal, D. S., & Moon Jr, J. R. 2017. Employee quality and financial reporting outcomes. *Journal of Accounting and Economics*, 64(1), 123-149.
- Carter, M. E., Franco, F., & Tuna, İ. 2019. Matching premiums in the executive labor market. *The Accounting Review*, 94(6), 109-136.
- Carvalho, D. 2018. How do financing constraints affect firms' equity volatility?. *The Journal of Finance*, 73(3), 1139-1182.
- Chang, X., Chen, Y., Wang, S. Q., Zhang, K., & Zhang, W. 2019. Credit default swaps and corporate innovation. *Journal of Financial Economics*, 134(2), 474-500.
- Chang, X., Fu, K., Low, A., & Zhang, W. 2015. Non-executive employee stock options and corporate innovation. *Journal of Financial Economics*, 115(1), 168-188.
- Chi, S. S., & Shanthikumar, D. M. 2017. Local bias in Google search and the market response around earnings announcements. *The Accounting Review*, 92(4), 115-143.
- Cornaggia, J., Mao, Y., Tian, X., & Wolfe, B. 2015. Does banking competition affect innovation?. *Journal of Financial Economics*, 115(1), 189-209.
- Cvijanović, D. 2014. Real estate prices and firm capital structure. *The Review of Financial Studies*, 27(9), 2690-2735.
- Dai, L., Parwada, J. T., & Zhang, B. 2015. The governance effect of the media's news dissemination role: Evidence from insider trading. *Journal of Accounting Research*, 53(2), 331-366.
- Davidson, R., Dey, A., & Smith, A. 2015. Executives' "off-the-job" behavior, corporate culture, and financial reporting risk. *Journal of Financial Economics*, 117(1), 5-28.
- Davis, A. K., Guenther, D. A., Krull, L. K., & Williams, B. M. 2016. Do socially responsible firms pay more taxes?. *The Accounting Review*, 91(1), 47-68.
- Dessaint, O., & Matray, A. 2017. Do managers overreact to salient risks? Evidence from hurricane strikes. *Journal of Financial Economics*, 126(1), 97-121.
- Di Giuli, A., & Kostovetsky, L. 2014. Are red or blue companies more likely to go green? Politics and corporate social responsibility. *Journal of Financial Economics*, 111(1), 158-180.
- Dougal, C., Parsons, C. A., & Titman, S. 2015. Urban vibrancy and corporate growth. *The Journal of Finance*, 70(1), 163-210.
- Doyle, J. T., & Magilke, M. J. 2009. The timing of earnings announcements: An examination of the strategic disclosure hypothesis. *The Accounting Review*, 84(1), 157-182.
- Duchin, R., & Sosyura, D. 2012. The politics of government investment. *Journal of Financial Economics*, 106(1), 24-48.
- Duchin, R., & Sosyura, D. 2013. Divisional managers and internal capital markets. *The Journal of Finance*, 68(2), 387-429.
- Dyregang, S. D., Lindsey, B. P., & Thornock, J. R. 2013. Exploring the role Delaware plays as a domestic tax haven. *Journal of Financial Economics*, 108(3), 751-772.
- Engelberg, J. E., & Parsons, C. A. 2011. The causal impact of media in financial markets. *The Journal of Finance*, 66(1), 67-97.
- Engelberg, J., & Parsons, C. A. 2016. Worrying about the stock market: Evidence from hospital admissions. *The Journal of Finance*, 71(3), 1227-1250.
- Engelberg, J., Gao, P., & Parsons, C. A. 2013. The Price of a CEO's Rolodex. *The Review of Financial Studies*, 26(1), 79-114.
- Farre-Mensa, J., & Ljungqvist, A. 2016. Do measures of financial constraints measure financial constraints?. *The Review of Financial Studies*, 29(2), 271-308.
- Fee, C. E., Hadlock, C. J., & Pierce, J. R. 2013. Managers with and without style: Evidence using exogenous variation. *The Review of Financial Studies*, 26(3), 567-601.
- Francis, J. R., Reichelt, K., & Wang, D. 2005. The pricing of national and city-specific reputations for industry expertise in the US audit market. *The Accounting Review*, 80(1), 113-136.
- Garcia, D., & Norli, Ø. 2012. Geographic dispersion and stock returns. *Journal of Financial Economics*, 106(3), 547-565.
- Glaeser, S. 2018. The effects of proprietary information on corporate disclosure and transparency: Evidence from trade secrets. *Journal of Accounting and Economics*, 66(1), 163-193.
- Gow, I. D., Wahid, A. S., & Yu, G. 2018. Managing reputation: Evidence from biographies of corporate directors. *Journal of Accounting and Economics*, 66(2-3), 448-469.
- Hasan, I., Hoi, C. K., Wu, Q., & Zhang, H. 2014. Beauty is in the eye of the beholder: The effect of corporate tax avoidance on the cost of bank loans. *Journal of Financial Economics*, 113(1), 109-130.

- Hasan, I., Hoi, C. K., Wu, Q., & Zhang, H. 2017. Does social capital matter in corporate decisions? Evidence from corporate tax avoidance. *Journal of Accounting Research*, 55(3), 629-668.
- Hauser, R. 2018. Busy directors and firm performance: Evidence from mergers. *Journal of Financial Economics*, 128(1), 16-37.
- He, G. 2015. The effect of CEO inside debt holdings on financial reporting quality. *Review of Accounting Studies*, 20(1), 501-536.
- Heider, F., & Ljungqvist, A. 2015. As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes. *Journal of Financial Economics*, 118(3), 684-712.
- Hilary, G., & Hui, K. W. 2009. Does religion matter in corporate decision making in America?. *Journal of Financial Economics*, 93(3), 455-473.
- Hillert, A., Jacobs, H., & Müller, S. 2014. Media makes momentum. *The Review of Financial Studies*, 27(12), 3467-3501.
- Hochberg, Y. V., & Lindsey, L. 2010. Incentives, targeting, and firm performance: An analysis of non-executive stock options. *The Review of Financial Studies*, 23(11), 4148-4186.
- Hoi, C. K., Wu, Q., & Zhang, H. 2013. Is corporate social responsibility (CSR) associated with tax avoidance? Evidence from irresponsible CSR activities. *The Accounting Review*, 88(6), 2025-2059.
- Hoi, C. K., Wu, Q., & Zhang, H. 2019. Does social capital mitigate agency problems? Evidence from Chief Executive Officer (CEO) compensation. *Journal of Financial Economics*, 133(2), 498-519.
- Hollander, S., & Verriest, A. 2016. Bridging the gap: the design of bank loan contracts and distance. *Journal of Financial Economics*, 119(2), 399-419.
- Houston, J. F., Jiang, L., Lin, C., & Ma, Y. 2014. Political connections and the cost of bank loans. *Journal of Accounting Research*, 52(1), 193-243.
- Houston, J. F., Lin, C., Liu, S., & Wei, L. 2019. Litigation risk and voluntary disclosure: Evidence from legal changes. *The Accounting Review*, 94(5), 247-272.
- Huang, A., Hui, K. W., & Li, R. Z. 2019. Federal judge ideology: A new measure of ex ante litigation risk. *Journal of Accounting Research*, 57(2), 431-489.
- Ivković, Z., & Weisbenner, S. 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, 60(1), 267-306.
- Ivković, Z., & Weisbenner, S. 2007. Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *The Review of Financial Studies*, 20(4), 1327-1357.
- Jennings, J., Lee, J., & Matsumoto, D. A. 2017. The effect of industry co-location on analysts' information acquisition costs. *The Accounting Review*, 92(6), 103-127.
- Jha, A., & Chen, Y. 2015. Audit fees and social capital. *The Accounting Review*, 90(2), 611-639.
- Karpoff, J. M., Schonlau, R. J., & Wehrly, E. W. 2017. Do takeover defense indices measure takeover deterrence?. *The Review of Financial Studies*, 30(7), 2359-2412.
- Kedia, S., & Rajgopal, S. 2011. Do the SEC's enforcement preferences affect corporate misconduct?. *Journal of Accounting and Economics*, 51(3), 259-278.
- Kedia, S., Koh, K., & Rajgopal, S. 2015. Evidence on contagion in earnings management. *The Accounting Review*, 90(6), 2337-2373.
- Kini, O., Shenoy, J., & Subramaniam, V. 2017. Impact of financial leverage on the incidence and severity of product failures: Evidence from product recalls. *The Review of Financial Studies*, 30(5), 1790-1829.
- Klasa, S., Ortiz-Molina, H., Serfling, M., & Srinivasan, S. 2018. Protection of trade secrets and capital structure decisions. *Journal of Financial Economics*, 128(2), 266-286.
- Kostovetsky, L. 2015. Political capital and moral hazard. *Journal of Financial Economics*, 116(1), 144-159.
- Kraft, A. G., Vashishtha, R., & Venkatachalam, M. 2018. Frequent financial reporting and managerial myopia. *The Accounting Review*, 93(2), 249-275.
- Kubick, T. R., Lockhart, G. B., Mills, L. F., & Robinson, J. R. 2017. IRS and corporate taxpayer effects of geographic proximity. *Journal of Accounting and Economics*, 63(2-3), 428-453.
- Kumar, A., Page, J. K., & Spalt, O. G. 2011. Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics*, 102(3), 671-708.
- Law, K. K., & Mills, L. F. 2015. Taxes and financial constraints: Evidence from linguistic cues. *Journal of Accounting Research*, 53(4), 777-819.
- Lerner, J. 2006. The new new financial thing: The origins of financial innovations. *Journal of Financial Economics*, 79(2), 223-255.
- Li, Y., Lin, Y., & Zhang, L. 2018. Trade secrets law and corporate disclosure: Causal evidence on the proprietary cost hypothesis. *Journal of Accounting Research*, 56(1), 265-308.

- Ljungqvist, A., Zhang, L., & Zuo, L. 2017. Sharing risk with the government: How taxes affect corporate risk taking. *Journal of Accounting Research*, 55(3), 669-707.
- Lou, D. 2014. Attracting investor attention through advertising. *The Review of Financial Studies*, 27(6), 1797-1829.
- Ma, M. 2017. Economic links and the spillover effect of earnings quality on market risk. *The Accounting Review*, 92(6), 213-245.
- McGuire, S. T., Omer, T. C., & Sharp, N. Y. 2012. The impact of religion on financial reporting irregularities. *The Accounting Review*, 87(2), 645-673.
- Morsfield, S. G., & Tan, C. E. 2006. Do venture capitalists influence the decision to manage earnings in initial public offerings?. *The Accounting Review*, 81(5), 1119-1150.
- Mukherjee, A., Singh, M., & Zaldokas, A. 2017. Do corporate taxes hinder innovation?. *Journal of Financial Economics*, 124(1), 195-221.
- Nikolaev, V. V. 2018. Scope for renegotiation in private debt contracts. *Journal of Accounting and Economics*, 65(2-3), 270-301.
- Ovtchinnikov, A. V., & Pantaleoni, E. 2012. Individual political contributions and firm performance. *Journal of Financial Economics*, 105(2), 367-392.
- Parsons, C. A., Sulaeman, J., & Titman, S. 2018. The geography of financial misconduct. *The Journal of Finance*, 73(5), 2087-2137.
- Reppenhausen, D. A. 2010. Contagion of accounting methods: Evidence from stock option expensing. *Review of Accounting Studies*, 15(3), 629-657.
- Ross, D. G. 2010. The “dominant bank effect.” How high lender reputation affects the information content and terms of bank loans. *The Review of Financial Studies*, 23(7), 2730-2756.
- Schwert, M. 2018. Bank capital and lending relationships. *The Journal of Finance*, 73(2), 787-830.
- Seasholes, M. S., & Zhu, N. 2010. Individual investors and local bias. *The Journal of Finance*, 65(5), 1987-2010.
- Shahrur, H. 2005. Industry structure and horizontal takeovers: Analysis of wealth effects on rivals, suppliers, and corporate customers. *Journal of Financial Economics*, 76(1), 61-98.
- Shevlin, T., Thornock, J., & Williams, B. 2017. An examination of firms’ responses to tax forgiveness. *Review of Accounting Studies*, 22(2), 577-607.
- Shive, S. 2012. Local investors, price discovery, and market efficiency. *Journal of Financial Economics*, 104(1), 145-161.
- Shue, K. 2013. Executive networks and firm policies: Evidence from the random assignment of MBA peers. *The Review of Financial Studies*, 26(6), 1401-1442.
- Stuart, T. E., & Yim, S. 2010. Board interlocks and the propensity to be targeted in private equity transactions. *Journal of Financial Economics*, 97(1), 174-189.
- Vashishtha, R. 2014. The role of bank monitoring in borrowers’ discretionary disclosure: Evidence from covenant violations. *Journal of Accounting and Economics*, 57(2-3), 176-195.
- Wang, Y., & Xia, H. 2014. Do lenders still monitor when they can securitize loans?. *The Review of Financial Studies*, 27(8), 2354-2391.
- Yost, B. P. 2018. Locked-in: The effect of CEOs’ capital gains taxes on corporate risk-taking. *The Accounting Review*, 93(5), 325-358.

Appendix B. Alternative Data Sources for State of Corporate Headquarters

This Appendix discusses Heider and Ljungqvist (2015) and the SEC header file as two alternative sources of data on state of corporate headquarters.

Heider and Ljungqvist

Heider and Ljungqvist (2015, HL) make manual corrections to the 2013 Compustat file which they have shared with other researchers. HL provide a list of corrections that would make the 2013 Compustat file accurate. While the corrections themselves are not problematic, it is problematic to apply these corrections to more recent Compustat files for two reasons.

- 1) These corrections were derived only on the specific sample used in HL: a sample that excluded financials and utilities, and focuses only on NYSE/NASDAQ/AMEX listed companies. Applying these corrections to the broader universe of firms on Compustat will not correct the state of headquarters for firms that were not in their sample.
- 2) These corrections are to the 2013 Compustat file. As a result, they will miss any subsequent changes in corporate headquarters and associated back-filling. For example, consider the firm that moved from Texas to California in 2015. In 2015, Compustat would back-fill all of the firm's observations back to 1968 with a California headquarters. Applying the corrections would not correct for this, because the data from 1968 onward was correct as of 2013 and no corrections were recorded. In 2015, all observations for this firm between 1968-2014 would be incorrect, and would indicate California.

Given these issues, HL share their corrections with the following caveats:

```
** This Stata do-file corrects Compustat's backfilled HQ states to actual  
historic HQ states for the period 1989-2011.  
* It is advisable to delete pre-1989 observations, as the code will  
otherwise backfill pre-1989 observations.  
* For similar reasons, it is advisable to delete post-2011 observations:  
the code will not correct post-2011 errors in Compustat.
```

We clarify it is not sufficient to simply delete post-2011 observations. Because of back-filling, all observations, even those before 2011, will be incorrect if the firm moved after 2013. The literature seems to be unaware of these issues—these issues notwithstanding—a large and growing number of published and unpublished papers continues to apply the HL corrections to recent Compustat files (e.g., Hanlon, Verdi, and Yost, 2020; Lai, Li, and Yang, 2020).

We assess error rates in the state of corporate headquarters if the HL corrections were applied to the current Compustat file. We repeat the analysis in Table 6, comparing the listed state of headquarters on each file against the firm's annual 10-K filing. Table B1 presents results. Column (1) presents error rates for the 2019 Compustat file after applying the HL corrections. Consistent with the corrections missing any change in headquarters and associated backfilling subsequent to 2013, we find error rates of 10.8% in 1996 that decline to 4.8% in 2013. The findings in Table B1 make clear that simply deleting observations after 2011 will not rectify the problem.

SEC Header

The SEC maintains a header file for each corporate filing, and this file contains the state of headquarters and business address. However, these header files are not always current, and are sometimes updated with a significant lag. For example, Bemis Company moved their headquarters from Minnesota to Wisconsin in 2006, but the SEC header file continues to show Minnesota as the state of headquarters through 2010. Column (2) of Table B1 presents error rates for the SEC header file. Of all data sources considered, the SEC header best replicates the state of headquarters listed on the firm's annual 10-K filing. Column (2) shows error rates for the SEC header data are consistently around 2%.

Table B1. Error Rates in Alternative Data Sources

Percentage of observations where the state of headquarters listed on the respective data source differs from the 10-K filing		
	(1)	(2)
Year	<i>2019 Compustat HL corrected</i>	<i>SEC Header</i>
1996	10.8%	2.1%
1997	11.1%	2.0%
1998	10.4%	1.8%
1999	10.7%	2.2%
2000	10.0%	2.7%
2001	9.8%	3.2%
2002	9.5%	2.7%
2003	9.0%	1.9%
2004	8.7%	1.8%
2005	8.5%	1.8%
2006	8.3%	1.9%
2007	8.3%	1.7%
2008	8.2%	1.9%
2009	7.5%	1.7%
2010	6.4%	1.8%
2011	6.2%	1.7%
2012	6.0%	1.4%
2013	4.8%	1.3%
Total	8.8%	2.0%

This table presents the error rates for alternative data sources for the state of corporate headquarters from 1996 to 2013. For each year, we calculate the percentage of observations where the state of corporate headquarters differs from that on the front page of the firm's annual 10-K filing. Column (1) reports results for the state of headquarters on the 2019 Compustat file after applying the Heider and Ljungqvist (2015) corrections. Column (2) reports results for the SEC Header file as of May 2020.