Economic Magnitudes Within Reason*

Zack Liu† Adam Winegar‡

November 2023

Abstract

A common method of calculating economic magnitudes is to multiply the regression coefficient of the variable of interest by its sample standard deviation. This method is often problematic in finance settings when researchers use granular fixed effects. We show that in many recently published finance papers and for many common finance variables, the sample standard deviation is much larger than the within-group variation that identifies the regression coefficient, and that within-group changes of this magnitude are rare. Without additional assumptions, this common approach can significantly inflate the economic magnitude of the identified effect and impact the comparison of effects among different variables of interest. We recommend using within-group measures of variation to improve the interpretation of economic magnitudes in this setting.

*We thank Jonathan Cohn, Alex Edmans, William Grieser, Erik Mayer, Øyvind Norli, Paul Povel, Sheridan Titman, Mitch Towner, Malcolm Wardlaw, and seminar participants at BI Norwegian Business School for comments. We thank Xi Chen for providing excellent research assistance.

†C.T. Bauer College of Business, University of Houston. E-mail: zliu@bauer.uh.edu. Address: 230E Melcher Hall, 4250 MLK Blvd, Houston, TX 77204. Phone: 713-743-4764

‡BI Norwegian Business School. E-mail: adam.w.winegar@bi.no.

The importance of empirical findings depends on both the statistical significance and the economic magnitude (McCloskey and Ziliak, 1996). In finance research, the workhorse regression specification is a fixed-effects regression with firm (or industry) dummies to account for time-invariant and often unobservable characteristics that may be related to the variable of interest and outcome variable. To determine the economic magnitude of the estimate, researchers commonly multiply the regression coefficient of the variable of interest by its sample standard deviation (or standardize the coefficient to represent a one standard deviation move). Ideally, this approach estimates the effect of a reasonable change in the underlying variable and allows for comparisons across variables of interest. In this paper, we argue that this approach of estimating economic magnitudes of fixed-effects regression coefficients often fails to achieve these goals in empirical finance research.

To illustrate the potential pervasiveness of this issue, we survey papers published in the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies in 2020 and 2021. In Table 1, we show the total number of papers published and the total number that run fixed-effect regressions. Within the second group, we show the total number and proportion of those papers that calculate economic magnitudes using the one standard deviation approach. Among the papers that run fixed-effect regressions, we find that 39%, 46%, and 48% of the papers published in the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies, respectively, use this approach. Although many of these papers use relatively complex and clever methods to identify the effect, little attention is paid to whether the counterfactual change is reasonable relative to the source of the identification.

The finance literature provides ample evidence supporting the use of granular fixed effects to control for unobserved heterogeneity.[1] By incorporating fixed effects into a regression model, including 2SLS regressions, the researcher chooses to focus on the within-group variation in the variables to estimate the regression coefficients.[2] In these regressions, the fixed effects often explain a significant amount of variation and affect the coefficient estimates, indicating large between-group differences in the both variable of interest and the outcome.

[1] Lemmon, Roberts, and Zender (2008) and Graham, Li, and Qiu (2012) examine the importance of fixed effects for leverage and CEO pay, respectively. Also, see Coles and Li (2023) for a survey of the explanatory importance of fixed effects for a wider array of corporate finance variables.

[2] For 2SLS, the within-group variation in the first-stage predicted values is bounded by the within-group variation of the variable of interest.

The primary concern with scaling by the sample standard deviation to generate counterfactuals is extrapolation. When the counterfactual change is significantly larger than the common variation in the data, the estimated counterfactual relies heavily on the underlying modeling assumptions (e.g., linearity), which may not be directly testable or defensible for large changes (King and Zeng, 2006). Furthermore, if the sample standard deviation for the variable of interest is much larger than typical within-group changes, then it is likely that some time-invariant group characteristics are responsible for the change. However, the relation between those unobserved characteristics and the outcome is not identified by the regression. Additionally, this standardization can create misleading comparisons if some variables have high levels of within-group variation while others have relatively low levels, which we show is the case for a common set of financial variables. Overall, without additional assumptions about the data-generating process, the use of sample standard deviations can greatly overstate the true economic magnitude when estimating the within-group effect.

In this paper, we first establish the prevalence and severity of this issue within finance research. Specifically, we replicate 14 papers from the set of papers we identify in Table 1 and find that 7 (3) of these papers report an economic magnitude estimate that is more than 50% (100%) larger than what it would be if the researchers use the identifying variation in the paper. Next, to demonstrate this issue is not unique to the replicated studies, we show in a more standardized setting that the sample standard deviation for commonly used finance variables is often two to three times larger than the within-firm standard deviation and that there is high variation in the ratio of within-firm standard deviation to sample standard deviation across the variables. Moreover, for these variables, we find that the within-firm differences of one sample standard deviation are rare and that even the 10-year peak-to-trough difference for most firms is much less than the sample standard deviation. Thus, for firm fixed-effects specifications, not only does a sample standard deviation change overestimate a reasonable within-firm change to the variable of interest, but the heterogeneity complicates comparisons across variables even in the same setting.

We propose several suggestions to better capture economic magnitudes using reasonable within-group variation. First, the researcher should specify the identifying variation used in the regression specification. Second, when appropriate, the researcher should report the identifying within-group variation by calculating and reporting the within-group standard deviation of the variable of interest. Third, the researcher can examine and report statistics about the empirical distribution of within-group changes for the variable of interest. In sum, the researcher should consider their specific application to determine the relevant counterfactual and defend their choice, instead of relying on a universal metric.[3]

The key takeaway is that while it is common and useful to estimate economic magnitudes using standard deviations for scaling, it is important to justify why the stated variation or counterfactual is reasonable. When gauging the economic magnitude of the variable of interest for a fixed-effects regression coefficient, the reasonable counterfactual should adhere to the following conditions. First, the counterfactual should be based on the within-group variation that is used to estimate the coefficient. Second, the counterfactual should occur regularly in the data. By satisfying these conditions, the counterfactual can be viewed as reasonable and likely to occur exogenously within the regression framework. Our recommendation to calculate economic magnitudes using within-group measures facilitates comparisons across variables by matching the degree of variation to the granularity of the estimated effects. In return, the researcher relies less on strong, implicit assumptions about the data-generating process when comparing effects.

While we are the first to highlight the severity of this issue among published finance papers, we are not the first to raise the more general issue of using the sample standard deviation as a scaling variable in fixed-effects regressions (Mummolo and Peterson, 2018; deHaan, 2021). We contribute to the broader discussion by showing that within-group moves greater than the sample standard deviation are rare and that heterogeneity in within-group variation complicates comparisons of economic magnitudes among variables.

There exist alternative methods for assessing economic significance. Some researchers argue for the use of R^2, or incremental R^2, to evaluate the importance of a variable (Frank and Goyal, 2009). Similar in spirit to R^2, Mitton (Forthcoming) recommends determining the economic significance by multiplying the coefficient of interest by the standard deviation of the variable of interest and dividing by the standard deviation of the outcome variable. While Mitton (Forthcoming) and our paper both critique the commonly used approach, our paper focuses on the interpretation of fixed-effect regression coefficient estimates, and we recommend using within-group measures of changes to match the variation used to identify

[3]This concern is not relevant for papers that use binary treatments and focus on the effect of the binary treatment directly. However, this concern is valid for papers that use binary treatments as a proxy for a continuous variable and is similar and related to concerns regarding the local average treatment effects and average treatment effects (Angrist and Imbens, 1994; Angrist and Pischke, 2009).

the regression coefficient. We show that these within-group measures represent reasonable within-group moves and thus using these measures to calculate economic magnitudes provides more reliable estimates of counterfactual changes to inform both researchers and policymakers. If a researcher wants to apply the Mitton (Forthcoming) intuition to quantify the economic significance of a fixed-effects coefficient, then we recommend using within-group standard deviation for both the variable of interest and the outcome variable. Even in this context, using within-group measures is important because we show that there is often wide heterogeneity in the ratio of within-group to sample standard deviation across variables, which in turn could re-order the relative importance rankings of these variables. Additionally, to determine the counterfactual change in the outcome variable, the within-group standard deviation still needs to be reported and discussed to provide context for the unitless Mitton (Forthcoming) measure.

This paper builds on the well-established literature on the advantages of using fixed-effects regressions in panel data settings (Deaton, 1985; Arellano and Honoré, 2001; Angrist and Pischke, 2009; Wooldridge, 2010).[4] More broadly this paper adds to the growing recent literature that critiques common econometric tools within finance research, especially for corporate finance. These include the recent re-evaluation of the application of difference-in-difference regressions (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021) and their use in finance (Baker, Larcker, and Wang, 2022), the validity of fixed effect assumptions (Grieser and Hadlock, 2019), methodological variations in variable selection and outlier treatments (Mitton, 2022), the repeated use of natural experiments (Heath et al., Forthcoming), and log-linear regressions (Cohn, Liu, and Wardlaw, 2022). These papers are primarily concerned with discussions of the identification of an effect and do not discuss the economic interpretation, the focus of this paper.

1 Motivating Example and Econometrics Primer

1.1. Motivating Example

To understand the issues discussed in this paper, consider a hypothetical study that seeks to estimate the effect of firm size, e.g., $log(Assets)$, on CEO pay. In this study, the regression of CEO pay on firm size should include firm fixed effects to account for the

[4] Roberts and Whited (2013) provides a survey of identification concerns in corporate finance.

fact that firms differ in compensation practices that are constant through time but cannot be measured or observed by the researcher (Graham, Li, and Qiu, 2012). This regression coefficient would estimate the within-firm effect of a change in firm size on CEO pay. To gauge the economic significance of this effect, the researcher could multiply the estimated coefficient by a reasonable within-firm change in size. For the period 1975 to 2018, the sample standard deviation of $log(Assets)$ is 2.422, meaning a positive sample standard deviation shock is a ten-fold increase in assets.[5] Within-firm changes of $log(Assets)$ of this magnitude are exceedingly rare. Thus, without additional assumptions, if the researcher calculates the economic magnitude by multiplying the coefficient of interest by the sample standard deviation of the variable, the resulting economic magnitude would be overstated compared to the effect of a more reasonable and commonly observed within-firm shock.

Next, consider comparing the impact of firm size and Tobin's Q on CEO pay. Again, the researcher follows the literature and estimates a regression with firm-level fixed effects. Now suppose that the two coefficients have the same magnitude and the researcher scales by the respective sample standard deviations. If firm size and Tobin's Q have similar sample standard deviations, then one might conclude that both variables are of similar importance and that a reasonable counterfactual of both variables would have the same effect. However, as we will show, a significantly larger portion of the variation in Tobin's Q is within-firm but a much smaller fraction of the variation in firm size is within-firm. Therefore, without stronger assumptions, the importance of Tobin's Q in explaining the within-firm variation in the CEO pay is significantly larger than that of firm size, and a reasonable counterfactual change for Tobin's Q would have a larger impact on CEO pay.

This motivating example illustrates the two primary concerns with the common approach to calculating economic magnitudes. First, the within-firm standard deviations are often much smaller than the sample standard deviation. Thus, the common approach may overestimate economic magnitudes relative to a reasonable benchmark. Second, there is heterogeneity among variables in within-firm standard deviations. Therefore, the common approach may incorrectly rank the importance of explanatory variables.

[5] $e^{2.422} - 1 = 10.268$.

1.2. *Econometrics Primer*

The standard ordinary least squares regression estimates coefficients that describe the linear relation between one or more independent variables $\mathbf{x}$ and an outcome variable y. For a set of observations indexed by i, the OLS regression has the form

$$y_i = \mathbf{x}_i \beta + \epsilon_i, \tag{1}$$

where β is a vector of coefficients and ϵ is a mean-zero error. The OLS regression will estimate β_j as the average change in y for a one-unit change in independent variable x_j.

While β_j is informative about the average relationship, β_j is not informative about how much of the variation in y each independent variable contributes because it does not include information about the variation in x_j. Thus, when interpreting OLS coefficients, researchers often calculate economic magnitudes by multiplying the regression coefficient by the standard deviation of that independent variable. Since the observations are assumed to be independent, the researcher is relying on the distribution of x_j to determine a reasonable counterfactual shift.

In many finance applications, researchers have access to panel data sets. To control for unobserved, invariant group differences, researchers add fixed effect dummies to their regressions. The OLS fixed-effects regression commonly takes the form

$$y_{i,t} = \mathbf{x}_{i,t} \beta + \gamma_{i,g} + \epsilon_{i,t}, \tag{2}$$

where g is the group that the observation is a member of (i.e., industry or firm) and t denotes the time of the observation. In this panel data context, each independent variable x can be decomposed into a within-group component and a between-group component. Assuming a homogeneous treatment effect, the OLS fixed-effects regression will estimate β as the average change in y for a one-unit change in the within-group component of x. While the OLS fixed-effects model can estimate the group dummies, it does not recover the relation between the between-group component of x and y. Since the OLS fixed-effects regression is identified within-group, unlike OLS without fixed effects, the reasonable counterfactual based on within-firm variation may be different too. We illustrate this concern using simulated data in Appendix A3.

While this discussion focuses on applications where a single fixed effect absorbs significant variation, the intuition and recommendations are applicable to other similar settings such as two-way (e.g., firm and year) or higher dimensional fixed effects (e.g., firm, industry $\times$ year). The main takeaway is that significant differences between within-group and sample standard deviation cause economic magnitudes calculated using the sample standard deviation to be too large relative to the identified within-group effect.

2 Applications to Real Data

2.1. Replications of Published Papers

Standardizing or using the sample standard deviation for calculating economic magnitudes is potentially misleading if there is a large difference between the within-group and sample standard deviations. Moreover, from Table 1, we find that among recently published papers 46% include economic magnitude calculations where the fixed-effect regression coefficient is multiplied by the sample standard deviation, but few papers justify the use of this counterfactual.[6] To assess whether using this measure of variation creates a potential problem, we replicate and analyze 23 variables of interest from 14 papers for which we have either available replication packages or access to the underlying data that we identify in the survey of recently published papers from Table 1.[7]

Starting with the papers published in the top three finance journals that utilize fixed-effects regressions and discuss economic magnitudes using the sample standard deviation, we identify papers with complete replication packages, rely on data from common financial database providers (e.g., CRSP, COMPUSTAT), or use data that is publicly available (e.g., Home Price Indices, Call Reports data).[8] From the initial set of papers, we identify papers that we could readily replicate and for which our replication approximately matches the reported summary statistics from the paper, using the code provided by the authors or follow the data-cleaning procedures described in the paper.[9] Note for the purposes of this

[6](Mitton, Forthcoming) finds that 65% of papers that report economic significance from 2000-2018 in empirical corporate with a continuous variable use this approach.

[7]Of these papers, only Griffin, Kruger, and Maturana (2021) explicitly mentions within-group standard deviations.

[8]Even though many journals now have data and code sharing policies, many of the papers in Table 1 were submitted prior to these requirements. Moreover, a large number of these papers require either proprietary data, special access, or hand-collected data, which severely limits our ability to replicate the underlying data.

[9]We denote whether we utilized the provided replication package in Table 2. For the replications without replication packages, we show how closely we match the summary statistics of the paper in Table A1 in the

paper, we do not attempt to replicate the complete findings of the paper, instead we only attempt to replicate the distribution of a key variable of interest. Overall, we are able to replicate a total of 23 key variables of interest from 14 papers.

Next, we calculate the sample standard deviation and the within-group standard deviation for these variables. For papers that utilize multiple or higher-dimensional fixed effects, we estimate the within-group standard deviation using the same set of fixed effects that correspond to the reported economic magnitude estimates of the given paper.[10] Using our estimates of the within-group standard deviation, we then calculate the ratio of the within-group standard deviation to the sample standard deviation (W-S Ratio). This ratio also represents the proportional change in the economic magnitudes from using within-group versus the sample standard deviation. Table 2 lists the journal, the paper citation, the variable of interest, the fixed effects, the sample standard deviation, the within-group standard deviation, the ratio of within to sample standard deviation, the reported economic magnitude in the paper, and the economic magnitude adjusted by the W-S Ratio.

Using the W-S Ratio, we classify the variables into four groups. First, 4 of 23 variables (4 papers) have a W-S Ratio that is less than or equal to 50%, and thus scaling using the sample standard deviation would result in economic magnitudes that are more than 100% larger than if estimated using the identifying variation of the variable. Second, 6 of 23 variables (3 papers) have a W-S Ratio that is between 50% and 66%, and thus the economic magnitudes are between 50-100% too large. For these two groups of variables, we view the difference between the within-group and sample standard deviation as severe. Hence, we have serious reservations about relying on the reported economic magnitudes.

Third, 3 of 23 variables (3 papers) have a W-S Ratio that is between 66% and 75%, and scaling using the sample standard deviation would result in economic magnitudes between 33-50% larger than if estimated with the identifying variation of the paper. While not as severe a difference as the first two groups, we still recommend a degree of caution for the reported economic magnitudes. Finally, 10 of 23 variables (6 papers) have a W-S Ratio that is greater than or equal to 75%, and thus the economic magnitudes would be only 33% larger if estimated using the identifying variation in the paper. For these papers, we consider

Appendix.

[10] To estimate within-group standard deviations, we regress the given variable of interest on the listed fixed effects and calculate the standard deviation of the residuals.

scaling using the standard deviation less worrisome.

There are two key takeaways from this analysis. First, it is often difficult to predict ex ante which variables will have within-group identifying variation significantly smaller than the sample standard deviation. Although it is generally true that more granular, higher-dimensional fixed effects will result in less within-group variation, some variables maintain substantial variation despite a relatively large number of fixed effects. The opposite could be true too, where the within-group identifying variation is substantially reduced even when there are relatively few fixed effects. Thus, as we will discuss in Section 3, researchers who utilize fixed-effects regressions should generally report the degree of identifying variation regardless of the extent of the fixed effects. Second, a substantial fraction of papers (7 of 14) have key variables where the identifying variation is substantially smaller than the sample standard deviation. Combined with the survey from Table 1, this evidence suggests that overstating economic magnitudes due to scaling choices is common in financial economics research.

2.2. Common Finance Variables

In order to further illustrate issues that arise when estimating economic magnitudes when using fixed effects regressions, we now turn to a more standardized setting to demonstrate the issues observed above are not unique to the replications. In this section, we illustrate that it is exceedingly common for the sample standard deviation of finance variables to be much greater than the within-firm standard deviation – the identifying variation for firm fixed-effect regressions. Moreover, we compare the ratio of the within-firm standard deviation to the sample standard deviation (W-S Ratio) for these common finance variables and find that their W-S Ratios vary greatly. Thus, not only should we be concerned about overstating economic magnitudes, but we also cannot rely on relative variable importance rankings based on economic magnitudes estimated using the sample standard deviation.

Using this standard sample, we can also check the reasonableness of a sample standard deviation or within-firm standard deviation change in the variable of interest. Using a normally distributed variable as a benchmark, approximately 68% of values will fall within one standard deviation of the mean, and the probability that an observation is more than one standard deviation away from another observation is approximately 48%.[11] It is well

[11] This is calculated as $1 - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\mathbb{1}(x-\sigma<y) - \mathbb{1}(x+\sigma<y))\phi(x)dx\phi(y)dy$, where ϕ represents the pdf of the standard normal distribution.

known that standard deviations may not provide a complete description of the dispersion or variability for distributions with large skewness or kurtosis. As we will show, the distribution of many variables is highly skewed and, moreover, a sample standard deviation change is not common. However, we find that within-group distributions are better behaved and that a within-group standard deviation change is relatively more common.

For this standardized setting, we identify the within-group and sample standard deviations for fourteen common financial and accounting variables. Table 3 reports the within-firm, within-industry, and sample standard deviations for these fourteen variables using a sample of COMPUSTAT firms from 1975-2018. We apply standard data filters and require firms to have at least one spell of five consecutive firm-year observations.[12] We select these variables because they are commonly used as either the variable of interest or as controls (Frank and Goyal, 2009; Mitton, Forthcoming). Though we primarily focus on firm fixed effects, we include results for industry fixed effects as well to illustrate that it is not ex ante clear how much within-group variation exists given the level of fixed effects. Similar to our replication analysis, several of these variables have a within-firm standard deviation that is 50-70% smaller than the sample standard deviation and present potential issues if the researcher uses the sample standard deviation as a reasonable counterfactual change in a firm's financial and accounting variables.[13] For instance, using the sample standard deviation of $log(Assets)$ to calculate the economic magnitude gives an effect almost three times larger than the economic magnitude calculated using the within-firm standard deviation. Moreover, comparing columns (4) to (6) with columns (7) to (9), the differences in standard deviation between the sample versus within-firm persist across time periods and generally increase for shorter time periods. Thus, this concern is likely larger for papers that use shorter sample periods.[14]

In Figure 1, we report the W-S Ratios for the fourteen variables from Table 3. Notably, the W-S Ratios vary greatly across variables even when we apply a common set of fixed effects, and the ratios are arguably hard to predict ex ante. For instance, $log(Sales)$ and $log(Assets)$

[12] We require firms to have five consecutive firm-year observations to get a reasonable measure of within-firm standard deviation. Results are similar if we remove this filter.

[13] The same reasoning and issues exist if researchers instead use the interquartile range as their metric of variation.

[14] We also note that increasing the dimensionality of the fixed effects will further decrease the within-group variation. See Table A5 in the Appendix.

have a relatively low ratio implying that the within-firm variation is substantially smaller than the overall variation, and thus most of the variation is due to between firm differences (i.e., cross-sectional variation). However, ROE and Book Leverage have a relatively high ratio implying that most of the variation is coming from within-firm differences (i.e., time-series variation). Therefore, comparing the relative ranking of the economic magnitudes of two variables based on the scaling of their fixed-effects regression coefficients using the sample standard deviation could yield misleading inferences without making the stronger assumption that the between-firm effect of the variable is similar to the within-firm effect.

The heterogeneity across variables in the W-S ratio also affects the Mitton (Forthcoming) measure of economic significance – multiplying the coefficient by the standard deviation of the variable of interest and dividing by the standard deviation of the dependent variable. One of the purposes of this measure is to allow for comparisons of different variables across similar regressions. However, the use of OLS fixed-effects regressions complicates this measure when there is heterogeneity in the identifying variation across variables. For example, Figure 1 shows that the W-S ratio of $log(Assets)$ is less than one-half that of other common variables like Tobin's Q or leverage. When comparing variables that are commonly used to explain profitability, the economic significance of $log(Assets)$ calculated using the sample standard deviation is more than twice as large as the economic significance of leverage and roughly 15% larger than the economic significance of Tobin's Q (see Mitton (Forthcoming) Table 5, column 7). However, repeating the same calculations using within-firm standard deviations would imply that the economic significance of Book Leverage and $log(Assets)$ are approximately equal (0.25) and that Tobin's Q is more economically significant than $log(Assets)$ (0.43).[15] Thus, the relative rankings are re-ordered after matching the variation of the economic significance estimates to the identifying variation of the fixed-effects regression. Finally, even if the economic significance calculations that divide by the sample standard deviation of the outcome variable yield similar results as the calculations that divide by within-group standard deviations, the within-group standard deviation still needs to be reported and discussed to provide context for the unitless measure.

An alternative view of a reasonable counterfactual change in a variable of interest is to

[15] This is calculated by multiplying the E_s^s measure for $log(Assets)$, Book Leverage, and Tobin's Q from Mitton (Forthcoming), Table 5, column 7 for profitability (ROA), by the variable's W-S ratio and dividing by the W-S ratio of profitability.

consider the frequency at which a given move occurs for any firm. One way to gauge this definition of reasonableness is to determine the proportion of observations for each firm that are more than a given difference from other firm-specific observations or the number of firm-year observations that experience a given move in the variable of interest over a fixed amount of time. For each variable in Table 3, we examine the proportion of observations that are more than one standard deviation away from any other observed values of the variable for a given firm. As a benchmark, we can compare the proportions to the normal distribution benchmark of 48%.

In Table 4 column (1), we report the percentage of observations that are more than one sample standard deviation away for a given observation of a given firm. Specifically, we link each observation to all other observations for the firm and calculate the proportion of these observation pairs that differ by at least one sample standard deviation. All variables have a smaller likelihood than would be expected under the normal distribution, and for several variables the likelihood is very low (e.g., less than 10%). This indicates that for these variables a sample-standard-deviation-sized change is very rare for most firm-level observations. In column (4), we report the same metric for the within-firm standard deviation. While observations of the same firm rarely differ by more than one sample standard deviation, the likelihood of a within-firm change greater or equal to the within-firm standard deviation is much closer to the normal distribution benchmark.

Another measure of within-firm changes is to examine the changes for adjacent or consecutive years. In Table 4 columns (2) and (5), for each variable, we report the percentage adjacent-year observations that differ by at least one sample standard deviation or one within-firm standard deviation, respectively. In columns (3) and (6), we report the percentage of firm-years that have a 10-year peak-to-trough difference in the variable of interest of at least one sample standard deviation or one within-firm standard deviation, respectively. For 13 of the 14 variables, less than 10% of their adjacent-year changes are greater than or equal to the sample standard deviation. Moreover, for the majority of variables, less than 25% of their 10-year peak-to-trough differences are greater than the sample standard deviation. For instance, only 22% of observations have a 10-year peak-to-trough difference in Tobin's Q as large as the sample standard deviation. Therefore, it is a strong assumption that a change as large as the sample standard deviation happens under conditions of all else equal.

Note that even within-firm movements greater than the within-firm standard deviation may still be less frequent than the normal distribution benchmark. There are two potential explanations for the lower frequency. First, the within-firm distribution of the variable is unlikely to be normally distributed. In Figure 2, we plot the within-firm distributions for Book Leverage, Tobin's Q, ROA, and $log(Sales)$ relative to their demeaned sample distributions. As established above, the within-firm standard deviation is lower than the sample standard deviation. While the within-firm distributions are closer to a normal distribution, there is still significant skewness and kurtosis. Even so, this does not mean the within-firm standard deviation is an incorrect measure of variation, but it is one explanation for why the frequency of changes is lower than the normal distribution benchmark.

Second, the within-firm standard deviation is an aggregate measure for the entire sample that may not be representative of the changes for any given firm. In contrast to Figure 1, Figure 3 illustrates the potential for heterogeneity across groups in the degree of within-firm variation. Specifically, for Book Leverage, Tobin's Q, ROA, and $log(Sales)$, we first group firms into decile bins according to the firm-level mean of the variable. Next, we plot the mean for the given bin plus or minus the average within-firm standard deviation for that bin. This figure captures the degree that different groups of firms vary around their firm-level means.[16] Thus, it is possible that high within-firm variation for some firms could obfuscate that many firms have only relatively small movements in the given financial variables.[17]

The main takeaways from this section are that common finance variables in a standardized setting often have within-firm variation that is less than the sample standard deviation and that there is significant heterogeneity in within-firm variation across variables. Thus, to determine the economic significance of the effect, the researcher should carefully consider the counterfactual they are trying to represent and provide a defense of whether it is reasonable.

[16] Solon, Haider, and Wooldridge (2015) and Gibbons, Suárez Serrato, and Urbancic (2019) examine how the fixed-effect coefficient is affected by heterogeneity in both within-group treatment effects and within-group variation in the variable of interest. From Figure 3, there are substantial differences in within-firm variation across groups, thus the interpretation of the OLS or 2SLS coefficient as an average partial effect for the population crucially depends on the homogeneity assumption. To help alleviate these concerns, when there is significant heterogeneity in within-firm variation, researchers should consider estimating the average treatment effect using the method of Gibbons, Suárez Serrato, and Urbancic (2019) and plot the statistical leverage of their observations across the distribution of their variable of interest or instrument.

[17] An alternative measure is to use the median of the firm-level standard deviation in the variable of interest. As we show in Table A6 in the Appendix, using this measure comes much closer to the normal distribution benchmark.

3 Recommendations

The goal of economic magnitude calculations is to understand how a reasonable change in the variable of interest will impact the outcome variable. There is no standard definition of a reasonable change in the variable of interest because the definition of a reasonable change depends on the specific application. Given that fixed-effects regressions are the workhorse model in finance research, we focus our recommendations on these applications.

The implicit assumption behind fixed-effects regressions is that the within-group variation in the variable is exogenous, while the between-group variation in the variable may be contaminated by other unmodeled factors. Thus, to gauge the amount of within-group variation, we recommend that researchers do the following:

1. Discuss the identifying variation and defend the choice of standardization or scaling.

2. Report the within-group standard deviation of the variable of interest.

3. Summarize the frequency of within-group changes of the variable of interest.

First, researchers should discuss the sources of their identifying variation. While fixed-effects regressions are useful in removing the effects of invariant group characteristics, this specification choice also constrains the variation that is used to identify the regression coefficient. When interpreting the regression coefficient, the researcher should be explicit about their choice of standardization or scaling and explain why this represents a reasonable counterfactual or generates a useful comparison across variables.[18]

Second, as we show in Section 2.2, the within-group standard deviation of finance variables is often much smaller than the sample standard deviation. Since the within-group variation is the variation used to identify the regression coefficient, researchers should report a statistic that summarizes this variation. To do so, the researcher should regress the variable of interest on all relevant fixed effects. Then, the researcher can calculate the within-group standard deviation by taking the standard deviation of the residuals from this regression.[19]

[18] Our critique also has potential implications for structural models where the goal of the model is to match moments, such as variance or standard deviation, from the underlying data. The researcher should be careful to match within-firm standard deviations or variances if they demean outcome variables to account for differences across firms, as suggested by Strebulaev and Whited (2017).

[19] This can be easily done for instance with the add-on package to the popular reghdfe package in Stata, from deHaan (2021) available here: https://github.com/ed-dehaan/sumhdfe. Another alternative is to directly regress the variable of interest on the fixed effects, and then examine the variation of the residuals.

Additionally, to visually asses the magnitude of the identifying variation relative to the sample and the appropriateness of using a standard deviation metric, the researcher can plot the within-group residuals similar to Figure 2.

Third, the researcher can report statistics about the empirical distribution of within-group changes. The researcher can either calculate the difference between each observation and all other observations in the group for the variable of interest or examine within-firm changes over a given time-period.[20] In addition, the researcher can plot the within-group distribution of the variable of the interest, similar to Figure 3. This may be revealing about the extrapolability of the results to the larger population. In general, summarizing the within-group variation helps validate the reasonableness of the counterfactual change.

There are a few alternatives to the within-group standard deviation that researchers can consider as well. While most of our statistical intuitions are based on symmetric or normal distributions, not all data is generated from these distributions. If the researcher is worried that the variable of interest is skewed, the researcher can instead report the interquartile range of the above residuals. Additionally, if there is significant heterogeneity in within-group standard deviations, the researcher can also report the median group's within-group standard deviation.

Finally, researchers often use different sets of fixed effects to isolate the identifying variation for the variable of interest. When dealing with multiple regression specifications, it is hard to provide a one-size-fits-all approach because the within-group variation depends on the specification. In general, the most conservative approach is to calculate the identifying variation using the most stringent set of fixed effects. If the coefficient of interest does not change significantly among the specifications, then either the within-group variation that is accounted for by additional fixed effects is not particularly important or the between- and within-group effects are similar. Thus, in these situations, the researcher can use the less stringent set of fixed effects to estimate within-group variation for economic magnitude calculations.

[20] If there are multiple levels of fixed effects (e.g., firm and year), the researcher should carefully define the counterfactual exercise (e.g., within-firm) and verify that the residuals vary in that dimension.

4 Conclusion

Finance researchers often use a sample standard deviation change in the variable of interest to gauge the economic significance of their regression coefficients. This method is often problematic when gauging the economic significance of fixed-effects regression coefficients unless stronger assumptions are made. As we demonstrate in many settings, this common approach can significantly overstate the economic magnitudes and change the relative importance of these variables. We suggest that researchers use the within-group standard deviation, or other measures of the identifying within-group variation, of the variable of interest to calculate the economic magnitudes of their regression estimates and be more explicit in their justification of the counterfactual according to their specific applications.

References

Angrist, Joshua D and Guido W. Imbens, 1994, Identification and estimation of local average treatment effects, *Econometrica* 62, 467–475.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion* (Princeton university press 2009).

Arellano, Manuel and Bo Honoré, Panel data models: some recent developments, *Handbook of econometrics*, volume 5, 3229–3296 (Elsevier 2001).

Babina, Tania, 2020, Destructive creation at work: How financial distress spurs entrepreneurship, *The review of financial studies* 33, 4061–4101.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang, 2022, How much should we trust staggered difference-in-differences estimates?, *Journal of Financial Economics* 144, 370–395.

Begley, Taylor A and Amiyatosh Purnanandam, 2021, Color and credit: Race, regulation, and the quality of financial services, *Journal of Financial Economics* 141, 48–65.

Berger, David, Nicholas Turner, and Eric Zwick, 2020, Stimulating housing markets, *The Journal of Finance* 75, 277–321.

Callaway, Brantly and Pedro H.C. Sant'Anna, 2021, Difference-in-differences with multiple time periods, *Journal of Econometrics* 225, 200–230, themed Issue: Treatment Effect 1.

de Chaisemartin, Clément and Xavier D'Haultfoeuille, 2020, Two-way fixed effects estimators with heterogeneous treatment effects, *American Economic Review* 110, 2964–96.

Chen, Tao, Hui Dong, and Chen Lin, 2020, Institutional shareholders and corporate social responsibility, *Journal of Financial Economics* 135, 483–504.

Choi, Darwin, Zhenyu Gao, and Wenxi Jiang, 2020, Attention to global warming, *The Review of Financial Studies* 33, 1112–1145.

Cohn, Jonathan B., Zack Liu, and Malcolm I. Wardlaw, 2022, Count (and count-like) data in finance, *Journal of Financial Economics* 146, 529–551.

Coles, Jeffrey L and Zhichuan F Li, 2023, An empirical assessment of empirical corporate finance, *Journal of Financial and Quantitative Analysis* 58, 1391–1430.

Deaton, Angus, 1985, Panel data from time series of cross-sections, *Journal of Econometrics* 30, 109–126.

deHaan, Ed, 2021, Using and interpreting fixed effects models, *Available at SSRN 3699777* .

Frank, Murray and Vidhan K. Goyal, 2009, Capital structure decisions: Which factors are reliably important?, *Financial Management* 38, 1–37.

Gibbons, Charles E., Juan Carlos Suárez Serrato, and Michael B. Urbancic, 2019, Broken or fixed effects?, *Journal Econometric Methods* 8.

Goodman-Bacon, Andrew, 2021, Difference-in-differences with variation in treatment timing, *Journal of Econometrics* 225, 254–277, themed Issue: Treatment Effect 1.

Graham, John R, Si Li, and Jiaping Qiu, 2012, Managerial attributes and executive compensation, *Review of Financial Studies* 25, 144–186.

Grieser, William D. and Charles J. Hadlock, 2019, Panel-data estimation in finance: Testable assumptions and parameter (in)consistency, *Journal of Financial and Quantitative Analysis* 54, 1–29.

Griffin, John M, Samuel Kruger, and Gonzalo Maturana, 2021, What drove the 2003–2006 house price boom and subsequent collapse? disentangling competing explanations, *Journal of Financial Economics* 141, 1007–1035.

Heath, Davidson and Christopher Mace, 2020, The strategic effects of trademark protection, *The Review of Financial Studies* 33, 1848–1877.

Heath, Davidson, Matthew C. Riggenberg, Mehdrad Samadi, and Ingrid M. Werner, Forthcoming, Reusing natural experiments, *Journal of Finance* .

Huang, Yi, Marco Pagano, and Ugo Panizza, 2020, Local crowding-out in china, *The Journal of Finance* 75, 2855–2898.

Kalmenovitz, Joseph, 2021, Incentivizing financial regulators, *The Review of Financial Studies* 34, 4745–4784.

King, Gary and Langche Zeng, 2006, The dangers of extreme counterfactuals, *Political Analysis* 14, 131–159.

Lemmon, Michael L, Michael R Roberts, and Jaime F Zender, 2008, Back to the beginning: persistence and the cross-section of corporate capital structure, *Journal of Finance* 63, 1575–1608.

Levine, Ross, Chen Lin, Mingzhu Tai, and Wensi Xie, 2021, How Did Depositors Respond to COVID-19?, *Review of Financial Studies* 34, 5438–5473.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, 2021, Measuring corporate culture using machine learning, *The Review of Financial Studies* 34, 3265–3315.

Liao, Gordon Y, 2020, Credit migration and covered interest rate parity, *Journal of Financial Economics* 138, 504–525.

McCloskey, Deirdre N and Stephen T Ziliak, 1996, The standard error of regressions, *Journal of Economic Literature* 34, 97–114.

Mitton, Todd, 2022, Methodological variation in empirical corporate finance, *Review of Financial Studies* 35, 527–575.

Mitton, Todd, Forthcoming, Economic significance in corporate finance, *Review of Corporate Finance Studies* .

Mummolo, Jonathan and Erik Peterson, 2018, Improving the interpretation of fixed effects regression results, *Political Science Research and Methods* 6, 829–835.

Roberts, Michael R and Toni M Whited, Endogeneity in empirical corporate finance1, *Handbook of the Economics of Finance*, volume 2, 493–572 (Elsevier 2013).

Saidi, Farzad and Daniel Streitz, 2021, Bank concentration and product market competition, *The Review of Financial Studies* 34, 4999–5035.

Solon, Gary, Steven J Haider, and Jeffrey M. Wooldridge, 2015, What are we weighting for?, *Journal of Human Resources* 50, 301–316.

Strebulaev, Ilya A. and Toni M. Whited, 2017, Dynamic models and structural estimation in corporate finance, *Available at SSRN 2091854* .

Wooldridge, Jeffrey M., *Econometric analysis of cross section and panel data* (MIT press 2010).

Zandberg, Jonathan, 2021, Family comes first: Reproductive health and the gender gap in entrepreneurship, *Journal of Financial Economics* 140, 838–864.

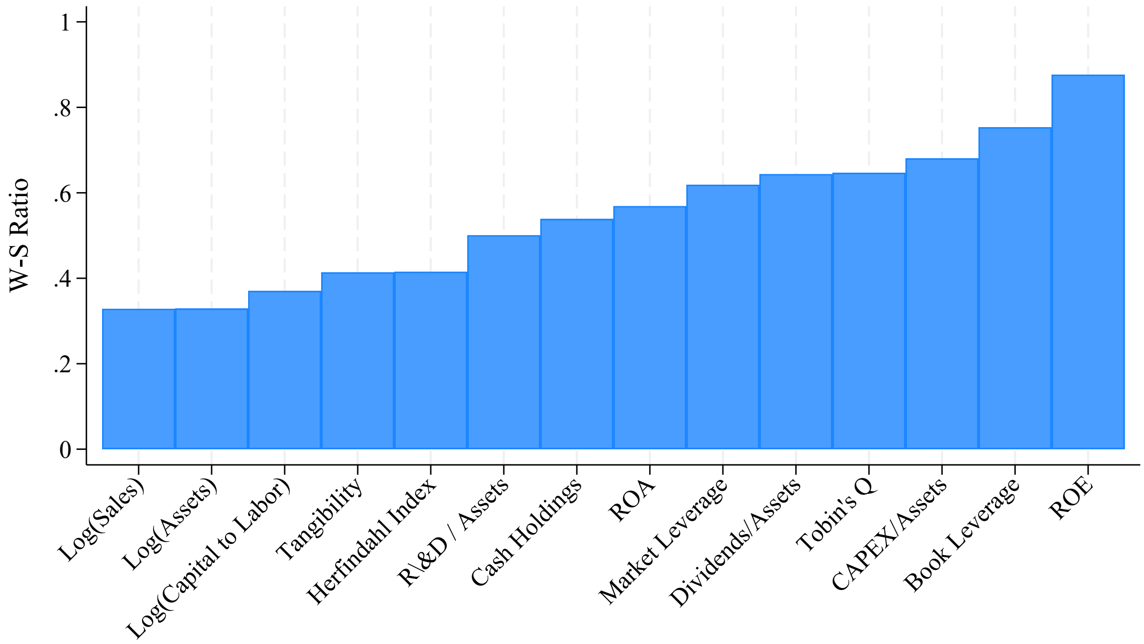

Figure 1: Within-firm to sample standard deviations This figure reports the ratio of the within-firm standard deviation to the sample standard deviation (W-S Ratio) for the variables in Table 3. The bars represent the ratio and the variables are sorted in ascending order. The main takeaway is that variables have wide heterogeneity in their W-S Ratio, but that the majority of variables have a within-firm standard deviation that is significantly smaller than their sample standard deviation.

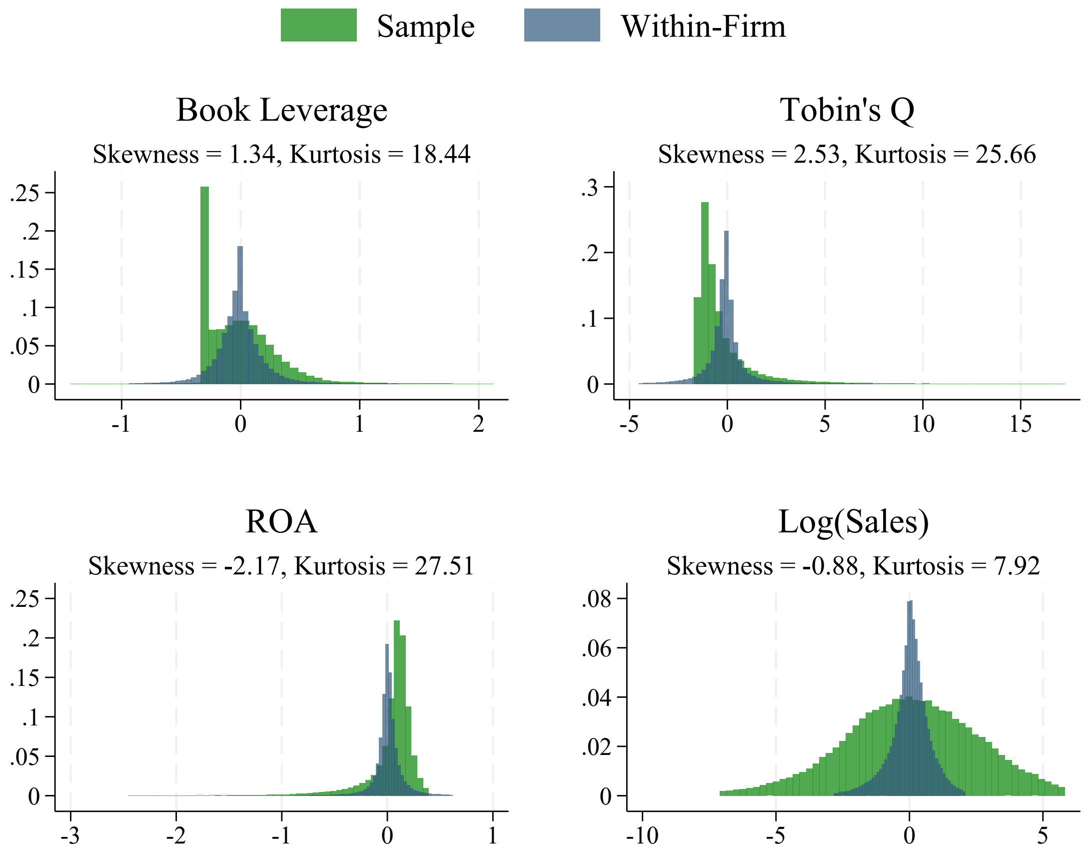

Figure 2: Sample vs. within-firm variation This figure reports the sample and within-firm distributions of the given variables. The green (blue) bars plot the sample (within-firm) distribution after subtracting the sample (firm) mean from each observation. Each bar represents the frequency that observations fall into the respective bin. The reported skewness and kurtosis are for the within-firm distributions. For graphing purposes, the first and ninety-ninth percentiles are excluded. The key takeaway of this figure is that the within-firm distributions have less variation than the sample distributions and that the within-firm distributions are relatively more normally distributed, as indicated by the skewness and kurtosis statistics.

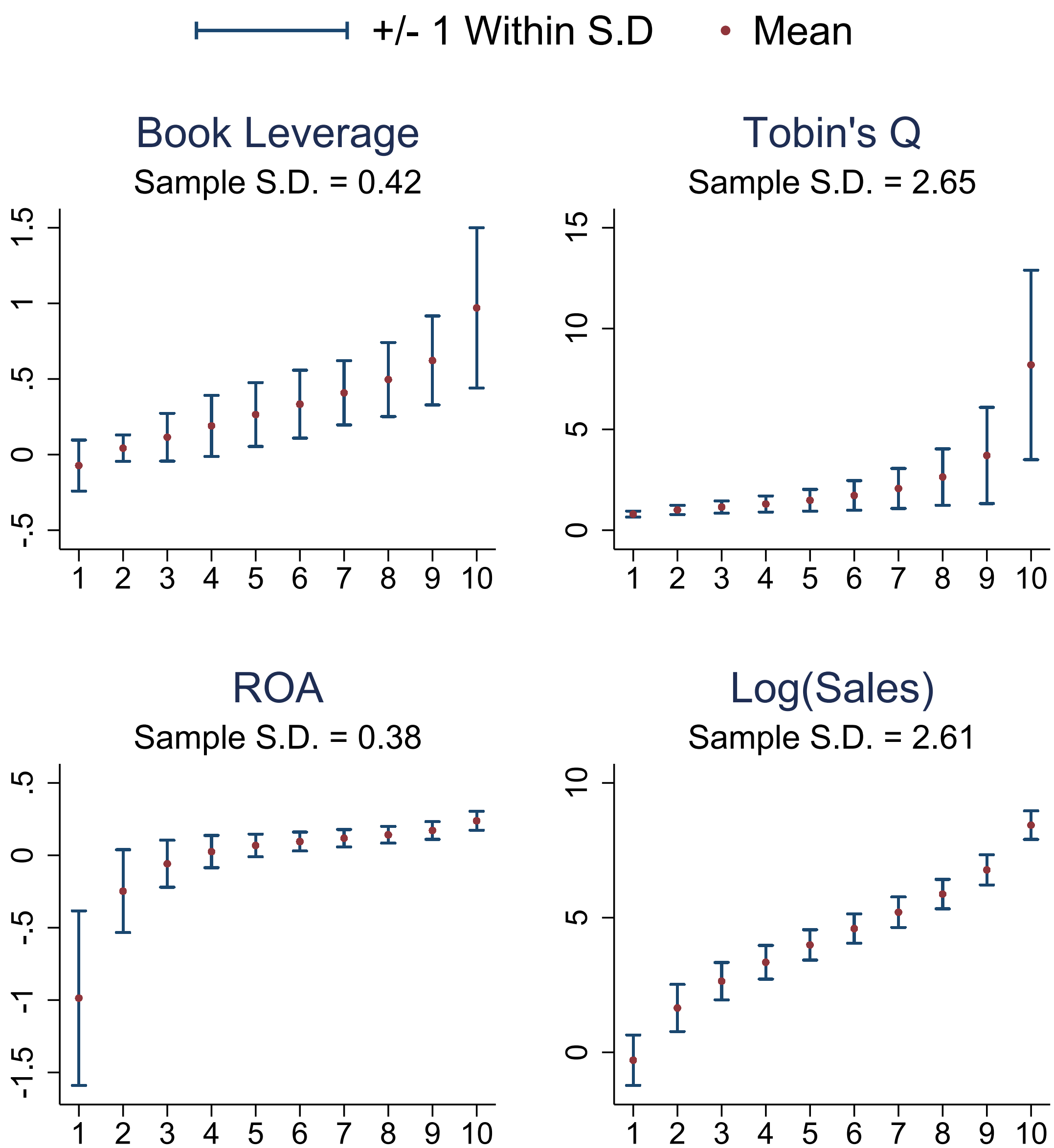

Figure 3: Within-firm variation across distribution This figure reports the within-firm variation for the decile groups of the given variables. The red dots plots average firm-level mean for each decile across firms of the given variable. The bars widths represent the average within-firm standard deviation in the given decile bin. The key takeaway is that for these variables there is a large degree of heterogeneity across firms in their degree of within-firm variation of the given variable.

Table 1: 2020 & 2021 publication statistics This table presents statistics on the prevalence of the common economic magnitude calculation in papers published by JF, JFE, and RFS in 2020 and 2021. This table reports the total number of published articles (column 1), the number of articles that use some version of panel fixed effects (column 2), and the number and proportion of articles that reference the effect of a one standard deviation move in a variable of interest conditional on using panel fixed effects (columns 3 and 4). This table highlights the common use of scaling by the standard deviation when using panel regressions with fixed effects.

Journal	Total Count	Papers with FE Regressions		
		Count	1-S.D. E.M.	Pct
	(1)	(2)	(3)	(4)
Journal of Finance	144	69	27	39.1%
Journal of Financial Economics	359	230	107	46.5%
Review of Financial Studies	279	172	83	48.3%
Overall	782	471	217	46.1%

Table 2: Within-standard deviation ratio for replicated papers This table presents the ratio of within-group standard deviation to sample standard deviation (W-S Ratio) for replicated papers published in the top 3 finance journals. For each paper, we use the author provided replication package or follow the data cleaning procedures described in the paper. † indicates papers with replication packages. Using these datasets, we report the sample standard deviation and the within-group standard deviation. The groups are determined by the original paper's fixed effect specification decisions. We report the variable of interest, the fixed effects, the sample standard deviation, the within-group standard deviation, the W-S Ratio, the economic magnitue given in the paper, and the adjusted economic magnitude based on the W-S Ratio. *** indicates W-S ratios below 50%, ** indicates 66%, and * indicates 75%. This table highlights that for many of the replicated papers the reported economic magnitudes can be substantially larger than it would be if estimated from the identifying variation.

	Citation	Variable of Interest	FE	Sample S.D. (1)	Within S.D. (2)	W-S Ratio (3)	E.M. rep. (4)	E.M. adj. (5)
RFS	Babina (2020)	Book Leverage	Firm	0.249	0.138	55.4%**	21%	11.6%
JFE	Begley and Purnanandam (2021)	log(Adjusted Gross Income)	Mortgage Bin, Population Bin, MSA	0.458	0.2896	63.2%**	11%	7.0%
JF	Berger, Turner, and Zwick (2020)†	First-Time Buyers/Total Individuals	CBSA	0.009984	0.008769	87.8%	2.6%	2.3%
JFE	Chen, Dong, and Lin (2020)	Distraction	Firm, Industry x Year	0.0457	0.0078	17.1%***	0.404	0.069
RFS	Choi, Gao, and Jiang (2020)†	Abnormal Temperature	Month	2.70	2.54	94.1%	16	15.1
JFE	Griffin, Kruger, and Maturana (2021)†	Subprime Share	MSA	0.058	0.05	86.2%	-3.7%	-3.2%
		Noncore Deposits		0.060	0.032	53.3%**	-3.2%	-1.7%
		Worse Originator		0.016	0.012	75.0%*	-3.6%	-2.7%
		Private Securitization		0.083	0.051	61.4%**	-2.4%	-1.5%
		Non-Owner Occupancy		0.091	0.07	76.9%	-0.4%	-0.3%
		Out-of-Town		0.079	0.044	55.7%**	-1.0%	-0.6%
		House price growth		0.067	0.041	61.2%**	-0.6%	-0.4%
RFS	Heath and Mace (2020)†	Famous Trademarks/Total	Firm, Industry x Year	0.7073	0.5001	70.7%*	43.0%	30.4%
JF	Huang, Pagano, and Panizza (2020)†	Local Debt to GDP	City, Year	13.70	5.199	37.9%***	1.1%	0.42%
RFS	Kalmenovitz (2021)†	Incentives	Year x Office x Grade	0.1193	0.0898	75.3%	16.7%	12.6%
RFS	Levine et al. (2021)†	log(COVID/Population)	Branch, State x Week, Bank x Week, Date	2.235	0.548	24.5%***	5.2%	1.3%
RFS	Li et al. (2021)†	Integrity	Industry, Year	1.325	1.235	93.2%	2.3%	2.2%
		Innovation		2.757	2.207	80.1%	0.7%	0.5%
		Respect		2.184	1.874	85.8%	2.4%	2.0%
		Quality		1.562	1.312	84.0%	1.1%	0.9%
JFE	Liao (2020)†	CIP basis	Currency, Quarter	15.01	11.46	76.3%	4.0%	3.1%
RFS	Saidi and Streitz (2021)	Bank-Industry HHI	Industry, Year	0.1737	0.1256	72.3%*	2.2%	1.6%
JFE	Zandberg (2021)	Abortion Ratio	State, Age, Industry, Year	0.075	0.033	44.0%***	5.9%	2.6%

Table 3: Sample and within variation in common COMPUSTAT variables This table presents the sample standard deviation, the within-industry (SIC 3-digit) standard deviation, and within-firm (GVKEY) standard deviation of the given variables and time periods. The sample period is from 1975-2018 and excludes financial firms and utilities. Additionally, we exclude firm-years with negative sales and firm-years missing one of the given variables and require firms to have at least one spell of five consecutive firm-year observations. All variables are winsorized at the 1% and 99% levels. All variables are defined in Appendix Table A2. This table highlights that for many common finance variables the within-firm standard deviation (column 3) is significantly smaller than the sample standard deviation. Thus, using the sample standard deviation for economic magnitude calculations overstates the effect relative to the identifying within-firm variation.

	1975-2018			1975-1996			1997-2018		
Variable	Sample (1)	Ind. (2)	Firm (3)	Sample (4)	Ind. (5)	Firm. (6)	Sample (7)	Ind. (8)	Firm (9)
Log(Assets)	2.422	2.227	0.797	2.104	1.879	0.578	2.485	2.252	0.621
Log(Sales)	2.614	2.341	0.859	2.326	2.029	0.623	2.733	2.398	0.724
Book Leverage	0.419	0.407	0.316	0.340	0.330	0.232	0.463	0.447	0.336
Market Leverage	0.245	0.225	0.152	0.245	0.226	0.135	0.242	0.220	0.142
Tangibility	0.246	0.172	0.102	0.230	0.164	0.086	0.254	0.172	0.096
Log(Capital to Labor)	1.634	1.133	0.605	1.387	0.977	0.507	1.719	1.120	0.533
CAPEX/Assets	0.073	0.063	0.049	0.075	0.067	0.050	0.070	0.058	0.044
ROA	0.381	0.363	0.217	0.244	0.235	0.143	0.443	0.421	0.235
ROE	0.956	0.943	0.838	0.783	0.774	0.645	1.052	1.037	0.910
Dividends/Assets	0.019	0.018	0.012	0.018	0.017	0.009	0.020	0.019	0.013
Cash Holdings	0.211	0.182	0.114	0.173	0.159	0.092	0.228	0.191	0.112
R&D / Assets	0.122	0.102	0.061	0.083	0.070	0.038	0.140	0.117	0.066
Tobin's Q	2.653	2.530	1.716	1.913	1.804	1.164	3.011	2.886	1.849
Herfindahl Index	0.142	0.081	0.059	0.132	0.069	0.052	0.146	0.059	0.047

Table 4: Frequency of firm-level moves of common COMPUSTAT variables This table reports the frequency of changes that occur for the common finance variables in Table 3. Column 1 reports the proportion of observations for a given firm that differs from any other observation for that firm by at least one sample standard deviation, column 2 reports the proportion of adjacent year observations for a given firm that differ by at least one sample standard deviation, and column 3 reports the frequency that the 10 year peak-to-trough difference for a firm is at least one sample standard deviation. Columns 4-6 repeat the analysis using within-firm standard deviation of the variables. The sample period is from 1975-2018 and excludes financial firms and utilities. Additionally, we exclude firm-years with negative sales and firm-years missing one of the given variables, and require firms to have at least one spell of five consecutive firm-year observations. All variables are defined in Appendix Table A2. This table highlights that common COMPUSTAT variables rarely have movements within firm that are the magnitude of the variables' sample standard deviation, indicating that using the sample standard deviation is relatively extreme relative to the identifying variation when using firm fixed effects.

	Sample S.D.			Within-Firm S.D.		
Variable	Prop. (1)	1-year (2)	10-year (3)	Prop. (4)	1-year (5)	10-year (6)
Log(Assets)	0.055	0.001	0.056	0.315	0.043	0.466
Log(Sales)	0.054	0.005	0.075	0.295	0.050	0.450
Book Leverage	0.137	0.058	0.259	0.197	0.083	0.370
Market Leverage	0.200	0.075	0.430	0.322	0.166	0.592
Tangibility	0.083	0.025	0.201	0.306	0.127	0.593
Log(Capital to Labor)	0.069	0.012	0.120	0.306	0.079	0.507
CAPEX/Assets	0.165	0.114	0.421	0.242	0.173	0.551
ROA	0.075	0.048	0.171	0.149	0.094	0.328
ROE	0.121	0.091	0.230	0.134	0.099	0.252
Dividends/Assets	0.086	0.033	0.151	0.134	0.050	0.219
Cash Holdings	0.139	0.063	0.343	0.276	0.161	0.558
R&D / Assets	0.062	0.040	0.123	0.110	0.073	0.195
Tobin's Q	0.092	0.055	0.224	0.147	0.092	0.335
Herfindahl Index	0.079	0.009	0.111	0.212	0.043	0.313

Appendix

1 Replication Summary

We replicate 14 papers for this study. For 9 of these paper, the replication package was available. For the other 5 papers, we follow the data collection and cleaning procedures outlined in the paper. Table A1 reports the number of observations, mean, and standard deviation of the variable of interest in the published paper and based on our replications.

2 Variable Definitions

We provide the definitions for the variables in Table 3 in Table A2.

3 Simulations

3.1. Simulations with linear relation

In order to demonstrate the intuition for our results, we run several simulations to illustrate that OLS fixed-effect regressions identify a within-group effect and that using sample standard deviation can overstate the effect of a reasonable change in the variable of interest.

We simulate panels of observations (x, y). Each simulated data set is 5000 observations with 50 groups (e.g., industries), and 100 observations per group (e.g., firm-years). For each observation i, t, where i denotes firm, g denotes a group, and t denotes time, we draw two random variables $a_{i,t}$ and $b_{i,g}$, each from an independent mean-0 normal distribution. We vary the standard deviation of a, the within-group variation, and the standard deviation of b, the between-group variation, across the different simulations. We then define

$$x_{i,t} = a_{i,t} + b_{i,g}. \tag{3}$$

Finally, we define y as

$$y_{i,t} = \beta_{within} * a_{i,t} + \beta_{between} * b_{i,g} + \epsilon_{i,t}, \tag{4}$$

where ϵ is i.i.d. standard normal. β_{within} and $\beta_{between}$ are defined for each simulation. This setup produces a group fixed effect in $x_{i,t}$ and allows us to evaluate how the within- and between-variation affect OLS regression coefficients with and without group fixed effects.

Table A3 Panel A reports the parameters for each simulation. We simulate 1000 datasets for each set of parameters. First, we run the regression with no fixed effects. Second, we run the regression with group fixed effects. Third, we standardize x using the sample mean and standard deviation, and then run OLS regressions with group fixed effects. Table A3 Panel B reports the average OLS coefficient estimates of the relation between x and y over the 1000 simulations. Table A3 Panel C reports the average sample standard deviation and within-group standard deviation of x and y. To calculate the within standard deviation, we demean the variable within each group, then take the standard deviation.[21] Table A3 Panel D reports the likelihood of a within-group change for x that is at least as large as either the sample or within-group standard deviation.

In simulation 1, the simulation parameters are $\beta_{within} = 1$, $\beta_{between} = 0.5$, $std(a) = 1$, and $std(b) = 3$. This baseline simulation shows several things. First, as expected, the fixed effects regression recovers an unbiased estimate of β_{within}. Second, given that the data is generated with a within standard deviation of one for x and $\beta_{within} = 1$, the economic magnitude of a one standard deviation increase in x within the group should be a one-unit increase in y. However, if we were to use the sample standard deviation of x of 3.123, we would incorrectly interpret that a one standard deviation increase in x within the group would be a roughly three-unit increase in y. Additionally, the coefficient on the standardized x regression would imply a similar three-unit increase in y, which is a rare event that occurs only 3% of the time. The three-unit economic magnitude is too large because the sample standard deviation is driven largely by between-group differences and β_{within} identified by the OLS fixed-effects regression is not equal to $\beta_{between}$.

Simulation 2 differs from simulation 1 in that $std(a) = 3$ and $std(b) = 1$. In this simulation, most of the variation in x comes from within the group. In this case, the sample standard deviation is very similar but not exactly the same as the within standard deviation. Thus, the economic magnitudes are similar if one uses the within-group standard deviation or the sample standard deviation. Additionally, in this case, the coefficients on the regres-

[21] The reported standard deviations are slightly lower than the input parameters because there is no finite sample adjustment.

sion with and without fixed effects are similar as well because there is less between-group variation.

Figure A1 plots the within-firm standard deviation in x for different deciles based on firm means of x. The left panel is a dataset generated using the simulation 1 parameters and the right panel is a dataset generated using the simulation 2 parameters. As we noted above for simulation 1, there is a large difference between the sample standard deviation and the within-group standard deviation of x. If you take the mean observation in the 6th decile bin and add (subtract) one sample standard deviation from it, the observation would likely change to the 9th (1st) decile bin rather than being contained within the 6th bin. In simulation 2, where most of the variation is within the group, the sample standard deviation and within standard deviation are similar. If we take the mean observation in the 6th decile and add or subtract one sample standard deviation from it, it will still fall within the bars of the 6th bin.

Simulation 3 differs from simulation 1 in that $\beta_{within} = 2$. In this simulation, the OLS fixed-effects regression recovers an unbiased estimate of β_{within} and that using the sample standard deviation and standardizing x give the same, incorrect interpretation. Simulation 4 differs from simulation 1 in that $\beta_{between} = -1.5$. In simulation 4, we see that fixed-effect regression recovers an unbiased estimate of β_{within} while the coefficient from the regression without fixed effects is impacted both by the between-group effect and the relative variation within and between groups. Simulation 5 differs from simulation 1 in that $std(b) = 0$. In this case, all of the variation in x comes from the within-group component. In this corner case, using the sample standard deviation is equivalent to using the within-group standard deviation. However, in this case, fixed effects are unnecessary since the regressions with and without fixed effects will yield the same coefficient.

Note that in these simulations, the relation between a and b, the components of x, and y is linear. In reality, there is no guarantee of a linear relation, and the true form of the relation, especially that between the group component and the outcome, could be quadratic, exponential, or non-parametric. Even in situations where the group component has a non-linear relation to the outcome, the OLS fixed-effects regression coefficient will give the correct estimate if the within-group component has a linear, or locally-linear, relation to the outcome. We simulate this situation in Appendix Section A3.2. Again, for within-group changes in x, the OLS fixed-effects regression coefficient will provide a good guide for the counterfactual effect. However, for larger changes that likely affect the group-level component of x, the OLS fixed-effects estimator will not provide a good guide to estimate the potential change in the outcome.

The main takeaway from this section is that if there is a significant difference between within-group and sample standard deviation, then economic magnitudes for OLS fixed-effect coefficients calculated using sample standard deviation may be too large. Additionally, standardizing the variable of interest does not fix this issue.[22]

3.2. Simulation with exponential relation

In Section A3, we simulate data where there is a linear relation between the x and y. In reality, the relation between these variables could take on any form. In this simulation, we specify an exponential relation between the industry component of x, $b_{i,g}$ and y. Keeping everything else the same, we define y as

$$y_{i,t} = \beta_{within} * a_{i,t} + exp(\beta_{between} * b_{i,g}) + \epsilon_{i,t}. \tag{5}$$

Table A4 Panel A reports the parameters for each simulation. We simulate 1000 datasets for each set of parameters. First, we run the regression with no fixed effects. Second, we run the regression with group fixed effects. Third, we standardize x using the sample mean and standard deviation, and then run OLS regressions with group fixed effects. Table A4 Panel B reports the average OLS coefficient estimates of the relation between x and y over the 1000 simulations. Table A4 Panel C reports the average sample standard deviation and within-group standard deviation of x and y. To calculate the within standard deviation, we demean the variable within each group, then take the standard deviation.[23] Table A4 Panel D reports the likelihood of a within-group change for x that is at least as large as either the sample or within-group standard deviation.

In simulation 1, the simulation parameters are $\beta_{within} = 1$, $\beta_{between} = 0.5$, $std(a) = 1$, and

[22] Mitton (Forthcoming) suggests scaling the standardized coefficient by the standard deviation of the outcome variable (y), with the interpretation that it represents the percentage change in standard deviations of y for a one standard deviation change in x. Although this measure has several advantages, it still suffers from the same critique if the scaling is not done within. For instance, scaling by the sample standard deviation of y in simulations 1, 4, and 5 of Table A3, will result in different estimates of the economic magnitude of the effect of x. Scaling by the within standard deviation of y, however, will produce a consistent effect as expected.

[23] The reported standard deviations are slightly lower than the input parameters because there is no finite sample adjustment.

$std(b) = 3$. This baseline simulation shows several things. First, as expected, the fixed effects regression recovers an unbiased estimate of β_{within}. Second, given that the data is generated with a within standard deviation of one for x and $\beta_{within} = 1$, the economic magnitude of a one standard deviation increase in x within the group should be a one-unit increase in y. However, if we were to use the sample standard deviation of x of 3.123, we would incorrectly interpret that a one standard deviation increase in x within the group would be a roughly three-unit increase in y. Additionally, the coefficient on the standardized x regression would imply a similar three-unit increase in y. The three-unit economic magnitude is incorrect because the sample standard deviation is driven largely by between-group differences and β_{within} identified by the OLS fixed-effects regression is not equal to $\beta_{between}$.

Compared to the simulations in Section A3, we see that the OLS fixed-effects regression coefficients remain the same. This is because the group fixed effect addresses, on average, the group-level relations between x and y, even if the relation is non-linear. If we are interested in estimating the counterfactual impact of a small change in x, we can assume that all else is equal and use the OLS fixed-effects regression coefficient to estimate the effect. However, if we want to know the impact of a larger change, one that likely is linked to a change in the group-level dimension of x, then the OLS fixed-effects regression cannot confidently measure the likely change in the outcome.

4 Granular FE's

In Table A5, we recreate Table 3 allowing for different granularity of fixed effects. From Table A5, we can see that increasing the granularity of fixed effects reduces the within-group variation. Including industry $\times$ year fixed-effects further reduces the within-group variation for some variables by a relatively large percentage (e.g., $log(Assets)$ or $log(Sales)$), whereas for others it has a relatively small effect (e.g., ROE or Book Leverage). For variables that are defined at the industry $\times$ year level, these fixed effects mechanically absorb all variation (e.g., Herfindahl Index).

5 Median within-firm variation

In Table A6, we present estimates of the median within-firm standard deviation and the analogous results from Table 4. We calculate the median within-firm standard deviation by first estimating the standard deviation of the residuals of the given variable when regressed on firm-level fixed effects. We then calculate the median standard deviation across all firms that have at least five consecutive firm-year observations.

In Table A6, we can see that the median within-firm standard deviation is often significantly smaller than the within-firm standard deviation. Moreover, using the median within-firm standard deviation often produces counterfactual moves that are more aligned with the theoretical benchmark. This metric will generally be more aligned than the sample within-firm standard deviation, if the individual firm-level within standard deviations are highly skewed. For instance, from Figure 3 we can see that Tobin's Q is highly skewed with the highest decile of firms having the largest within-firm variation. Thus, for most firms the within-firm standard deviation is still a relatively large change relative to the variation that most firms have in the data, with only a 14.7% chance that any two firm-level observations having a difference more than a one within-firm standard deviation. However, by using the median within-firm standard deviation this probability increases to roughly 39.2%.

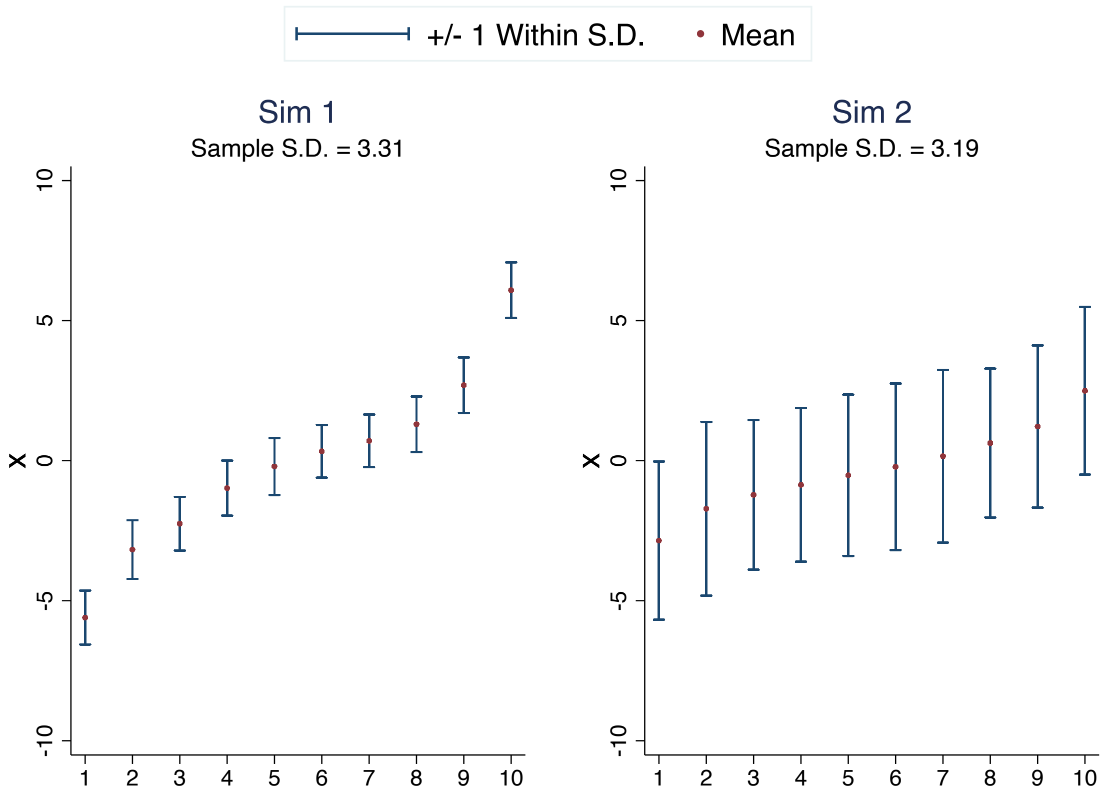

Figure A1: Within-firm variation by decile This figure reports the mean of x for each decile across firms (red dots). The bars represent the given mean plus/minus average within-firm standard deviation in the given decile bin. The data corresponds to simulation 1 and 2 of Table 3.

Table A1: Replication summary statistics This table presents summary statistics for the variable of interest. For each paper for which we replicate and did not have a replication package, we first report the summary statistics presented in the published version of the paper. Then, we present the summary statistics for our replications. We report the number of observations, mean, and standard deviation. Although there are slight differences in values, in general for each paper we are fairly close to the published summary statistics for the variable of interest.

Citation	Variable of Interest	Dataset	N (1)	Mean (2)	S.D. (3)
Babina (2020)	Book Leverage	Actual	20000	0.26	0.215
		Replication	20470	0.385	0.249
Chen, Dong, and Lin (2020)	Distraction	Actual	28020	0.163	0.045
		Replication	18140	0.160	0.046
Begley and Purnanandam (2021)	log(Adjusted Gross Income)	Actual	16309	10.93	0.44
		Replication	16387	10.93	0.46
Saidi and Streitz (2021)	Bank-Industry HHI	Actual	5250	0.25	0.18
		Replication	5293	0.272	0.17
Zandberg (2021)	Abortion Ratio	Actual	NA	19.3%	7.5%
		Replication	1805481	19.9%	7.7%

Table A2: Definitions of variables. The table contains the definitions of the variables in Table 3.

Variable	Definition
Log(Assets)	Natural log of total assets (`ln(at)`)
Log(Sales)	Natural log of sales (`ln(sale)`)
Book Leverage	Ratio of total debt (`dltt+dlc`) over total debt plus total book equity (`dltt + dlc + seq`)
Market Leverage	Ratio of total debt (`dltt+dlc`) over total debt plus preferred stock and market value of common equity (`dltt + dlc + pstk + csho*prcc_f`)
Tangibility	1 minus the ratio of current and intangible assets (`act + intan`) over total assets (`at`)
Log(Capital to Labor)	Natural log of the ratio of net property, plants, and equipment over employment (`ln(ppent / emp)`)
CAPEX / Assets	Capital expenditures (`capx`) over total assets (`at`)
ROA	Operating income (`oibdp`) over total assets (`at`)
ROE	Operating income (`oibdp`) over total common equity (`ceq`)
Dividends / Assets	Ratio of dividends (`dvc`) to total assets (`at`)
Cash Holdings	Ratio of cash (`ch`) plus short term investments (`ivst`) over total assets (`at`)
R&D / Assets	Ratio of R&D (`xrd`) over total assets (`at`)
Tobin's Q	Ratio of market value of common equity (`prcc_f*csho`) plus book value of debt (`at-ceq`) less deferred taxes (`deferred`) over total assets (`at`)
Herfindahl Index	Sum of squared market share $\left(\frac{sale}{\sum sale}\right)^2$ for each SIC3 code and fiscal year

Table A3: Simulation results This table presents the simulation results. Panel A gives the input parameters for each set of simulations. Panel B reports the average coefficient estimates for OLS with no fixed effects, OLS with fixed effects, and OLS with fixed effects and standardized x. Panel C reports the average population and within-group standard deviation for each set of simulations. Panel D reports the likelihood of a within-group change for x that is at least as large as either the sample or within-group standard deviation.

Panel A: Simulation Parameters

	(1)	(2)	(3)	(4)	(5)
β_{within}	1.000	1.000	2.000	1.000	1.000
$\beta_{between}$	0.500	0.500	0.500	-1.500	0.500
std(a)	1.000	3.000	1.000	1.000	1.000
std(b)	3.000	1.000	3.000	3.000	0.000

Panel B: Coefficient Estimates

	(1)	(2)	(3)	(4)	(5)
coeff (no fe)	0.552	0.951	0.656	-1.237	0.999
coeff (w/ fe)	1.001	1.000	2.000	1.001	0.999
coeff (std w/ fe)	3.125	3.159	6.263	3.128	1.000

Panel C: Population and Within Standard Deviation

	(1)	(2)	(3)	(4)	(5)
Sample S.D. of x	3.123	3.159	3.131	3.126	1.000
Within S.D. of x	0.995	2.985	0.995	0.995	0.995
Sample S.D. of y	2.049	3.201	2.685	4.661	1.414
Within S.D. of y	1.407	3.147	2.225	1.407	1.407

Panel D: Likelihood of Change Within Group

	(1)	(2)	(3)	(4)	(5)
Sample S.D.	0.030	0.452	0.029	0.030	0.475
Within S.D.	0.477	0.477	0.477	0.477	0.477

Table A4: Simulation results This table presents the simulation results. Panel A gives the input parameters for each set of simulations. Panel B reports the average coefficient estimates for OLS with no fixed effects, OLS with fixed effects, and OLS with fixed effects and standardized x. Panel C reports the average population and within-group standard deviation for each set of simulations. Panel D reports the likelihood of a within-group change for x that is at least as large as either the sample or within-group standard deviation.

Panel A: Simulation Parameters	(1)	(2)	(3)	(4)	(5)
β_{within}	1.000	1.000	2.000	1.000	1.000
$\beta_{between}$	0.500	0.500	0.500	-1.500	0.500
std(a)	1.000	3.000	1.000	1.000	1.000
std(b)	3.000	1.000	3.000	3.000	0.000

Panel B: Coefficient Estimates	(1)	(2)	(3)	(4)	(5)
coeff (no fe)	1.420	0.958	1.532	-2.50e+04	0.999
coeff (w/ fe)	1.001	1.000	2.000	1.001	0.999
coeff (std w/ fe)	3.125	3.159	6.263	3.128	1.000

Panel C: Population and Within Standard Deviation	(1)	(2)	(3)	(4)	(5)
Sample S.D. of x	3.123	3.159	3.131	3.126	1.000
Within S.D. of x	0.995	2.985	0.995	0.995	0.995
Sample S.D. of y	6.649	3.219	7.000	1.87e+05	1.414
Within S.D. of y	1.407	3.147	2.225	1.407	1.407

Panel D: Likelihood of Change Within Group	(1)	(2)	(3)	(4)	(5)
Sample S.D.	0.030	0.452	0.029	0.030	0.475
Within S.D.	0.477	0.477	0.477	0.477	0.477

Table A5: Comparison of within standard deviations for different fixed effects This table presents the sample standard deviation and the within-firm standard deviation for different levels of fixed effects: firm, firm & year, and firm & year × industry, as well as the overall sample standard deviation.

		Within S.D.		
Variable	Sample S.D. (1)	Firm FE (2)	Firm & Year FE (3)	Firm & Ind. × Year FE (4)
Log(Assets)	2.420	0.799	0.632	0.592
Log(Sales)	2.610	0.861	0.723	0.684
Book Leverage	0.420	0.317	0.315	0.305
Market Leverage	0.240	0.152	0.149	0.139
Tangibility	0.250	0.102	0.101	0.096
Log(Capital to Labor)	1.630	0.607	0.554	0.521
CAPEX/Assets	0.070	0.050	0.048	0.045
ROA	0.380	0.217	0.216	0.211
ROE	0.960	0.839	0.839	0.818
Dividends/Assets	0.020	0.012	0.012	0.012
Cash Holdings	0.210	0.114	0.113	0.110
R&D / Assets	0.120	0.061	0.061	0.060
Tobin's Q	2.650	1.719	1.701	1.645
Herfindahl Index	0.140	0.059	0.058	0.000

Table A6: Frequency of firm-level moves of common COMPUSTAT variables - median within-firm This table presents within-firm standard deviation (column 1), the median within-firm standard deviation (column 2), the proportion of observation for a given firm that differs from any other observation for that firm by plus or minus the median firm's standard deviation (column 3), and the frequency that the peak-to-trough movement for a firm over the given time period exceeds the median firm's standard deviation (columns 4 and 5). The sample period is from 1975-2018 and excludes financial firms and utilities. Additionally, we exclude firm-years with negative sales and firm-years missing one of the given variables, and require firms to have at least one spell of five consecutive firm-year observations.

	Within-Firm S.D.	Med. Within-Firm S.D.	Frequency of changes using Med. Within-Firm S.D.		
Variable			Prop.	1-year	10-year
	(1)	(2)	(3)	(4)	(5)
Log(Assets)	0.799	0.453	0.470	0.120	0.692
Log(Sales)	0.861	0.482	0.458	0.126	0.699
Book Leverage	0.317	0.137	0.410	0.212	0.670
Market Leverage	0.152	0.113	0.401	0.242	0.671
Tangibility	0.102	0.066	0.436	0.222	0.749
Log(Capital to Labor)	0.607	0.374	0.445	0.162	0.703
CAPEX/Assets	0.050	0.028	0.386	0.300	0.742
ROA	0.217	0.077	0.405	0.274	0.719
ROE	0.839	0.214	0.380	0.261	0.653
Dividends/Assets	0.012	0.000	0.370	0.365	0.439
Cash Holdings	0.114	0.071	0.391	0.269	0.692
R&D / Assets	0.061	0.001	0.425	0.399	0.508
Tobin's Q	1.719	0.544	0.392	0.276	0.702
Herfindahl Index	0.059	0.016	0.482	0.237	0.669