Journal *of* Accounting Research

Using and Interpreting Fixed Effects Models

MATTHIAS BREUER ID* AND ED DEHAAN ID†

ABSTRACT

Fixed effects (FE) have emerged as a ubiquitous and powerful tool for eliminating unwanted variation in observational accounting studies. Unwanted variation is plentiful in accounting research because we often use rich data to test precise hypotheses derived from abstract theories. By eliminating unwanted variation, FE reduce concerns that omitted variables bias our estimates or weaken test power. FE are not costless, though, so their use should be carefully justified by theoretical and institutional considerations. FE also transform samples and variables in ways that are not immediately apparent, and in doing so affect how we should interpret regression results. This primer explains the mechanics of FE and provides practical guidance for the informed use, transparent reporting, and careful interpretation of FE models.

---

1. Introduction

Fixed effects (FE) have emerged as a ubiquitous and powerful tool for improving identification in accounting research.[1] Similar to methods such as standard control variables, matching, discontinuity designs, and transforming variables into abnormal values (e.g., abnormal accruals), the goal of FE is to eliminate unwanted variation in regression tests. Variation is "unwanted" if it is not part of our theory underlying the cause-and-effect relation of interest. By eliminating such variation, we reduce concerns that omitted variables bias our estimates (e.g., when they are correlated with both the dependent and independent variable) or weaken our test power (e.g., when they increase the residual variation of the dependent variable).

Although FE can improve identification by flexibly purging unwanted variation, they are no panacea. Their use needs to be justified by theoretical and institutional considerations, and the misuse of FE can hurt identification rather than help it. FE also transform the underlying sample and variables in ways that may not be immediately apparent, and in doing so affect how we should interpret regression results.

This primer provides an introduction to the informed use, transparent reporting, and careful interpretation of FE models, specifically focusing on using FE for identifying cause-and-effect relations.[2]

1.1 FE AND IDENTIFICATION IN ACCOUNTING RESEARCH

The prevalence of FE in modern accounting research likely reflects the complex function that accounting plays in businesses and the economy. For example, accounting interfaces numerous corporate functions (e.g., finance, sales, engineering), and the quality of accounting reports depends on complex interactions between firms' fundamentals (e.g., operating performance), reporting decisions (e.g., earnings management), and external forces (e.g., market conditions). To cut through this complexity, our theories necessarily abstract from many economic forces to study the unique informational effects of accounting itself; for example, modeling the effects of reporting decisions on market value, abstracting from operating decisions. By contrast, our rich, observational data typically reflect all of those abstracted forces; for example, thousands of heterogeneous firms

---

[1] Among archival studies in the *Journal of Accounting and Economics*, *Journal of Accounting Research*, and *The Accounting Review* published in 2019, 48% of papers have a model with a high-frequency FE grouping such as for each firm or firm-period. More broadly, 96% of papers have a model with at least one FE grouping (e.g., industries) and 84% of papers have a model with at least two FE groupings (e.g., industries and years). These statistics are even higher than those reported by Amir et al. [2016], who review a broad set of accounting journals in an earlier period.

[2] FE can also be used in descriptive studies, in prediction models, or for measurement purposes (e.g., for measuring CEO styles in Bertrand and Schoar [2003]), in which case purely statistical consideration (e.g., representativeness or overfitting) may take precedence over the theoretical and institutional considerations stressed in this primer.

over decades, and market prices that impound innumerable operating decisions and information. Our data usually therefore include extensive unmodeled variation that raises serious concerns about omitted variable bias and lowers test power. FE can purge broad swaths of unwanted variation to home-in on variation that can be interpreted through the lens of our theories.

Although the empirical challenges of accounting research are not new, the growth in FE usage likely stems from recent developments in and around our field. First, although accounting research since the 1960s has accumulated evidence speaking to many of the field's questions, much of the earlier evidence, as a result of the challenging identification, permits multiple explanations. As the field matures, there is a natural role for studies that reexamine existing questions using alternative and more advanced identification strategies, including strategies leveraging FE.[3] Second, the increasing use of FE likely reflects our field's increasing recognition of the importance of causal identification for informing policy and practice, and the importance of carefully leveraging institutional insights in research designs (e.g., Leuz [2018, 2022]). Third, advances in empirical and computational methods now enable the use of high-frequency, multidimensional FE that were simply impracticable in the recent past (e.g., Correia [2016], Correia, Guimarães, and Zylkin [2020]).

1.2 PAPER SYNOPSIS

FE are a set of indicator variables ($\mathbb{1}_g$) that uniquely identify groups of observations; for example, observations for each subject (i) or time period (t).[4] FE are a type of control variable and, just like any other control, the indicators can be included as independent variables in an ordinary least squares (OLS) regression:

$$y_{i,t} = \beta x_{i,t} + \alpha_g \mathbb{1}_g + \epsilon_{i,t}, \tag{1}$$

where α_g is the corresponding set of coefficient estimates for each FE. For notational convenience, we omit the set of indicators ($\mathbb{1}_g$) when representing FE in regressions. We also discuss subjects in the context of *firms* and time periods as *years*, although the issues we discuss generalize to other FE groupings and data structures. In the following regression, α_i^1 represent

---

[3] Of course, there remains perpetual demand for studies of new hypotheses, in which case, our identification expectations should be inversely related to the novelty and importance of the question being examined.

[4] Formally, we can denote group FE as follows: $\mathbb{1}_g = \mathbb{1}\left(\{i, t\} \in \mathcal{I}_g\right)$ where $\mathcal{I}_g$ is the collection of indexes of observations that belong to group g. Our discussion broadly applies to any indicator variable that is used as a control, though our focus is on using indicators to control for categorical variables. Our discussion is most pertinent in cases where the set of FE indicators comprises an appreciable fraction of the sample size.

firm FE and α_t^2 represent year FE:[5]

$$y_{i,t} = \beta x_{i,t} + \alpha_i^1 + \alpha_t^2 + \epsilon_{i,t}. \tag{2}$$

Section 2 explains the mechanical functions of FE in regressions. Sections 2.1 and 2.2 begin by providing intuition for how FE implicitly de-mean both the dependent variable (Y) and independent variables (X) by each FE group's sample average. In doing so, FE eliminate average differences in Y and X that exist *across* FE groups, and estimate regression coefficients using only the variation that exists *within* each FE group. By focusing on within-group variation, FE can substantially reduce the heterogeneity of observations used to estimate β. FE are especially useful in cases where across-group heterogeneity is not integral for understanding our question of interest but is thought to be correlated with unobserved factors. In those cases, the use of FE can make us incrementally more confident that we have eliminated correlated omitted variables, which threaten to confound our inferences.[6]

Section 2.3 explains how FE can either increase or decrease test power, depending on how much variation they eliminate in Y versus X. Explaining variation in Y increases power and reduces the risk of false negatives (i.e., failure to detect an effect when there is an effect), so is generally helpful. Explaining variation in X decreases test power, but if the eliminated X variation correlates with otherwise uncontrolled determinants of Y, then a reduction in power is the necessary price of using FE to avoid omitted variable bias. Section 2.4 discusses extensions and special cases.

Our discussions elucidate two key differences between FE and other econometric methods such as standard controls or matching. The first is that FE are particularly convenient because they allow researchers to flexibly eliminate unwanted variation across FE groups without needing to measure all the specific factors that drive that unwanted variation. Instead, a researcher needs only to be able to identify the groups that observations belong to (e.g., a given firm or year). The second difference is that FE do not require a separate design stage in which researchers explicitly transform data or drop observations (Imbens and Rubin [2015]). For example, a researcher drops observations in matched-sample designs or

---

[5] Researchers use various practices to represent FE in equations. One practice is to show a set of coefficients and named FE: "$\beta x_{i,t} + \sum \alpha_i Firm_i + \epsilon_{i,t}$." Other studies write-out the FE: "$\beta x_{i,t} + FirmFE + \epsilon_{i,t}$." We represent FE as α to reflect our later point that FE can be thought of as unique intercepts for each group (given that intercepts are often represented by α), and we use superscripts to differentiate FE groupings (i.e., firm and year). We could also use different symbols (e.g., α_i and η_t) for the distinct FE groupings.

[6] Confidence in our inferences is always a matter of degree: The stronger we expect the theoretical and institutional arguments for the identification strategy to be, the more plausible it is that we can attribute an *association* between Y and X to the *causal effect* of that particular X instead of other correlated forces. In this vein, FE can increase our confidence by reducing concerns about omitted variable bias, a common source of endogeneity. By contrast, FE cannot easily address other sources of endogeneity, such as simultaneity or reverse causality (see, e.g., Roberts and Whited [2013]).

explicitly calculates abnormal accruals, and these procedures are usually clearly described in sample selection tables and descriptive statistics. FE implicitly perform sample refinements and variable transformations in a single step while estimating a regression, which is convenient and reduces researcher subjectivity.

Our discussions also highlight that the single-step convenience of FE can introduce problems that may not be immediately apparent. One issue is that FE are so easily included in modern software that researchers can neglect to carefully consider the specific concerns that FE are intended to address. Without careful consideration, we risk using FE in ways that undermine rather than improve identification; for example, hurting test power by including FE when a linear control would be equally effective. Other issues stem from the fact that FE can significantly alter which observations contribute to coefficient estimates; for example, FE groups with little or no within-group variation in the X of interest play a limited role in estimating β. FE can therefore mask the effective sample used to identify results, and can exacerbate generalizability concerns about whether treatment effects would be similar among FE groups that have little variation in X. Finally, focusing solely on within-group variation also has implications for interpretation of coefficient estimates (e.g., by changing from levels to deviations), affects the quantification of the economic magnitudes (e.g., by changing the effective standard deviation of X), and alters approaches to outlier adjustments (e.g., truncating within-group outliers rather than full-sample outliers).

To avoid said problems, section 3 provides guidance and best practices for the informed use and transparent reporting of FE. Section 3.1 suggests to clearly motivate the use of FE in research designs, and describe the resulting identifying variation. Section 3.2 provides guidance on sample-construction issues and choices pertinent to FE regressions. Section 3.3 recommends additional statistics to make transparent the sample and variable transformations implicitly performed by FE regressions. Section 3.4 concludes with diagnostic and robustness tests.

1.3 CONTRIBUTION AND LIMITATIONS

The contribution of this primer is to introduce new accounting researchers to FE and fill in knowledge gaps for experienced researchers. Knowledge gaps are understandable given that many econometrics texts devote few pages to FE and provide minimal discussion of issues that are germane in accounting studies. For example, Roberts and Whited [2013] provide an overview of methods for addressing endogeneity but only a couple of pages on the tradeoffs involved in considering FE models. Similarly, Angrist and Pischke [2008] briefly discuss FE and related tradeoffs in the context of difference-in-differences (DiD) designs. Wooldridge [2010] explains FE mechanics but does not get into many of the conceptual issues we discuss below. A few recent studies in the accounting and finance literature discuss select issues of FE models. Gormley and Matsa [2013] and Amir et al. [2016], for example, show that industry and firm

FE, respectively, can reduce bias arising from unobserved cross-sectional heterogeneity, and Armstrong et al. [2022] and Jennings et al. [2024] show that including unnecessary FE can exacerbate bias in certain cases. Our primer complements those studies by providing a broader review of FE considerations in accounting research.

Our primer also has several limitations. First, our explanations gloss over some econometric details and special cases, and instead focus on intuition for the things that are most relevant in accounting research. Second, we ignore or brush over many other econometric issues that should be considered in designing tests, such as the effects of measurement error or violations of OLS assumptions. Third, we only briefly talk about FE in nonlinear models such as Logit or Poisson, which have more complex issues than OLS. Lastly, we selectively refer to examples for certain FE models and practices in accounting studies, many of which hail from our own work. Our selections reflect our familiarity with those studies and should not be taken to mean that those studies are the only good examples in the literature. We provide references for further reading throughout the paper.

2. What FE Do

Our discussion of the functioning of FE proceeds in four parts. Section 2.1 discusses the mechanics of FE in linear regressions. Sections 2.2 and 2.3 discuss the effects of FE on the bias and variance of regression estimates, respectively. Section 2.4 discusses extensions and additional considerations.

2.1 THE MECHANICS OF REGRESSIONS WITH FE

We start by illustrating the mechanics of FE using a simple OLS regression of a dependent variable (Y) on one independent variable (X). We assume that our panel data set consists of subjects i (e.g., firms) over time periods t (e.g., years). We then extend to the case of a regression with multiple Xs.

2.1.1. Regression Without and with a Constant. Consider the Y and X data that are histogrammed in panels A and B of figure 1. The data pertain to three firms represented by different shading, but for now we will consider the data together.

Without a constant, the regression equation of Y on X and the closed-form solution of the OLS coefficient estimate are as follows:

$$y_{i,t} = \beta x_{i,t} + v_{i,t}, \tag{3}$$

$$\hat{\beta} = \arg\min_{\beta} \sum (y_{i,t} - \beta x_{i,t})^2, \tag{4}$$

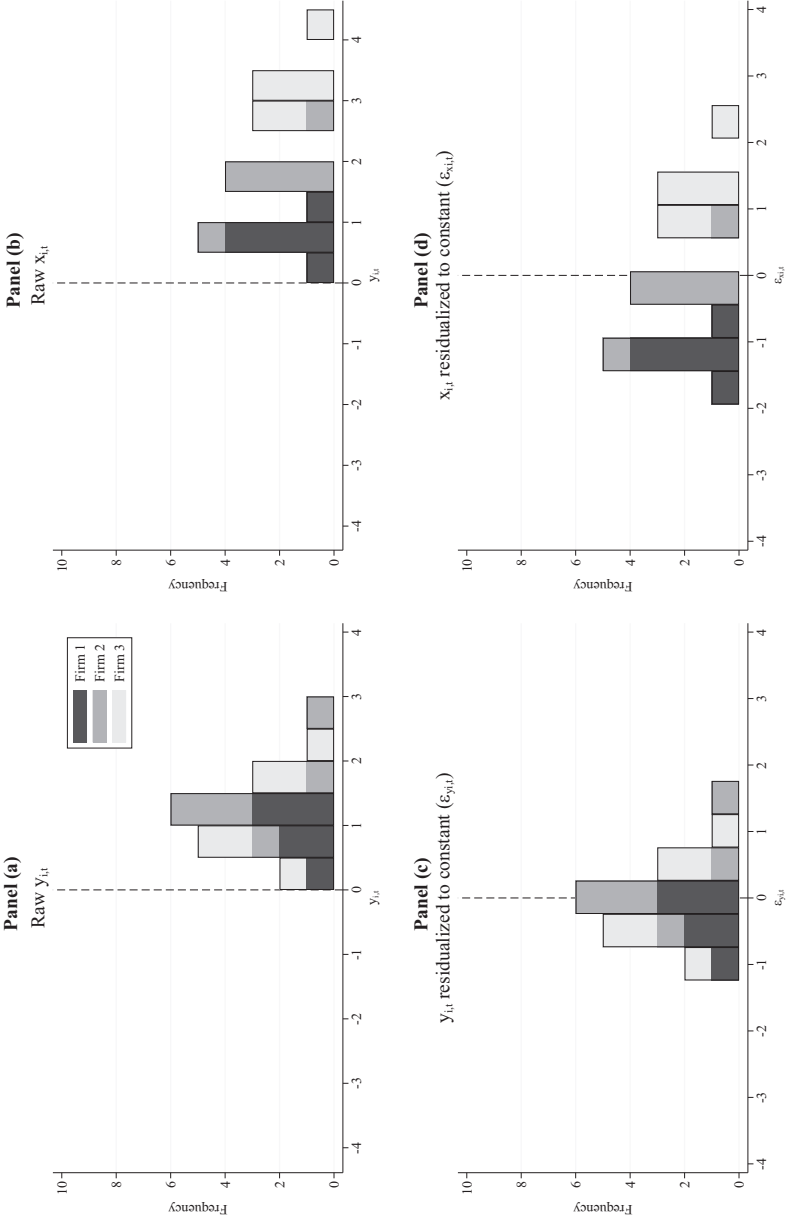$$\hat{\beta} = \frac{\sum y_{i,t} x_{i,t}}{\sum x_{i,t}^2}. \tag{5}$$

FIG. 1.—Histograms. The figure shows the distributions of Y and X before and after adjustments for a constant or firm FE. The underlying data comprise three firms, which contribute six observations each. The firm-specific data are represented through different shadings. Panel A plots the raw distribution of Y ($y_{i,t}$). Panel B plots the raw distribution of X ($x_{i,t}$). Panel C plots the distribution of X after residualizing to the constant ($\epsilon_{x_{i,t}}$). Panel D plots the distribution of X after residualizing to the constant ($\epsilon_{x_{i,t}}$). Panel E plots the distribution of Y after residualizing to firm FE ($\epsilon_{y_{i,t}}$). Panel F plots the distribution of X after residualizing to firm FE ($\epsilon_{x_{i,t}}$). The dotted lines in the graphs depict the location of zero on the x-axis.
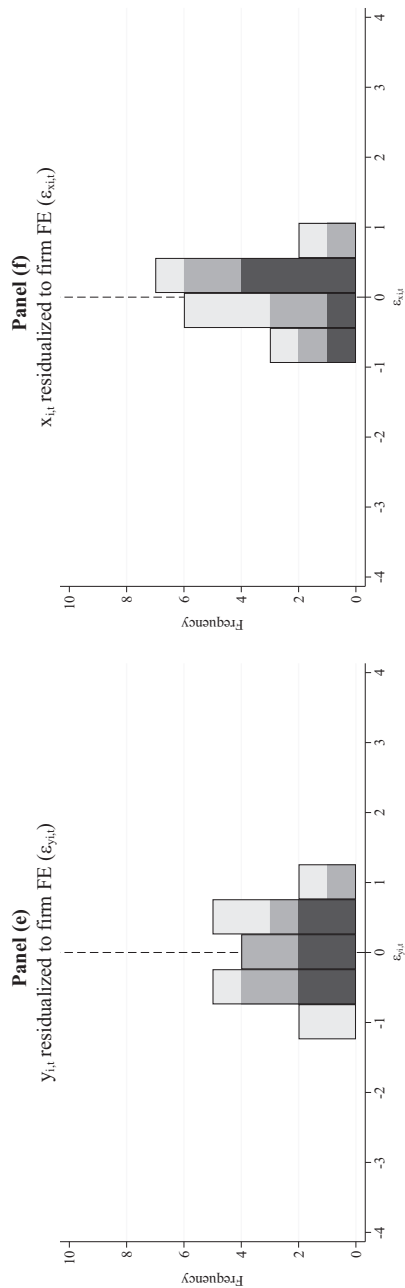
FIG. 1.—*Continued*

Without a constant, the linear specification forces the regression slope to run through the origin; that is, it predicts Y to be zero ($\hat{y}_{i,t} = 0$) when X is zero ($x_{i,t} = 0$). For example, panel A of figure 2 plots the joint distribution of our X and Y data, and the red line in panel B is the fitted regression slope that intersects the origin $(0,0)$. In our example data, though, the data-generating process includes a nonzero intercept, so the slope in panel B is biased (upwards, in this case).

To avoid such bias, we usually add an intercept to the OLS specification, represented by α, which allows the regression line to intersect the y-axis at somewhere other than $(0,0)$. We do so by adding a constant, which is an indicator that takes a fixed value of one for all observations in the sample. The constant is a special version of a FE, one with only one group ($G = 1$). Hence, a regression with a constant is a special case of a FE regression. Besides being a special case, it is an instructive case because many researchers are familiar with regressions with a constant. The corresponding regression and closed-form solution of the OLS coefficient estimates are given by:

$$y_{i,t} = \alpha + \beta x_{i,t} + \epsilon_{i,t}, \tag{6}$$

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta}{\arg\min} \sum (y_{i,t} - \alpha - \beta x_{i,t})^2, \tag{7}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \tag{8}$$

$$\hat{\beta} = \frac{\sum (y_{i,t} - \bar{y})(x_{i,t} - \bar{x})}{\sum (x_{i,t} - \bar{x})^2}. \tag{9}$$

The red line in panel C of figure 2 plots the regression line for the case with a constant. The estimate of the constant (i.e., $\hat{\alpha}$) is roughly 1.2, meaning that the line now intersects the y-axis at roughly 1.2. In contrast to panel B without a constant, the regression slope with a constant is now slightly negative.

Equations (8) and (9) provide two important insights. First, the estimate of the constant, $\hat{\alpha}$, is set so that the point of means $(\bar{x}, \bar{y})$, instead of the origin $(0,0)$, lies on the regression line. To see this, equation (8) can be rearranged to: $\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$. Given $\hat{\beta}$, $\hat{\alpha}$ is chosen such that this regression line holds for the point of means. Second, by including a constant, the $\hat{\beta}$ coefficient from equation (9) now captures the comovement of the *de-meaned* Y and X. Specifically, $y_{i,t}$ and $x_{i,t}$ in equation (9) are each de-meaned by their sample averages, that is, $(y_{i,t} - \bar{y})$ and $(x_{i,t} - \bar{x})$, while they were not in equation (5).[7] Both insights together illustrate that the constant recenters the observations around the point of means by de-meaning Y and X. As a result, the point of means, essentially, becomes the new (dotted) origin for the regression as shown in panel C of figure 2.

---

[7] Another way to see this second insight is to use equation (8) to substitute $\bar{y} - \beta\bar{x}$ for α in equation (6). Substituting and rearranging produces a regression without a constant but with de-meaned Y and X: $(y_{i,t} - \bar{y}) = \beta(x_{i,t} - \bar{x}) + \epsilon_{i,t}$.
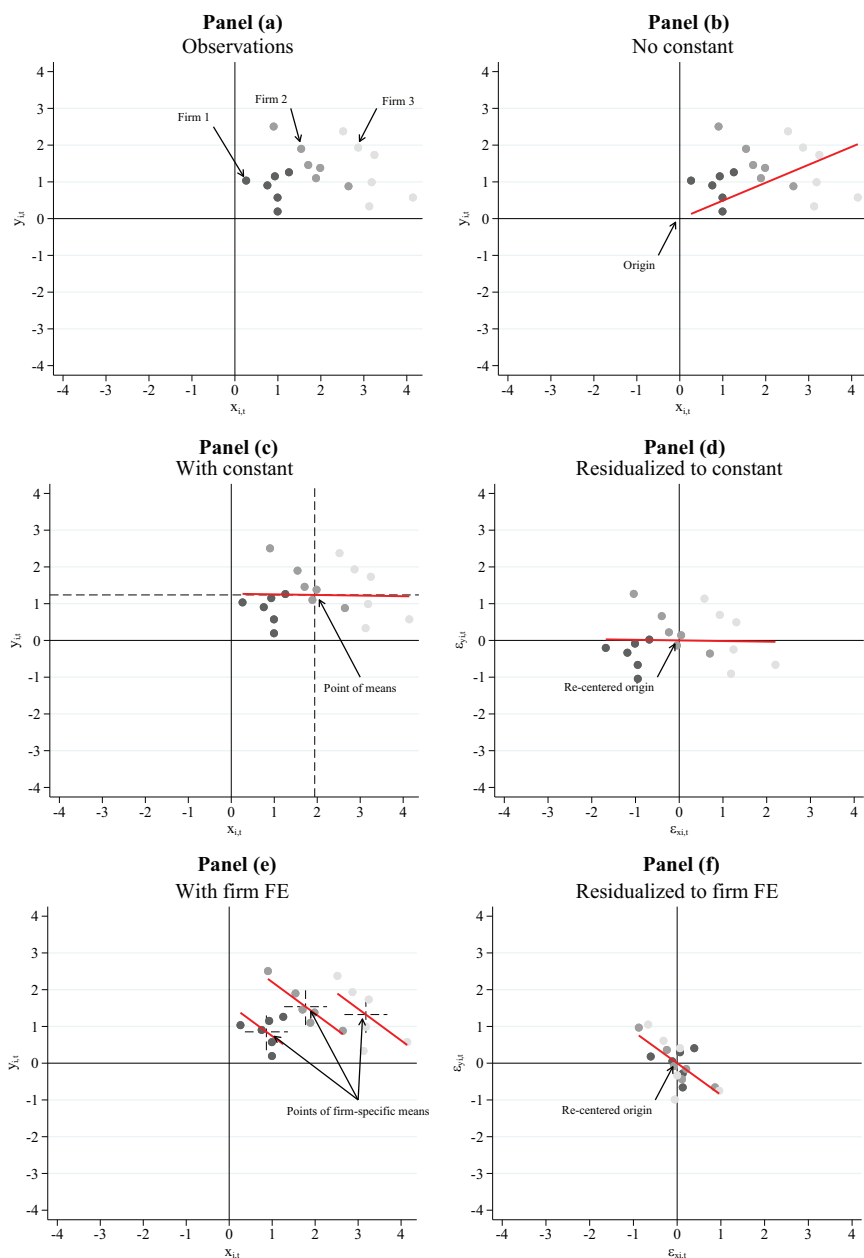
FIG. 2.—Regressions. The figure shows regressions of Y on X with and without a constant or firm FE. The underlying data comprise three firms, which contribute six observations each. Panel A plots the data ($y_{i,t}$ and $x_{i,t}$) in a scatter plot. Panel B plots the regression line of a regression of $y_{i,t}$ on $x_{i,t}$ without a constant in red. Panel C plots the regression line of a regression of $y_{i,t}$ on $x_{i,t}$ with a constant in red. Panel D plots the regression line of a regression of the residuals of $y_{i,t}$ ($\epsilon_{i,t}^{y}$) on the residuals of $x_{i,t}$ ($\epsilon_{i,t}^{x}$) after residualizing to the constant in

red. Panel E plots the regression line(s) of a regression of $y_{i,t}$ on $x_{i,t}$ with firm FE (for each firm) in red. Panel D plots the regression line of a regression of the residuals of $y_{i,t}$ ($\epsilon_{i,t}^y$) on the residuals of $x_{i,t}$ ($\epsilon_{i,t}^x$) after residualizing to firm FE in red. The dotted black lines show the intercept(s) implied by the constant (panel C) or firm FE (panel E). Firm FE imply group-specific intercepts, hence, the three distinct intercepts in panel E.

The case of the univariate regression with a constant shows that including a single FE (i.e., a constant) implicitly de-means both Y and X before running the regression. This insight reflects the *projection theorem*, which states that the following two-step procedure recovers the same $\hat{\beta}$ coefficient as an OLS regression of Y on X (Frisch and Waugh [1933], Lovell [1963]). The first step is to calculate the residuals of regressions of Y and X on a constant each. This step de-means Y and X:

$$y_{i,t} = \alpha^y + \epsilon_{i,t}^y, \tag{10}$$

$$x_{i,t} = \alpha^x + \epsilon_{i,t}^x, \tag{11}$$

where α^y and α^x denote the respective constant values. The regression estimates of those values, $\hat{\alpha}^y$ and $\hat{\alpha}^x$, correspond to the in-sample means of Y ($\bar{y}$) and X ($\bar{x}$), respectively. After partialling out those estimates (i.e., deducting $\hat{\alpha}^y$ from $y_{i,t}$ and $\hat{\alpha}^x$ from $x_{i,t}$), we obtain the residuals $\hat{\epsilon}_{i,t}^y$ and $\hat{\epsilon}_{i,t}^x$, which, in essence, represent the de-meaned Y ($y_{i,t} - \bar{y}$) and X ($x_{i,t} - \bar{x}$).

Panels C and D of figure 1 show the distributions of residuals $\hat{\epsilon}_{i,t}^y$ and $\hat{\epsilon}_{i,t}^x$. As can be seen, residualizing the Y and X to a constant shifts their distributions to be centered around zero. Panel D of figure 2 shows that the joint distribution of the residualized values is now centered around the origin.

The second step of the projection theorem is to regress the residuals of Y ($\hat{\epsilon}_{i,t}^y$) on the residuals of X ($\hat{\epsilon}_{i,t}^x$):

$$\hat{\epsilon}_{i,t}^y = \beta \hat{\epsilon}_{i,t}^x + \nu_{i,t}, \tag{12}$$

$$\hat{\beta} = \frac{\sum \hat{\epsilon}_{i,t}^y \hat{\epsilon}_{i,t}^x}{\sum \hat{\epsilon}_{i,t}^{x\,2}} = \frac{\sum (y_{i,t} - \bar{y})(x_{i,t} - \bar{x})}{\sum (x_{i,t} - \bar{x})^2}. \tag{13}$$

The resulting coefficient estimate, $\hat{\beta}$, is equal to the coefficient estimate from regressing the untransformed Y on X and a constant as in equation (9).[8] This equivalence can also be seen in panel D of figure 2: The

---

[8] To obtain the same standard error as in a one-step approach, we need to adjust for the degrees of freedom used up in the first step.

TABLE 1
Fixed Effects and Constant

Group	Constant	$\mathbb{1}_1$	$\mathbb{1}_2$	$\mathbb{1}_3$
1	1	1	0	0
1	1	1	0	0
2	1	0	1	0
2	1	0	1	0
3	1	0	0	1
3	1	0	0	1

The table presents an excerpt of the X matrix for three groups with two observations per group. The matrix includes a column of ones to identify the constant. It also includes three group-level indicators denoting the group FE.

fitted regression line has exactly the same slope as in the untransformed data with a constant in panel C.[9]

The equivalence demonstrated by the projection theorem illustrates that the inclusion of a constant essentially transforms both Y and X by deducting their means before estimating the regression. Importantly, this equivalence extends beyond the simple univariate regression example to regressions with multiple Xs (e.g., we can residualize another X in the first step and include its residuals in the second-step regression). The insights gained from the univariate example also extend to the case of regressions with constants for multiple groups ($G > 1$), which we discuss next.

2.1.2. Regression with FE. FE regressions essentially break up the single constant into G separate constants, one for each group (e.g., firm or year). We can illustrate this disaggregation by looking at an excerpt of the matrix of Xs containing the constant and FE, represented by $\mathbb{1}_g$, for each of three groups (see table 1). The sum of the FE is equal to the constant: $Constant = \mathbb{1}_1 + \mathbb{1}_2 + \mathbb{1}_3$. FE allow each group to have a distinct intercept, so we replace the single α with a set of G constants.

$$y_{i,t} = \alpha_g + \beta x_{i,t} + \epsilon_{i,t}. \tag{14}$$

And the new coefficient estimates are:

$$\hat{\alpha}_g = \bar{y}_g - \hat{\beta}\bar{x}_g, \tag{15}$$

$$\hat{\beta} = \frac{\sum (y_{i,t} - \bar{y}_g)(x_{i,t} - \bar{x}_g)}{\sum (x_{i,t} - \bar{x}_g)^2}. \tag{16}$$

[9] In the accounting literature, Chen, Hribar, and Melissa [2018] use this projection theorem to advocate for a one-step estimation of accruals models. The projection theorem is also useful for removing the effects of correlated variables when plotting relations between Y and X. For example, deHaan et al. [2021] plot the relation between index funds' fees and disclosure readability, but first residualize both variables to year FE to remove the effects of time trends. Or, Kim [2022] plots the relation between banks' loan loss provisions and local housing prices, but first residualizes both variables to controls and FE to better visually isolate the relation of interest.

The key difference from the regression with a single constant is that equations (8) and (9) subtracted overall sample means of Y and X, whereas equations (15) and (16) subtract *group-specific* means. In essence, although the single constant recenters the entire sample around the origin, FE separately recenter each group's observations around the origin.

To illustrate this point, panels E and F of figure 1 present the same data from panels A and B, but after residualizing Y and X to firm FE. The effect of residualizing to firm FE is to separately recenter each firm's observations around zero. The recentering essentially means that each firm obtains its own intercept, illustrated in panel E of figure 2. Panel F of the same figure shows that fitting a FE regression line on the recentered data produces a strongly negative slope, which is very different from the roughly flat line without FE in panel D. In this case, the difference is due to potentially confounding heterogeneity across the different firms; that is, that firms with larger values of X, on average, also exhibit larger values of Y.

One key insight here is that $\hat{\beta}$ in FE regressions focus on within-group variation in X (i.e., $x_{i,t} - \bar{x}_g$) and Y (i.e., $y_{i,t} - \bar{y}_g$).[10] Variation resulting from differences in means across groups is implicitly eliminated.[11] This is why researchers often say that FE "restrict analyses to within-group variation in X." As a result of this transformation, $\hat{\beta}$ can be interpreted as capturing predicted changes in Y in response to deviations of X from a group's mean.

A second insight is that the data residualized to FE in panel E and F of figure 1 are more narrowly distributed than the raw recentered data in panels C and D. This difference becomes especially important when using the distributional properties of X (e.g., its standard deviation) to characterize economic magnitudes, which we further discuss in section 3.

A third insight from equation (16) is that FE regressions transform the raw variation in Y and X to within-group variation *and* estimate the resulting regression coefficient in one step. This one-step approach can be more convenient than manually preprocessing data (e.g., calculating "abnormal" values or deducting industry means) and can avoid inconsistent adjustments (e.g., Gormley and Matsa [2013]). But this one-step approach, without further detail on the transformed sample statistics, is also less transparent.

Finally and as a practical matter, because the group FE are finer partitions of the constant, the constant and the set of G FE are collinear, preventing us from estimating both in the same model.[12] To remedy the collinearity issue,

---

[10] The coefficient estimates of the variable of interest in FE regressions are a linear combination (or weighted average) of within-group estimates. Accordingly, the coefficient estimates in FE regressions also use cross-group variation (e.g., in the size of the group-specific estimates), strictly speaking.

[11] Other differences in distributions across groups (e.g., variance) are not eliminated.

[12] For this reason, for example, a coarse *Post* indicator for years after the introduction of the Sarbanes-Oxley Act (SOX) cannot be included together with finer year FE (e.g., Engel, Hayes, and Wang [2007]). The finer FE would subsume the *Post* indicator. As a result, broader

we must drop either the constant or any one of the group FE. If the constant is dropped, each $\hat{\alpha}_g$ estimate captures the intercept of its group. If instead one of the FE is dropped, $\hat{\alpha}$ captures the intercept for the left-out group, and each remaining group FE, $\hat{\alpha}_g$, captures the difference between the left-out group's intercept and group g's intercept.[13] Either way, the coefficient estimates on other Xs are unaffected, so the choice is irrelevant as long as the FE are merely used as controls.[14]

2.1.3. Regression with Multiple FE Groupings and Controls. The insights from the above example extend to the case of multiple sets of FE. For example, we might include FE for each firm i and each year t:

$$y_{i,t} = \alpha_i^1 + \alpha_t^2 + \beta x_{i,t} + \epsilon_{i,t}. \tag{17}$$

Heuristically, the firm FE deduct the firm-specific mean from each of the variables, whereas the year FE simultaneously deduct the year-specific mean from the variables.[15] As a result of the two-way FE, the double-de-meaned X only varies if a given firm experiences a deviation from its mean level that is different from the average deviation of all firms in the same year.

Two sets of FE can sometimes be subdivided or combined to form finer groupings. For example, in a data set with multiple observations for each firm within a year, firm and year FE could be combined into a single, finer firm-year FE grouping. Or, year FE could be subdivided into finer date FE (i.e., month/day/year). In both cases, though, the finer FE groupings are subsets of the coarser FE groupings, so both could not be used in the same model.

Finally, additional control variables, $z_{i,t}$ can also be added to equation (17) with little complication. FE will de-mean $z_{i,t}$ in the same way as $x_{i,t}$. As long as $z_{i,t}$ varies within the FE groups, OLS will estimate a regression coefficient. Section 2.2 discusses the case where $z_{i,t}$ does not vary within FE groups.

2.1.4. The Effects of FE on Contributing Observations. By limiting the analysis to within-group variation, FE can also change whether and how

---

time-trends cannot be easily controlled for with year FE *if* we are interested in a treatment, like SOX, that was introduced for all our sample firms around the same time (e.g., Leuz [2007]).

[13] Consider, for example, the FE estimate for group $g = 2$. In case the constant is dropped, the FE estimate corresponds to the conditional expectation of Y for the case where the group FE takes the value of one and all other Xs are zero: $E[y_{i,t}|g = 2] = \alpha_2$. If the constant is included, this conditional expectation takes the following form: $E[y_{i,t}|g = 2] = \alpha + \alpha_2$. In this case, the FE estimate (α_2) is the difference between the conditional expectation of Y ($\alpha + \alpha_2$) and the constant (α).

[14] The constant estimate reported by statistical software may vary across packages. In some packages (e.g., STATA's "AREG"), the constant estimate represents the average intercept across all FE groupings.

[15] The double-de-meaning calculation is cumbersome in unbalanced panels and not typically amenable to simple closed-form expressions. See Greene [2002, p. 293].

observations affect coefficient estimates. FE groups with just a single observation, for example, become irrelevant for estimating $\hat{\beta}$. These "singleton" groups are subsumed by the group FE given that $x_{i,t} = \bar{x}_g$, $y_{i,t} = \bar{y}_g$, and the same is true for any other regressors. Although singletons do not affect coefficients, their presence can understate standard errors. We discuss the impact on standard errors in section 2.4.6, and provide guidance on dealing with singletons in section 3.

FE groups with multiple observations but no within-group variation in the X of interest (so-called "no-variation" groups) affect standard errors but may or may not affect $\hat{\beta}$. No-variation groups have no effect on $\hat{\beta}$ in the simple case of one X and one set of FE. For the more common case of multiple Xs or FE groupings, no-variation groups can contribute to estimating $\hat{\beta}$ as long as they have variation in at least one regressor. They do so by helping to estimate the coefficients on control variables and other sets of FE, which in turn influence the estimate of $\hat{\beta}$. In this regard, no-variation groups act much like control groups in matching designs or simple DiD models.[16] However, although control groups are usually easily observable in other methods, no-variation groups in FE models are only identifiable using supplementary diagnostics. For a specific example of no-variation groups and detailed discussion of DiD models with two sets of FE, see section 2.4.1. Section 3 provides guidance on how to identify and handle no-variation groups.

Besides altering the effective sample used to estimate $\hat{\beta}$, FE regressions can also change which observations have the largest influence on $\hat{\beta}$. Extreme observations in the tails of the raw distributions of Y and X, for example, can fall in the middle of the distributions of the within-group de-meaned Y and X. Thus, the most influential observations in FE regressions tend to be those that have the largest within-group variation in Y or X, as opposed to those that are outliers in the raw data. Relatedly, variables that are abnormally distributed (e.g., truncated or highly skewed) in the raw data can be more normally distributed in de-meaned data. As further discussed in section 3, these changes have implications for the interpretation of the coefficient estimates, approaches to outlier treatments, and choices about variable transformations and functional forms.

2.2 BIAS

Omitted variable bias occurs if a determinant of Y is correlated with X and omitted from the regression specification. Consider, for example, the following data-generating process where $x_{i,t}$ is the X of interest:

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta z_{i,t} + \epsilon_{i,t}. \tag{18}$$

---

[16] No-variation groups can also be "always treated" firms in DiD models, in which case, they are not "control observations" in the typical sense.

TABLE 2
Fixed Effects and Confounder

Group	$z_{i,t}$	$\mathbb{1}_1$	$\mathbb{1}_2$	$\mathbb{1}_3$
1	0.2	1	0	0
1	0.2	1	0	0
2	0.1	0	1	0
2	0.1	0	1	0
3	−0.3	0	0	1
3	−0.3	0	0	1

The table presents an excerpt of the X matrix for three groups with two observations per group. The matrix includes a column showing the (group-varying) values of confounder $z_{i,t}$. It also includes three group-level indicators denoting the group FE.

If we omit $z_{i,t}$ from the regression (e.g., because it is unobservable to us), the coefficient estimate for $x_{i,t}$ is as follows:

$$y_{i,t} = \alpha + \beta x_{i,t} + v_{i,t}, \tag{19}$$

$$\hat{\beta} = \beta + \underbrace{\delta \frac{\sum (x_{i,t} - \bar{x})(z_{i,t} - \bar{z})}{\sum (x_{i,t} - \bar{x})^2}}_{bias}. \tag{20}$$

The estimate, $\hat{\beta}$, is different from the true value, β (i.e., is biased), *if* $z_{i,t}$ is a determinant of $y_{i,t}$ (i.e., $\delta \neq 0$) *and* covaries with $x_{i,t}$ (i.e., $\sum (x_{i,t} - \bar{x})(z_{i,t} - \bar{z}) \neq 0$).

If $z_{i,t}$ is the same for all observations within each group (i.e., $z_{i,t} = z_g$), then group FE can perfectly eliminate the omitted variable bias in equation (20). We can see this intuition in two ways.

First, as discussed, including FE in a regression amounts to including a separate indicator variable for each group. If $z_{i,t}$ is the same for all observations within group g, then the FE form a perfect linear combination of $z_{i,t}$. For example, in the X matrix with $G = 3$, shown in table 2, $z_{i,t}$ is perfectly explained by the following linear combination of the FE:

$$z_{i,t} = 0.2 \times \mathbb{1}_1 + 0.1 \times \mathbb{1}_2 + -0.3 \times \mathbb{1}_3. \tag{21}$$

The FE subsume all of the variation in $z_{i,t}$, so $z_{i,t}$ can be safely omitted from the regression. In fact, because OLS regressions cannot accommodate an X that is a linear combination of other Xs, the FE and $z_{i,t}$ could not be included together even if $z_{i,t}$ was observable.

The second way to understand the intuition of FE is through the within-group de-meaning calculations from section 2.1.2. FE have the effect of de-meaning Y and all Xs, including $z_{i,t}$. If $z_{i,t} = z_g$, the bias term in equa-

tion (20) becomes:

$$\delta \frac{\sum (x_{i,t} - \bar{x}_g)(z_{i,t} - \bar{z}_g)}{\sum (x_{i,t} - \bar{x}_g)^2} = \delta \frac{\sum (x_{i,t} - \bar{x}_g)(0)}{\sum (x_{i,t} - \bar{x}_g)^2} = 0. \quad (22)$$

Because $(z_{i,t} - \bar{z}_g) = 0$, the numerator is zero and, hence, the bias term is effectively eliminated.

The above examples show that we can account for the unobservable and omitted $z_{i,t}$ by controlling for group FE. In this vein, FE are just another form of a control variable. Like all controls, FE help rule out alternative explanations by controlling for confounding differences across groups. Unlike standard linear controls though, FE do not assume a linear relationship between the levels of Y and controls.[17] Instead, they estimate this relationship nonparametrically, allowing each group to be associated with a different level of Y (i.e., to have a different intercept).

As a result, group FE not only control for one confounding group-specific factor, but instead control for all factors that vary only at the group level. This is true for both observed and unobserved factors. The latter renders FE a potent approach to reducing correlated omitted variable concerns in cases where group-level factors threaten to confound the estimation but cannot easily be measured. All we need to know to use FE is the group each observation is classified into.[18]

The above example shows the usefulness of FE for the extreme case in which the omitted $z_{i,t}$ is completely fixed within each group. Omitted factors likely often vary within FE groups in practice, but FE can still help reduce omitted variable bias as long as the within-group variation of Y and X are thought to be less correlated with omitted factors than the across-group variation. Section 3.4 further discusses the tradeoffs in using finer FE groupings to further mitigate within-group variation.

2.3 VARIANCE

FE eliminate omitted variables by controlling for confounding variation in Y and X. As in panels E and F of figure 1, the residualized Y and X after controlling for FE have more concentrated distributions, and in panel F of figure 2, the plotted observations are more tightly clustered around the origin. Although this reduction in variation is desirable to eliminate bias, it should not be surprising that removing variation can also affect standard errors and test power, and often for the worse. The tradeoff between bias and

---

[17] FE rely on appropriate groupings and additive separability though. That is, FE require that groups are chosen such that $z_{i,t}$ is unrelated to $y_{i,t}$ *within* groups and that there are no unaccounted interactions across groups or with other determinants of $y_{i,t}$.

[18] Firms operating in different industries or countries, for example, may face vastly different institutional environments that are hard to comprehensively measure and control for. As long as the institutional environments do not change significantly over our sample period, however, we can easily account for this unwanted variation via industry or country FE.

variance of regression estimates is a familiar one, inherent in many empirical methods (e.g., controlling, matching, regression discontinuity designs (RDD)).

FE can both increase or decrease test power, depending on the specifics of the setting and model. Here, we discuss test power and estimating standard errors in the case of observations with independently and identically distributed error terms. Issues relating to dependent observations are discussed in section 2.4.6.

The standard error for the coefficient of the X of interest is given by:

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\left(\frac{1}{N-p-1}\right)\left(\frac{\sum \hat{\epsilon}_{i,t}^2}{\sum (x_{i,t}-\bar{x})^2 (1-\hat{R}_x^2)}\right)}, \tag{23}$$

where $\hat{\epsilon}_{i,t}^2$ is the variance of the regression residual (i.e., the unexplained variation of Y); N is the sample size adjusted for the $p+1$ degrees of freedom used up by the intercept and p other regression parameters (i.e., other controls, including FE); $\sum (x_{i,t}-\bar{x})^2$ is the sample variation in $x_{i,t}$; and $\hat{R}_x^2$ is the fraction of variation in $x_{i,t}$ that is explained by all other controls and FE. Specifically, $\hat{R}_x^2$ is the unadjusted R^2 of regressing $x_{i,t}$ on the other controls and FE.[19]

The standard error is a central input in determining the statistical significance of the coefficient estimate. Equipped with the coefficient and standard-error estimates, the t-statistic is:

$$t_{\hat{\beta}} = \frac{\hat{\beta}-\beta_0}{\hat{\sigma}_{\hat{\beta}}}, \tag{24}$$

where β_0 is the null hypothesis for the slope coefficient, usually zero. The t-statistic provides a standardized statistic, which can be mapped into a p-value. Under standard assumptions, the mapping from the t-statistic to the p-value occurs via the (inverse cumulative) t-distribution with $N-p-1$ degrees of freedom. The smaller the standard error and the larger the degrees of freedom, the greater the statistical power of the test.[20]

The equations above illustrate that the precision of $\hat{\beta}$ and power to reject the null increase when the following increase:

- $\sum (x_{i,t} - \bar{x})^2$: the variation of $x_{i,t}$;

---

[19] The term $\sum (x_{i,t} - \bar{x})^2 (1 - \hat{R}_x^2)$ therefore represents the residual variation in $x_{i,t}$ after controlling for the share of variation explained by other regressors. As discussed in section 2, a multiple regression with controls and FE, by the projection theorem, yields the same coefficient estimate as the regression of residualized values of Y and X, where the residualized values are the error terms of regressions of Y and the X of interest on the other Xs (including FE). Similarly, the standard error equation for the coefficient of interest is close to the basic equation for the univariate case. The key differences are just that the multivariate standard error only uses the residual variation of $y_{i,t}$ ($\sum \hat{\epsilon}_{i,t}^2$) (after accounting for all other Xs) and the residual variation of $x_{i,t}$ ($\sum (x_{i,t} - \bar{x})^2 (1 - R_x^2)$) (after accounting for all other Xs) instead of their raw variation.

[20] On the uses and misuses of p-values and significance tests, see, for example, McShane et al. [2019a] and Imbens [2021].

- N: the number of observations.

And power decreases when the following increase:

- $\sum \hat{\epsilon}_{i,t}^2$: the unexplained variation of $y_{i,t}$;
- $\hat{R}_x^2$: the share of the variation of $x_{i,t}$ explained by other Xs and FE;
- p: the number of regression parameters.

FE affect several of the determinants of the standard error and p-value, so their impact on power is ambiguous. They can increase power by shrinking $\sum \hat{\epsilon}_{i,t}^2$, but they can decrease power by increasing $\hat{R}_x^2$ and p. Although the same is true for any control variable, the key difference is that a single control usually has a much smaller impact on these determinants than does a potentially large set of FE (e.g., indicators for thousands of firms). Unlike other controls, FE can also reduce power by decreasing N if singletons or no-variation groups are dropped.[21]

The impact of FE on test power depends on the correlation structure between Y ($y_{i,t}$), the X of interest ($x_{i,t}$), and other regressors ($z_{i,t}$). To illustrate the impact of FE in various cases, we revisit the following basic setup in which z_g may or may not be a confound:

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta z_g + \epsilon_{i,t}. \tag{25}$$

z_g is unobservable, so we instead control for it using FE for group g:

$$y_{i,t} = \alpha_g + \beta x_{i,t} + \epsilon_{i,t}. \tag{26}$$

In case 1, z_g is *not* a determinant of $y_{i,t}$ (i.e., $\delta = 0$ in equation (25)) and is *not* correlated with $x_{i,t}$. In this case, z_g is irrelevant, and FE are unnecessary. Including FE will decrease power by increasing p and likely $\hat{R}_x^2$ (which is the unadjusted r-squared, so can increase when any regressor is added). Thus, unnecessary FE will increase the risk that $\hat{\beta}$ is statistically insignificant even though $x_{i,t}$ does affect $y_{i,t}$ in truth (a false negative or Type 2 error).

In case 2, z_g is *not* a determinant of $y_{i,t}$, so does not need to be controlled for, but it *is* correlated with $x_{i,t}$. Unnecessarily including FE again hurts power by increasing p. In addition, the FE will have a larger positive impact on $\hat{R}_x^2$ now that $z_{i,t}$ does explain $x_{i,t}$. As the FE unnecessarily eliminate all variation in $x_{i,t}$ that can be explained by *any* group-level factor, this positive impact on $\hat{R}_x^2$ can be substantial. Thus, test power decreases, increasing the risk of a Type 2 error.

Case 3 is the opposite of case 2: z_g *is* a determinant of $y_{i,t}$ but is *not* correlated with $x_{i,t}$. Like in case 1, FE are not necessary to eliminate bias, and their inclusion increases p and likely $\hat{R}_x^2$, reducing power. At the same time, though, FE can *increase* power by reducing the unexplained variation in $y_{i,t}$

---

[21] We discuss the handling of singletons and no-variation groups further in section 3. Dropping observations also reduces $\sum (x_{i,t} - \bar{x})^2$ and $\hat{R}_x^2$, but the two effects offset one another.

($\sum \hat{\epsilon}_{i,t}^2$). The power-increasing reduction in $\sum \hat{\epsilon}_{i,t}^2$ can more than offset the power-decreasing effects of FE, in which case, FE can be used to increase test power, especially if the number of observations is large relative to the number of groups.[22] Finer FE (e.g., year-month instead of year) can also be used, but at some point the power-decreasing effects of finer FE will more than offset the power-increasing effects. Also, using increasingly fine FE can exacerbate concerns about singletons and no-variation groups.[23] We discuss corresponding diagnostics in section 3.4.1.

Lastly, we consider the case 4 in which z_g is both a determinant of $y_{i,t}$ and correlated with $x_{i,t}$. This case is arguably the most common case in empirical accounting research. FE are necessary to avoid omitted variable bias, but including FE comes with all of the power-increasing and -decreasing effects of the prior three cases. The net effect could be a power reduction and Type 2 error. The alternative, though, is to omit FE and risk a Type 1 error (a false positive).

Although including FE in case 4 can generate a Type 2 error, Type 2 errors are typically less concerning than Type 1 errors, for several reasons. First, conservative research designs and inferences provide a corrective force for the publication process's inherent bias toward overrejecting the null (e.g., Bloomfield, Rennekamp, and Steenhoven [2018]). Second, failing to reject the null due to limited residual variation in our X of interest does not necessarily indicate that X has no impact on Y. It merely suggests that, given all the other plausible first-order explanations that overlap with our variation (or explanation), we do not have sufficient *independent* variation to make strong inferences about the isolated impact of our X of interest. Such honest recognition about the shortcomings of the data should lead us to search for better settings and cleaner variation to credibly test our theory or effect of interest. It should not be used as an excuse to neglect other explanations (i.e., omit controls or FE).

In sum, we typically face a tradeoff when using FE: They often simultaneously reduce both bias and power. In the usual case in which we are

---

[22] deHaan, Li, and Zhou [2023a], for example, examine employee job searches (Y) following earnings announcements (X), and use firm-quarter FE to isolate variation in search during earnings announcement weeks relative to the trailing weeks within the same firm-quarter. The firm-quarter FE eliminate substantial variation in Y across firms and quarters, but are orthogonal to X given that each firm-quarter has exactly one earnings announcement. Thus, the firm-quarter FE in deHaan, Li, and Zhou [2023a] do not reduce the variation in the X of interest, so solely have the effect of increasing test power.

[23] Random effects models can also be used in cases where z_g is correlated with Y but uncorrelated with $x_{i,t}$. Random effects models are uncommon in accounting, likely because our $z_{i,t}$ variables are often thought to at least somewhat correlate with $x_{i,t}$. Also, Angrist and Pischke [2008] argue that the assumptions required of random effects models are too strong relative to other approaches. More recently, however, extension such as correlated random effects (or coefficients) models, pioneered by Mundlak [1978] and Chamberlain [1984], have gained popularity as an approach to uncover and account for treatment-effect heterogeneity (e.g., Suri [2011], Cabanillas et al. [2018], de Chaisemartin and Lei [2023]). See section 2.4 for further discussion of treatment-effect heterogeneity for FE models.

concerned about correlated omitted variables (case 4), we tend to include FE and err on the side of a Type 2 error. Sometimes FE may not be used to reduce bias but only to improve power (case 3). In other cases, FE are unnecessary but reduce power, so should be avoided (cases 1 and 2). In practice, though, it is impossible to conclusively differentiate between cases 4 versus 1 and 2, so we must rely on researcher judgment and knowledge of the specific setting and context.[24]

2.4 EXTENSIONS AND ADDITIONAL CONSIDERATIONS

This section introduces extensions and additional considerations. Corresponding diagnostics and remedies are provided in section 3. Casual readers may prefer to skip directly to section 3.

2.4.1. Two-Way FE in DiD Models. In a simple DiD design, a subject- or cohort-level indicator takes the value of one for groups that are affected by a treatment, and zero for control groups: $Treat_i = \mathbb{1}_1\{Treat_i = 1\}$. A time-varying indicator takes the value of one for periods after the treatment: $Post_t = \mathbb{1}_2\{Post_t = 1\}$. The X of interest is an interaction between the two: $x_{i,t} = Treat_i \times Post_t$. To simplify the following discussion, we will continue to use firms (i) as subjects over years (t):

$$y_{i,t} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 Post_t + \beta\left(Treat_i \times Post_t\right) + \epsilon_{i,t}. \qquad (27)$$

The DiD coefficient, $\hat{\beta}$, estimates the change in Y of the treated firms from the pre- to the posttreatment period, above-and-beyond the concurrent change of control firms. Specifically, $\hat{\beta}$ abstracts from pretreatment-level difference in Y across treatment and control firms (captured by α_1) and from pre/post differences in Y for the control firms (captured by α_2).[25] The constant (α_0) captures the level of the left-out category, that is, the control firms in the preperiod.

A "generalized DiD" model replaces the single indicators for *Treat* and *Post* with firm and year FE. Doing so can improve specificity by allowing each individual firm and year to have its own intercept, and can provide greater flexibility in settings with multiple treatment cohorts or dates.[26] For example, whereas α_1 in equation (27) accounts for the overall average difference in Y between treatment and control firms in the pretreatment

---

[24] Early examples in the accounting literature discussing the efficiency implications of FE can be found in Beaver et al. [1989] and Liu and Ryan [1995] in the context of capital-market reactions to banks' financial-reporting information.

[25] See section 4 of Roberts and Whited [2013] for a detailed introduction to DiD models and a discussion of the close relation between DiD and first-differences specifications.

[26] The literature on the adoption of International Financial Reporting Standards (IFRS) demonstrates the progression of DiD designs in accounting (e.g., Brüggemann, Hitz, and Sellhorn [2013], Leuz and Wysocki [2016], De George, Li, and Shivakumar [2016]). Earlier studies often use simple DiD designs around a common implementation date (e.g., in the EU in 2005). Later studies use generalized DiD designs to exploit the staggered adoption of IFRS around the globe, and in doing so replace the coarse *Treat* indicator with country or firm FE.

TABLE 3
Fixed Effects and Difference-in-Differences (DiD)

Panel		X Matrix Excerpt			Firm-Level De-meaning		Year-Level De-meaning		Residual
Firm	*Year*	$\mathbb{1}_i$	$\mathbb{1}_t$	$x_{i,t}$	$\bar{x}_i$	$\tilde{x}_{i,t} = x_{i,t} - \bar{x}_i$	$\bar{\tilde{x}}_t$	$\hat{x}_{i,t} = \tilde{x}_{i,t} - \bar{\tilde{x}}_t$	$\hat{\epsilon}^x_{i,t} = \hat{x}_{i,t}$
1	1	1	0	0	0.5	−0.5	−0.25	−0.25	−0.25
1	2	1	1	1	0.5	0.5	0.25	0.25	0.25
2	1	0	0	0	0.0	0.0	−0.25	0.25	0.25
2	2	0	1	0	0.0	0.0	0.25	−0.25	−0.25

The table presents the *X* matrix for a simple DiD model with two firms (one treated, one control firm) and two observations (one preperiod, one postperiod observation) plus the successive de-meaning steps implied by the two-way FE / DiD model. For simplicity, we do not show the inclusion of a constant (which captures the level of the control firm in the preperiod) here. $\mathbb{1}_i$ denotes the firm FE, which is equivalent to a treatment indicator ($Treat_i$) in this simplified setup with two firms. $\mathbb{1}_t$ denotes the year FE, which is equivalent to a postperiod indicator ($Post_t$) in this simplified setup with two periods. $x_{i,t}$ is the DiD treatment (i.e., $Treat_i \times Post_t$). $\bar{x}_i$ is the firm-level mean of the treatment ($x_{i,t}$). $\tilde{x}_{i,t}$ is the firm-level de-meaned treatment ($x_{i,t} - \bar{x}_i$). $\bar{\tilde{x}}_t$ is the year-level mean of the firm-level de-meaned treatment ($\tilde{x}_{i,t}$). $\hat{x}_{i,t}$ is the year- and firm-level de-meaned treatment ($\tilde{x}_{i,t} - \bar{\tilde{x}}_t$). $\hat{\epsilon}^x_{i,t}$ is the residual of a regression of the treatment ($x_{i,t}$) on the firm and year FE. It is equivalent to the doubly de-meaned treatment ($\hat{x}_{i,t}$).

period, firm FE allow each individual firm to have its own average *Y* before the treatment, captured by α_i^1:

$$y_{i,t} = \alpha_i^1 + \alpha_t^2 + \beta x_{i,t} + \epsilon_{i,t}. \tag{28}$$

In this design, each firm has its own intercept and the main effects of the $Post_t$ and $Treat_i$ indicators are subsumed by the time and firm FE.[27]

By design, the *X* variable of interest, $x_{i,t}$, is equal to zero for all observations for the control firms in DiD models; that is, the control firms are no-variation groups per section 2.1.4. Control firms therefore only (but importantly) contribute to estimating $\hat{\beta}$ by helping to estimate the counterfactual trend that the *Y* of treated firms (by the parallel trends assumption) would have followed had they not received the treatment. As always, control groups must be carefully selected to ensure that they are good counterfactual comparisons for the treatment groups.

To better understand the intuition of how no-variation control firms contribute to $\hat{\beta}$, we can look at an example of two firms and two years in table 3. Firm 1 receives a treatment in year 2, so has $x_{i,t}$ equal to zero in year 1 and one in year 2. Firm 2 is the control, so has $x_{i,t}$ equal to zero in both years (i.e., is a no-variation group). After applying firm FE to deduct firm-level

---

[27] Many studies explicitly note that the main effects in their DiD designs are subsumed by the chosen FE structure (e.g., Jayaraman and Kothari [2016], Amiram et al. [2018]). If the main effects of $Post_t$ and $Treat_i$ in equation (28) are not subsumed, then the *t* and/or *i* FE are improperly defined. The time FE must form a linear combination of *Post*, such that *Post* does not vary within any time grouping. For example, in a firm-quarter panel, year FE will not subsume *Post* if the treatment begins in the middle of a year. Similarly, in a generalized DiD model with cohort instead of firm FE, the cohort FE must perfectly differentiate between treatment and control firms, else *Treat* will not be subsumed.

means, the middle columns of table 3 show that the resulting $\tilde{x}_{i,t}$ for firm 2 is still zero in both years. After additionally applying year FE to deduct the year-level means, the right columns show that the double-de-meaned $\hat{x}_{i,t}$ for firm 2 is now 0.25 in year 1 and -0.25 in year 2. Thus, the *residualized* $x_{i,t}$ does vary within both the treated and the control firm. Accordingly, as the treatment status of the treated firm varies over time, so does the *relative* treatment status of the control firm, $\hat{x}_{i,t}$. In this example, $\hat{x}_{i,t}$ changes by $+0.5$ for the treatment firm from year 1 to 2, and by -0.5 for the control firm. The relative difference is therefore $1 (= +0.5 - (-0.5))$.

The above example illustrates that, in two-way FE regressions, even observations without variation in one of the dimensions still contribute to the identification of the coefficient on the X of interest. The same holds when the other control is not a second set of FE but is instead a standard linear control. As long as the linear control varies within the group FE, the observations without variation in the X of interest still indirectly contribute to the estimated coefficient on X by helping to improve the estimated coefficient on the other control. The better the control coefficient estimate, the better our estimate of the coefficient of interest.[28]

The insights from the simple model with one treatment date generalizes to the case of multiple treatment cohorts. They also apply to other multidimensional data (e.g., country-industry instead of firm-time data). By controlling for any differences at the level of the main dimensions (e.g., country and industry), multidimensional FE can comprehensively account for first-order factors correlated with the X of interest. In doing so, FE home in on the more subtle, and hopefully cleaner, variation due to the interaction of the distinct dimensions.[29]

2.4.2. Heterogeneous Effects in Generalized DiD Models. Although the generalized DiD can exploit multiple treatment cohorts and staggered treatment dates to improve identification, staggered treatment dates can complicate the interpretation of the $\hat{\beta}$ coefficient of interest. Specifically, complications arise when treatment effects are "heterogeneous," meaning that treatment effects differ across cohorts or over time. For example, treatment effects may differ for the types of firms that get treated early versus late,

---

[28] For this reason, for example, Daske et al. [2008] include never-treated control firms from countries such as the United States (U.S.) in their generalized DiDs regressions to study the impact of IFRS adoption. As long as the relations between Y and control variables (e.g., time effects and other Xs) are similar in the U.S. as in the other countries that adopt IFRS, including the U.S. helps estimating a better conditional expectation for Y. If, by contrast, U.S. firms were thought to experience distinct changes (e.g., the introduction of SOX) or react differently to common changes (e.g., a global financial crisis), they should be excluded from the study.

[29] In this vein, multidimensional FE are frequently used in Bartik or shift-share designs, which exploit the differential change of subjects (e.g., local industries) with distinct predetermined characteristics (e.g., industrial structure) in response to aggregate trends (e.g., in industry-wide employment) (e.g., Goldsmith-Pinkham, Sorkin, and Swift [2020], Breuer [2022]).

or may continue to become stronger for a few years after the treatment takes place.

In case of heterogeneous treatment effects, the DiD coefficient, $\hat{\beta}$, captures a weighted average of various two-by-two DiD comparisons (e.g., never vs. early, early vs. late; de Chaisemartin and D'Haultfœuille [2020], Goodman-Bacon [2021]). This weighted average can result in an estimator that is strongly biased relative to other sensible summaries of the heterogeneous effects. The underlying issue is that, in case of staggered treatment dates, a given group can be a treated cohort in one period and a control cohort in a later period. If the treatment effect varies across cohorts or over time, this switching can lead to some cohorts receiving negative weights in the weighted-average DiD estimator, because the weights are a function of the residualized treatment indicator. Table 3, for example, shows that the residualized treatment indicator can be negative. In table 3, this is not a problem because it is only negative for the treated firm in the preperiod. If, by contrast, it is negative for a treated firm in the postperiod (e.g., because an early treatment cohort is used as a control later on), the negative weight can flip the sign of the weighted-average treatment estimator. This, for example, can be the case if the cohort with the negative weight exhibits a stronger response than others. In this case, the weighted-average treatment effect can take a negative sign even if all cohorts experience a positive treatment effect (or vice versa).

The issue of heterogeneous effects is complex and extends to a variety of empirical settings. For further discussion specifically in the context of DiD models with two-way FE, see Barrios [2021] and Baker, Larcker, and Wang [2022].

2.4.3. Spillovers in FE Models. Spillover effects arise if a treatment affects both the treated and control firms (e.g., Glaeser and Guay [2017], Berg, Reisinger, and Streitz [2021]). Spillovers bias the treatment-effect estimate upwards (downwards) if they affect control firms in the opposite (same) direction as the treatment effect on the treated firms.[30] Spillover-related biases are particularly plausible in disclosure studies, given that we know that one firm's disclosures are often relevant for economically connected peer firms (e.g., Foster [1981]). Spillover biases tend to become more severe when treatment and control firms are more closely connected to one another; for example, if a treatment and control firm are in the same industry and region, it is more likely that the actions (e.g., disclosures) of the treatment firm spillover to affect the control firm.

FE can aggravate spillover bias by restricting analyses to more comparable treatment and control firms. Although we generally prefer more comparable control firms, in the case of local spillovers, restricting to local control

---

[30] Spillovers violate the stable unit treatment variance assumption (SUTVA) and therefore distort $\hat{\beta}$ in both cross-sectional comparisons of treated and control firms as well as in DiD designs.

firms reduces concerns about bias due to confounding differences but potentially adds bias due to an indirect impact of the treatment on the control firms (Berg, Reisinger, and Streitz [2021]).

2.4.4. Interacted FE. FE can be interacted with other Xs to allow the relation between Y and Xs to differ across FE groups.[31] Such heterogeneity in the relation between X and Y is commonplace in accounting data and research. Popular accruals models, for example, allow for differences in accruals processes across industry-years or even firms (e.g., Jones [1991], DeFond and Jiambalvo [1994], Chen, Hribar, and Melissa [2018], Breuer and Schütt [2023]). Similarly, Ball and Nikolaev [2022] allow for differences across firms in how various determinants map into firms' earnings. A given change in sales ($z_{i,t}$), for example, can increase earnings ($y_{i,t}$) more for one type of firm than another, depending on the firms' business models and competitive positions. Firm FE by themselves only account for level difference in Y so do not account for such heterogeneous relationships between Y and its determinants. Instead, we have to additionally interact the firm FE ($\mathbb{1}_i$) with the respective determinant ($z_{i,t}$):

$$y_{i,t} = \alpha_i + \beta x_{i,t} + \gamma_i(z_{i,t} \times \mathbb{1}_i) + \epsilon_{i,t}. \qquad (29)$$

This interacted FE regression estimates a separate γ_i coefficient for each firm.[32]

Interacted FE play a key role in cases when the X of interest is a differential slope. Those cases include studies examining earnings-response coefficients (ERCs), investment sensitivities, and, more generally, the effects of moderating variables (e.g., Jollineau and Bowen [2023]). A study, for example, might examine whether ERCs differ after a treatment, and do so with a regression of returns (CAR) on earnings surprise (UE) and its interaction ($UE \times Treat$). The coefficient of interest is that on $UE \times Treat$. Firm FE would account for firm-specific–level differences in CAR and UE, but an interaction of firm FE with UE is required to allow each firm to have a unique returns-earnings relation.[33]

Interacting high-frequency (e.g., firm) FE with X can consume many degrees of freedom, exacerbate issues stemming from FE removing useful variation in X (section 2.3), and increase the number of sample singletons

---

[31] See deHaan et al. [2023b] for deeper discussion on interacting controls, including FE, with X variables of interest.

[32] Interacting firm FEs with $z_{i,t}$ is akin to running regressions by firm and taking the average $\hat{\beta}$. In a model with additional determinants, though, the firm FE would also need to be interacted with each determinant to be equivalent to running regressions by firm.

[33] The same issue applies in DiD tests of differential slopes, in which the coefficient of interest is a triple interaction $UE \times Treat_i \times Post_t$. In those cases, we include all main effects and two-way interactions in the regression specification to interpret the triple interaction as the incremental slope of interest. If we use FE to substitute the main effects of $Treat$ and $Post$, full control also requires those FE to be interacted with UE.

and no-variation FE groups (section 2.1.4). Interacting high-frequency FE is often not practicable, in which case, one remedy is to use coarser groups. In the context of an ERC model, for example, interactions between UE and firm FE might be replaced with interactions of UE and treatment cohort or industry FE (e.g., Gipper, Leuz, and Maffett [2019], Blanke-spoor et al. [2019]). Coarser groupings may not control confounding heterogeneity, though, so it is important to be transparent about such potential shortcomings.

An alternative remedy for impracticable high-frequency FE interactions is to attempt to impose more structure on the coefficient distribution. Bayesian hierarchical modeling, for example, can account for rich firm-level heterogeneity while reducing concerns about limited within-group data (Breuer and Schütt [2023]). Instead of estimating coefficients only using within-group data, it also uses information about the distribution of coefficients across groups. It thus does not discard groups with limited data or try to estimate many coefficients from few observations of a given group. Instead, it combines within- and across-group data following Bayes' rule, putting more emphasis on cross-group information for groups with limited data (and vice versa). Whether coarser groupings or more structure is preferable depends on the extent to which we worry about the remaining within-coarse-group heterogeneity vis-à-vis the accuracy and transparency of the additional structure.

2.4.5. Dynamics in FE Models. An obvious problem is that FE, which account for constant time-invariant factors, cannot easily account for within-group dynamic changes over time. Industry FE, for example, cannot account for changes in the composition of firms listed on stock markets, which can bias studies of listed firms' accounting quality due to entry of riskier firms over time (Fama and French [2004], Srivastava [2014]). Using finer FE can sometimes help (e.g., firm FE would help abstract away from industry composition changes), but even firm FE are unlikely to fully purge unwanted heterogeneity as panels lengthen and subjects evolve over time (Millimet and Bellemare [2023]). Hence, even finer FE (e.g., firm-decade FE) may be needed when dealing with highly variable subjects or long sample periods.[34]

A more subtle but likely pervasive problem is that estimates of parameters of interest in linear models that include subject FE can be strongly biased in the presence of dynamics or feedback effects (e.g., Nickell [1981], Imai and Kim [2019]). Current values of firms' size, financing, and accounting choices, for example, are typically related to past values of those variables and, similarly, past unobservable factors. In this case, firm FE can introduce correlations between the de-meaned variables and unobservable factors, re-

---

[34] Bai [2009] provides an approach to address issues resulting from heterogeneous responses to common time-varying factors (e.g., the state of the economy) through interactive FE (see also section 2.4.4)

sulting in biased estimates. The bias arises because in-sample means combine and reflect the (recent) dynamics in firms' variables, which are driven by observable but also unobservable factors. By deducting those in-sample firm-level means from each observation, the de-meaned values of X of each period are implicitly related to each other and the past, current, and future error terms (i.e., unobservable factors). This problem is particularly acute in panels with relatively short time series, where in-sample firm-level means are strongly affected by firms' recent dynamics. The absence of dynamics, hence, is an implicit but often neglected assumption of models with subject (e.g., firm) FE. Specifically, "strict exogeneity" requires that past values of X do not directly influence current values of Y, and past values of Y do not affect current values of X (Imai and Kim [2019]).

2.4.6. Dependent Observations and Clustered Standard Errors. Unlike assumed in section 2.3, regression errors are typically not independent and identically distributed, but instead often exhibit substantial temporal and/or cross-sectional dependence. The error term $v_{i,t} = \epsilon_{i,t} + \delta z_g$, for example, is correlated within group g due to the omitted variable z_g. In this special case, dependence is completely fixed within groups, so the group FE eliminate bias of both the coefficient estimate and the standard-error estimate.[35]

In many cases, though, error dependence varies even within FE groups.[36] Consider, for example, a time-varying, autocorrelated omitted variable ($z_{g,t} = z_g + \rho z_{g,t-1} + e_{i,t}$). This type of within-group heterogeneity prevents group FE from fully accounting for the dependence of the error term.

In some cases, FE can even induce dependence (see section 2.4.5). Consider, for example, a model that is properly specified and has independent errors in the absence of FE. Including FE deducts the group-level mean of the error term ($\bar{\epsilon}_g = \frac{1}{N_g} \sum \epsilon_{i,t} \; \forall \; i, t \in g$) from the already independent error: $\epsilon_{i,t} - \bar{\epsilon}_g$. In the case of small group sizes, deducting the group's mean error links all group-level errors with each other, and therefore can induce group-level dependence. For larger groups, this issue is less pressing because each individual error term tends to have a negligible impact on the group-level mean.

As a result, FE are typically used in conjunction with approaches to adjust standard errors for dependence (e.g., heteroskedasticity and autocorrelation consistent (HAC) estimators; Conley, Goncalves, and Hansen [2018]). A popular approach is clustered standard errors, which account for arbitrary dependence among observations in a given cluster. Clustered stan-

---

[35] Beaver et al. [1989], for example, discuss the usefulness of FE in addressing cross-sectional correlation in returns and serial correlation in banks' financial-reporting information.

[36] Abadie et al. [2013] highlight that, when using FE groupings at the cluster level, heterogeneity in the treatment effect is a requirement for needing a clustering adjustment. As noted in section 2.4.1, such heterogeneity is frequently observed, complicating both the interpretation of the coefficient of interest and its standard error calculation.

dard errors assume independence across clusters, so we typically want to use relatively broad cluster groupings to allow for more dependence, although having too few groups can also cause problems.[37] Notably, this contrasts with the choice of groupings for FE, for which we often prefer finer groupings. FE groups should ideally be "nested" within the clusters (i.e., do not cross the clusters) to help eliminate any effects of FE-induced dependence discussed in the prior paragraph.

Following Petersen [2008], we illustrate clustered standard errors for the case of a balanced firm-time panel data. We can cluster standard errors at the firm level to account for serial dependence in such data. Under the simplifying assumption of identically distributed error terms, the clustered standard errors are given by:[38]

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\left(\frac{1+(T-1)\rho_x\rho_\epsilon}{MT-p-1}\right)\left(\frac{\sum \hat{\epsilon}_{i,t}^2}{\sum(x_{i,t}-\bar{x})^2(1-R_x^2)}\right)}, \tag{30}$$

where M is the number of firms; T is the number of time periods (such that $N = MT$); ρ_x is the dependence of the residual variation of X (after accounting for controls and FE); and ρ_ϵ is the dependence of the error term.

The equation above shows that the clustered standard errors adjust the effective sample size downward *if* there is dependence in both the residual variation of X (ρ_x) and the error term (ρ_ϵ) within a given cluster. The higher the dependence, the lower the effective sample size used. In the extreme case of perfect dependence (e.g., duplicated observations: $\rho_x = \rho_\epsilon = 1$), the effective sample size collapses to M, the number of clusters (here, firms):[39]

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\left(\frac{1}{M}\right)\left(\frac{N}{N-p-1}\right)\frac{\sum \hat{\epsilon}_{i,t}^2}{\sum(x_{i,t}-\bar{x})^2(1-R_x^2)}}. \tag{31}$$

Accordingly, the number of clusters is a key determinant of the clustered standard error, so it is important to exclude clusters that do not contribute to estimating $\hat{\beta}$. In particular, this is why it is important to drop

---

[37] Conley, Goncalves, and Hansen [2018] suggest using a sample-splitting approach akin to Fama and MacBeth [1973] instead of clustered standard errors to account for dependence in case we have (i) a few broad clusters, (ii) FE nested within the clusters, and (iii) variation in the X of interest within each cluster.

[38] The special case of identically distributed errors facilitates comparison to the prior standard error formula (equation (23)). Clustered standard errors do not require homoskedasticity to be valid, but rather explicitly allow for heteroskedasticity by calculating the covariance between the (residualized) X and the errors.

[39] Equation (31) obtains for the special case of $\rho_x = \rho_\epsilon = 1$ and a balanced panel. It illustrates the intuition why M, the number of clusters, rather than N, the number of observations, is the relevant sample size in dependent data. For the general clustered standard error formula with the appropriate degree of freedom adjustments in the presence of FE and an unbalanced panel (e.g., giving rise to singletons), please refer to Cameron and Miller [2015] and Correia [2015].

singletons in FE regressions, as detailed in Correia [2015] and discussed in section 3 below.

2.4.7. FE in Nonlinear Models. FE can create particular problems in nonlinear models, for example, binary outcome models such as Logit. In a firm-time panel, for example, the maximum-likelihood estimator (MLE) is typically biased and inconsistent when using firm FE. This "incidental parameters problem" is particularly severe if the time dimension is small and fixed (e.g., Neyman and Scott [1948], Lancaster [2000]). Accordingly, the use of FE in nonlinear models requires special attention, and is a topic generally outside our scope.

A common strategy for avoiding issues with FE in nonlinear models is to use OLS even for discrete outcomes. Such linear probability models can be suitable if we are interested in estimating the average marginal treatment effect instead of recovering a structural choice parameter or making predictions that respect the boundaries of the outcome (Angrist and Pischke [2008]). It is important to realize though that, in nonlinear models, we essentially have to deal with heterogeneous treatment effects by construction. As discussed in section 2.4.2, those heterogeneous effects can result in hard-to-interpret average effects. Accordingly, linear probability models with FE can avoid some issues of FE in nonlinear models but come with their own (related) problems.

The use of FE in nonlinear models is an active area of research among econometricians, and recent advances are starting to make the use of FE in nonlinear models more practicable. For example, Correia, Guimarães, and Zylkin [2020] develop an implementable Poisson model with high-frequency FE.

2.4.8. Measurement Error. FE can eliminate measurement error if the error is relatively constant within FE groupings, or can exacerbate measurement error if the error is more pronounced within FE groupings than across those groupings. For example, measurement error in proxies for 10-K readability may vary year-over-year despite substantial stickiness in the true readability of firms' 10-K filings, so within-firm changes in readability proxies are plausibly noisier than are cross-sectional differences. This issue is reminiscent of the overdifferencing concern in Cochrane [2012] and is further discussed in Jennings et al. [2024].[40] In general, the directional bias caused by measurement error is highly context specific and difficult to sign. The same holds for the effect of FE on measurement error bias. As a result, in the presence of both omitted variable and measurement error concerns, our choice of the appropriate FE structure requires careful consideration

---

[40] Jennings et al. [2024] illustrate a case in which firm FE are unnecessary because they are independent of Y, and show that erroneously including FE can exacerbate bias from measurement error and generate false-positive results.

of the relative importance of the two potential biases, the relative impact of FE on those distinct biases, and how those biases interact.[41]

3. Guidance on Diagnostics and Reporting

Our above discussion highlights that FE can help draw more credible inferences but do not come without drawbacks and limitations. Here, we propose good practices for the informed use and transparent reporting of FE.

3.1 RESEARCH DESIGN

3.1.1. Motivation for Using FE. FE can help the identification of cause-and-effect relations. They can purge unwanted variation and home in on variation that we can interpret through the lens of our theory of interest. In particular, FE can be useful when unwanted variation varies across identifiable groups and threatens to induce bias (e.g., because it is correlated with X and Y) or to weaken test power (e.g., because it creates residual variation in Y). What qualifies as "unwanted" variation and at what level the unwanted variation manifests (e.g., at the firm or industry level) are matters of ex ante reasoning.

Papers should clearly articulate what variation is needed (or "wanted") to test their research questions and how their research design choices, including their FE structure, help isolate this variation. This discussion should be informed by theory and institutional insights. Theory, for example, can help spell out the exact counterfactual or variation that papers seek to isolate, and can point to alternative explanations that papers need to control for. Institutional insights, in turn, can uncover relevant settings that provide the variation of interest and allow abstracting from other forces (i.e., "unwanted" variation). The use of theoretical and institutional insights for identification of cause-and-effect relations is central to what is commonly referred to as the *design-based approach* to empirical research (e.g., Card [2022], Leuz [2022]).[42]

Papers should motivate their FE choices in the context of their broader research designs. In combination with theory, institutional insights, and

---

[41] Garber and Klepper [1980] note that, by the theorem of the second best, addressing one source of bias in the presence of multiple specification problems may worsen our inferences rather than help it.

[42] We may, at times, find it difficult to differentiate between various possible research designs and/or to derive the best design based on theory. This difficulty tends to reflect the high degree of uncertainty we face often in empirical accounting research due to the combination of abstract theories and highly contextual data. In this case, we may want to take a more modest, exploratory approach, which describes patterns in the data (e.g., to inform future theory development) rather than an approach that aims to test specific theories and causal hypotheses (e.g., Wasserstein, Schirm, and Lazar [2019], Lavine [2019], Kallapur [2022], Breuer [2023]).

other methods (e.g., DiD designs), FE can allow papers to limit the possibility that other confounding forces are at play and strengthen the credibility that the reported results reflect the theory of interest.[43] Including FE without careful thought, by contrast, is unlikely to increase confidence in a paper's inferences, even if the FE are fine or high dimensional.

3.1.2. Discussion of Identifying Within-FE Variation. Papers should explicitly discuss the extent and sources of identifying variation in Y and X that remain after accounting for FE.[44] The remaining within-FE variation should be expected to be cleaner or more interpretable than the raw variation. If there are not good reasons to expect that the within-FE variation is cleaner, FE should likely be avoided.[45]

3.2 SAMPLE CONSTRUCTION

FE can change which and how observations contribute to estimating $\hat{\beta}$ (section 2.1.4). The following help transparently report the effective sample used in FE regressions[46]:

3.2.1. Singletons. Singletons have just one observation within a FE group, do not affect coefficient estimates, and can bias standard errors, so should be dropped before running a FE regression. Singletons must be identified iteratively in regressions with multiple FE groupings, meaning that one first drops singletons for the first FE grouping, and then drops singletons for the second FE grouping, and then repeats until no singletons remain.[47] In

---

[43] To study the impact of securities regulation, Christensen, Hail, and Leuz [2016], for example, not only exploit cross-country variation in the timing of regulatory reforms but also leverage the institutional insight that the reforms only apply to firms within each country that are listed on regulated exchanges. Including country-quarter FE therefore allows Christensen, Hail, and Leuz [2016] to alleviate concerns about confounding country-level events around regulation adoptions. In a similar vein, Granja [2018] uses state-year FE to exploit that reforms in bank-disclosure regulation not only varied in terms of timing across states but also applied to different types of banks within each state. Both studies provide detailed discussions how the institutional details inform their research design, including their FE structure, and help with identification.

[44] To identify the consequences of PCAOB inspections, Shroff [2020], for example, explicitly discusses and reports the within-country variation in PCAOB inspections.

[45] In this vein, we stress that within-firm temporal variation is *not* necessarily cleaner than cross-sectional variation, so firm FE specifications are *not* by default preferable to other specifications (e.g., Zhou [2001], Nikolaev and van Lent [2005]). For example, disclosure choices are sticky in theory and practice, so cross-sectional research designs may be more appropriate in many voluntary disclosure settings (e.g., Guay, Samuels, and Taylor [2016]). Similarly, the equilibrium effects of disclosures and accounting choices can take several years to achieve, in which case, cross-sectional designs are again likely appropriate (e.g., Blankespoor, deHaan, and Marinovic [2020], Breuer, Hombach, and Müller [2022]).

[46] Several of these statistics are generated by "SUMHDFE" in Stata, available on Ed deHaan's GitHub repository: https://github.com/ed-dehaan/sumhdfe.

[47] Iterating is required because dropping singletons in the second FE grouping can create new singletons in the first FE grouping. FE software such as Sergio Correia's reghdfe Stata package can iteratively drop singletons: http://scorreia.com/software/reghdfe

studies where FE are used in all models, it may be useful to drop obvious
singletons when constructing the sample (e.g., firms with only one obser-
vation in firm FE models). Singletons can differ across tests with different
available observations, though, so should still be identified and dropped
before each regression. The presence of numerous singletons can indicate
that the available data may not be granular enough to speak to your theory
and research design.

3.2.2. FE Groups with No Variation in the X of Interest. So-called "no-
variation FE groups" have multiple observations but no variation in the X
of interest, so only affect $\hat{\beta}$ by acting as control groups to help estimate
relations between Y and other determinants. Much like control groups in
matching and standard DiD designs, no-variation FE groups are only help-
ful if they share a similar data generating process as the groups that do have
variation in X. If they do, their inclusion in the sample improves test power
and specificity. In some cases such as DiD models, no-variation FE groups
are actually an essential component of the research design (especially in
the case of never-treated groups in staggered DiD). If no-variation groups
have dissimilar determinants of Y than do other group, though, their in-
clusion can bias $\hat{\beta}$.

Papers should report the number of no-variation groups and observa-
tions within the sample.[48] If a material number exists, researchers should
carefully consider the setting, what generates variation in the X of inter-
est, and whether no-variation groups are appropriate controls for groups
that do have variation in X. It is ultimately a judgment call as to whether
no-variation groups are good controls for other groups. This call needs to
be made based on theoretical and institutional considerations but several
analyses can help inform our judgment ex ante (e.g., based on pretreat-
ment data) or, at least, help assess the robustness of our estimates to the
respective sample selection choices ex post. These diagnostics are similar
to those used in matching and DiD designs:

- Similar to the covariate balance tests used in matching methods, ex-
amine descriptive statistics for no-variation groups versus other groups.
Significant differences in means can indicate that the groups may be
fundamentally different from one another.
- Regress Y on the controls and FE excluding the X of interest, sep-
arately for the no-variation groups and other groups. Significant dif-
ferences in coefficient estimates can indicate that the two subsamples
have different data generating processes.
- Run a regression with and without no-variation groups. If $\hat{\beta}$ is simi-
lar, then the no-variation groups are unlikely to be confounds (even

---

[48] A method for identifying no-variation groups is to calculate the standard deviation in X
within each group, and then tag observations with a standard deviation of zero. Singletons
and observations with missing values of Y and other determinants must be dropped first.

if results become less significant, which is expected due to decreased sample size). If $\hat{\beta}$ differs, then it is important to gauge which result is likely more correct.

If no-variation groups are dissimilar from other groups, then one option is to retain only those no-variation groups that are similar to other groups, for example through matching or balancing (e.g., Iacus, King, and Porro [2012], Roberts and Whited [2013], McMullin and Schonberger [2020]). Another option is to drop no-variation groups. Doing so will reduce test power and may render a model inestimable (e.g., in the case of DiD models), but these can be necessary consequences of not having good control observations.

3.2.3. FE Groups with No Variation in Y. FE groups with no variation in Y *do* directly contribute to estimating $\hat{\beta}$ in linear models.[49] Intuitively, if FE groups do have variation in X but do not have variation in Y, then a regression of Y on X and FE will produce $\hat{\beta} = 0$ for those groups. In cases where the Y could have varied within the FE group, this estimate can reflect that X does not affect Y. In cases where Y could not have varied (e.g., because it is not measured at a more granular level than the FE), the zero estimate, by contrast, arises even if X actually does affect Y.

Whether to drop groups without variation in Y often depends on whether it is interesting to examine only groups with variation. For example, roughly 70% of Compustat firms have zero R&D expense in recent years. In a study of the determinants of R&D expense, it plausibly makes sense to drop banks (which rarely have material R&D), but it likely would not make sense to drop consumer product firms (for which immaterial R&D is likely a strategic choice). Either way, it is important to be explicit about how sample selection choices affect the interpretation of results.

3.3 REPORTING AND INFERENCES

The inferences drawn from FE regressions should be interpreted in the context of the effective variation used in the estimation:

3.3.1. Reporting of Residualized Variables. Because FE implicitly transform variables into within-FE variation, standard descriptive statistics should be supplemented with descriptive statistics for each variable residualized to the FE. We recommend reporting both the raw standard deviation of each variable and the within-FE standard deviation of each variable. The smaller the within-FE standard deviation relative to the raw standard deviation, the more variation is purged by the FE. Examining additional descriptive statistics for the residualized variables is useful in cases where FE change a variable's distribution substantially (e.g., when FE convert a skewed variable to

---

[49] FE groups with no variation in Y have a more complex role in non-linear models. In some cases (e.g., Logit), they cannot contribute to estimating coefficients, whereas in other cases, they contribute only under specific conditions (e.g., FE groups with fixed nonzero values in a Poisson model; Correia, Guimarães, and Zylkin [2020]).

something more normally distributed), can make it clear which covariates are completely absorbed by the FE structure (because the within-FE standard deviation is zero), and can aid in the identification of within-group outliers (discussed more below).[50]

Observing that relatively little within-FE variation remains in the X of interest does not per se indicate that the FE structure is too fine, but it can raise several issues. First, the remaining within-FE variation may not be economically meaningful. Second, the within-FE variation may be too small to identify an effect that exists in truth (i.e., a Type 2 error). This is especially likely if the FE absorb a lot of variation in X that does not also correlate with omitted determinants of Y. Third, the remaining variation may be largely driven by measurement error. Fourth, it suggests caution in generalizing the results.

Observing that relatively little within-FE variation remains with Y is less concerning, and is often desirable to improve specificity and power. A criticism of FE explaining a lot of variation in Y is that the remaining variation is not economically meaningful. However, many accounting variables explain relatively little of the variation in Ys of interest (Lev [1989]). Given the choice between X increasing a regression R^2 from 1% to 2% in a model without FE versus 90% to 91% in a model with FE, the latter is likely preferable due to the reduced risk of omitted variables and potentially increased power. One way to be transparent about a regression's explanatory power coming from FE versus other regressors is to supplement the reported regression R^2 with the "within-R^2," which is the portion of the R^2 generated by the explanatory variables other than FE.

3.3.2. Interpreting the Magnitude of $\hat{\beta}$. Researchers often characterize the economic magnitude of $\hat{\beta}$ in terms of a distributional statistic, such as the standard deviation or interquartile range of X or Y. FE change the variation used to estimate $\hat{\beta}$, so change how we should make such characterizations.

In relation to X, a common way to characterize the magnitude of $\hat{\beta}$ is as follows: "a one-standard-deviation increase in X is associated with a $\hat{\beta} \times \sigma_x$ increase in Y," where $\hat{\beta}$ is replaced with the estimate value and σ_x is the sample standard deviation of X. Such characterizations are often used when a "one-unit" change in X has an unclear interpretation or is an unrealistically small or large change, in which case a one-standard-deviation change is more meaningful. Because FE regression estimates are based on the within-FE variation in X, the raw variation in X is no longer a meaningful unit of change for characterizing magnitudes. The economic magnitude of $\hat{\beta}$ should instead be characterized using the within-FE variation

---

[50] Within-FE distributional statistics can be calculated by first regressing each variable on the FE, and then calculating the distributional statistics of the residuals. An alternative method for characterizing remaining within-FE variation is as $(1 - R^2)$, where R^2 is from regressions of each variable on the set of FE. In the absence of singletons, the square root of $(1 - R^2)$ is equal the ratio of standard deviations. The two differ in the presence of singletons, which is why we recommend the ratio of standard deviations.

in X, which is never larger than the pooled sample variation, and is often much smaller.[51] For example, one can instead use the standard deviation of the residualized X. This approach avoids extrapolating the estimated effect above and beyond the variation in X that is actually used in the FE regression (Mummolo and Peterson [2018]).[52]

Some studies use the distributional properties of Y to characterize economic magnitudes; for example, "a one-unit increase in X is associated with a $\hat{\beta}$ increase in Y, which amounts to $\frac{\hat{\beta}}{\sigma_y}\%$ of one standard deviation of Y," where σ_y is the sample standard deviation of Y. The within-FE standard deviation in Y will tend to be smaller than the pooled standard deviation, so magnitudes will appear larger when characterized using within-FE variation in Y.

Whether or not to use the within-FE variation of Y depends on whether the within-FE variation in Y is an economically interesting benchmark. For example, imagine a randomized trial in which a model with firm FE estimates that treatment X increases firms' asset values by \$0.1 billion. The standard deviation of assets in a recent two-year CRSP/Compustat sample (winsorized) is roughly \$31 billion, so one could conclude that "the treatment increases firm value by an average of \$0.1 billion, which amounts to $0.1/31 = 0.03\%$ of one standard deviation of asset value." The same data set shows that the within-firm standard deviation of asset value is just \$2 billion, so one could characterize the same result as "the treatment increases asset value by \$0.1 billion, which amounts to $0.1/2 = 5\%$ of one within-firm standard deviation of asset value." These two characterizations are fundamentally different from one another: The former is about variation we observe in a population, whereas the latter is about within-firm year-over-year changes in value. The magnitudes are also quite different from one another, with the latter appearing much larger. Whether the latter is appropriate depends on whether you view the within-firm variation in asset value as an economically meaningful phenomenon to explain.[53]

The distinction between these two scenarios is subtle but important. We characterize X in terms of distributional statistics to provide better insights about a realistically sized change in X that we observe in the data (i.e., within our FE structure). Characterizing X in terms of unrealistically large changes risks extrapolation bias so should be avoided. We characterize Y in

---

[51] A similar issue occurs in matching, but is easier to identify. For example, a researcher should not use the prematched sample standard deviation in X to characterize an effect size, but should instead calculate the standard deviation only within the subset of observations retained in the matched sample.

[52] Recent examples of studies in the accounting literature using the within-FE variation to assess economic magnitudes include, for example, Glaeser and Omartian [2022] and Sani, Shroff, and White [2023].

[53] Other studies characterize $\hat{\beta}$ in terms of a percentage of the mean of Y. In models with FE, the residualized mean of Y is zero by construction, so such characterizations must be done using the raw variable's mean. In doing so, it is important to limit the sample to the observations effectively contributing to the estimation of the coefficient of interest.

terms of distributional statistics to provide better insights about the importance of a one-unit change in X for the phenomenon we are examining, so the researcher is freer to choose the basis for comparison. Whatever distributional properties are used, it is important to be clear about the motivation and effects thereof.

3.3.3. Generalizability. As with many econometric methods, FE involve a tradeoff between improving identification by removing unwanted variation (i.e., improving internal validity) and potentially sacrificing generalizability to the unexamined portion of the population (i.e., undermining external validity). Specifically, because FE use only within-group variation to estimate effects, empirical results cannot directly speak to whether similar cause-and-effect relations exist across FE groups.[54]

Although generalizability concerns are not a valid reason for omitting FE that are otherwise thought to be necessary, it is important to discuss reasons (e.g., based on theory or institutional knowledge) why a study's within-group estimates may or may not generalize across groups or to other settings. As always, transparently describing the context of the study's empirical results and potential generalizability to other settings will help readers understand the need for further research. Concerns about the generalizability of FE estimates (e.g., Armstrong et al. [2022]) call for further research in other settings (i.e., replications and reexaminations; Bloomfield, Rennekamp, and Steenhoven [2018]) and for caution in drawing inferences from any individual study (e.g., McShane et al. [2019b], McShane and Gelman [2022]).

3.4 OTHER ISSUES, DIAGNOSTICS, AND ROBUSTNESS TESTS

3.4.1. Finer FE. Finer FE (e.g., year-month instead of year FE) can increase credibility by further reducing concerns about potentially unwanted variation (e.g., Amir et al. [2016]), but can also exacerbate all the issues we have discussed about power, contributing observations, and generalizability. Accordingly, finer FE are not always preferred for main specifications, but can nevertheless be informative as a robustness test.[55] If finer FE hurt power but leave $\hat{\beta}$ largely unaffected, we can take some comfort in that the

---

[54] In studying auditor effects, Cameran, Campa, and Francis [2022], for example, discuss the tradeoff inherent in using client FE to separate auditor characteristics from client characteristics. Client FE help control for potentially confounding client characteristics, but also limit the estimation sample to a few clients with auditor switches. To allow readers to assess the generalizability of the study's findings, Cameran, Campa, and Francis [2022] report the share of observations without switches and also report results without client FE.

[55] Daske et al. [2013], for example, use country, industry, and year FE in their main specification to test for IFRS adoption effects. In robustness tests, they substitute the country and industry FE with firm FE, which are finer partitions of (or nested in) countries and industries. In a similar vein, Dekeyser et al. [2021] use location, industry, and year FE in their main specification. To assess robustness to finer FE, they also report specifications with combinations of those FE (e.g., location-industry or location-industry-year FE).

additional variation eliminated by the finer FE is unlikely to be a significant source of omitted variable bias. If finer FE significantly affect $\hat{\beta}$, by contrast, we need to carefully consider the possibility that omitted factors may confound our main estimates, especially if the corresponding power of the estimate does not dramatically decline.[56] Those diagnostics can be informative, ex post, about the credibility of our estimates and their robustness with respect to other potentially unaccounted sources of variation. We stress though that those statistical diagnostics do not allow researchers to choose their optimal FE level. As discussed before, this choice needs to be made ex ante, informed by theory and institutional insights.

3.4.2. Heterogeneous Effects in DiD Models. Heterogeneous effects can bias DiD models estimated via two-way FE specifications (section 2.4.2), so should be carefully evaluated when using such models. Barrios [2021] and Baker, Larcker, and Wang [2022] summarize methods for gauging potential biases by exploring the weights assigned to various treatment cohorts and the dispersion of the cohorts' treatment effects. Observing few negative weights and limited heterogeneity in treatment effects alleviates concerns about bias in two-way FE specifications.[57] Otherwise, modified DiD specifications may be called for. Cengiz et al. [2019], de Chaisemartin and D'Haultfœuille [2020], Sun and Abraham [2021], and Borusyak, Jaravel, and Spiess [2024], for example, propose alternative estimators to correct for the potential bias of simple two-way FE estimators.

3.4.3. Spillovers in FE Models. Spillovers from treated to control firms can introduce bias that can be aggravated through FE (section 2.4.3). We recommend gauging potential spillover bias in FE models, especially if theory predicts that spillovers likely occur. Following Berg, Reisinger, and Streitz [2021], one way to do so is to examine whether a treatment effect estimate is sensitive to varying levels of FE. If a treatment effect changes systematically when using finer FE, the effect may be driven by positive spillovers to local control firms. Alternatively, one can try to explicitly control for spillovers to alleviate concerns about their confounding influence (e.g., Berg, Reisinger, and Streitz [2021], Huber [2023]).

---

[56] This diagnostic is in the spirit of Oster [2019], who suggests comparing the change in coefficient estimates relative to the corresponding change in explained variation (R^2) across specifications with and without (finer) FE. For an implementation of Oster's diagnostic test as a post-estimation program in STATA after using "reghdfe," see Arthur Morris's GitHub repository https://github.com/ArthurHowardMorris/psacalc_supports_reghdfe.

[57] Further discussion about heterogeneity in treatment responses and ways to explore and accommodate them can be found in Goodman-Bacon [2021], Heckman, Urzua, and Vytlacil [2006], Cornelissen et al. [2016], Brinch, Mogstad, and Wiswall [2017], Mogstad and Torgovitsky [2018], and Mogstad, Santos, and Torgovitsky [2018]. Although the issue of heterogeneous responses is often discussed in the context of DiD models, the resulting inferential problems are more general.

3.4.4. Dynamics in Panel Data with FE. Section 2.4.5 discusses two complications from dynamic effects in FE models, the first of which is that subjects' characteristics are unlikely to be completely fixed over time. When subjects' characteristics evolve over time, subject FE become less effective as panel periods lengthen. In such cases, finer FE (e.g., subject-decade FEs), first-difference estimators, or rolling estimators may be more effective (Millimet and Bellemare [2023]).

The second complication from dynamic effects is when the strict exogeneity assumption is violated, for example, when assignment of the treatment of interest depends on past or future outcomes. In such cases, Imai and Kim [2019] recommend that researchers try to measure and control for time-varying confounders rather than use subject FE to adjust for unobserved time-invariant confounders.[58] In a similar vein, Plümper and Troeger [2019] recommend that, in case of dynamics and feedback effects, researchers should try to explicitly model and account for the dynamic relationships (e.g., via dynamic panel estimators; Arellano and Bond [1991]) instead of rely on subject FE as a default solution. Accordingly, researchers typically do not want to use subject FE when dealing with dynamic models (e.g., autoregressive models with lagged values of Y among the regressors).

3.4.5. Outlier Treatment. FE can change which observations are outliers and potentially have the largest influence on $\hat{\beta}$ (section 2.1.4). The most influential observations in FE regressions are likely to be those with the most extreme within-FE variation, not those with the most extreme values in the raw distribution. Truncating or winsorizing raw variables will potentially fail to mitigate observations with high leverage in FE regressions, in which case, it is useful to truncate or winsorize the residualized variables instead. For example, identify outliers in the residualized data (e.g., the top/bottom 1%), and then truncate or winsorize those outliers before regressing Y on X and the FE (and controls).[59] One can also use traditional leverage diagnostics to identify influential observations after running a regression with FE, or can use alternative estimation techniques that are designed for outlier data (e.g. Leone, Minutti-Meza, and Wasley [2019], Gassen and Veenman [2021]).

---

[58] When dynamic effects are thought to exist and appropriate subject-level controls cannot be measured and included, a valid research design may simply not be available.

[59] For applications of this approach in the accounting literature, see, for example, Breuer [2021] and Breuer, Leuz, and Vanhaverbeke [2021]. This approach to outlier treatment is akin to the idea behind robust regressions. A notable choice after identifying outliers is whether to: (i) run the subsequent regression using the raw Y and X and including FEs (and controls); or (ii) run the subsequent regression using the residualized Y and X without FE. In the first approach, the FE coefficients are reestimated, so is particularly suitable if one expects that outliers may distort the FE coefficient estimates.

4. Conclusion

FE can be a powerful tool for improving identification, but their use should be carefully considered in light of each study's institutional setting and broader research design. This paper provides an introduction to FE models, tailored to the context of causal inference in accounting research. We hope that better understanding FE mechanics and drawbacks can help improve researchers' choices in designing FE regressions and interpreting the results thereof.

REFERENCES

ABADIE, A.; S. ATHEY; G. W. IMBENS; and J. M. WOOLDRIDGE. "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138 (2023): 1–35. https://doi.org/10.1093/qje/qjac038.

AMIR, E.; J. M. CARABIAS; J. JONA; and G. LIVNE. "Fixed-Effects in Empirical Accounting Research." 2016. Available at SSRN https://ssrn.com/abstract=2634089.

AMIRAM, D.; A. KALAY; A. KALAY; and N. B. OZEL. "Information Asymmetry and the Bond Coupon Choice." *The Accounting Review* 93 (2018): 37–59. ISSN 0001-4826. https://doi.org/10.2308/accr-51852.

ANGRIST, J. D., and J.-S. PISCHKE. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press, 2008.

ARELLANO, M., and S. BOND. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies* 58 (1991): 277–97. ISSN 00346527. https://doi.org/10.2307/2297968.

ARMSTRONG, C.; J. D. KEPLER; D. SAMUELS; and D. TAYLOR. "Causality Redux: The Evolution of Empirical Methods in Accounting Research and the Growth of Quasi-Experiments." *Journal of Accounting and Economics* 74 (2022): 101521. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2022.101521.

BAI, J. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77 (2009): 1229–79. ISSN 00129682. https://doi.org/10.3982/ECTA6135.

BAKER, A. C.; D. F. LARCKER; and C. C. WANG. "How Much Should We Trust Staggered Difference-in-Differences Estimates?" *Journal of Financial Economics* 144 (2022): 370–95. ISSN 0304-405X. https://doi.org/10.1016/j.jfineco.2022.01.004.

BALL, R., and V. V. NIKOLAEV. "On Earnings and Cash Flows as Predictors of Future Cash Flows." *Journal of Accounting and Economics* 73 (2022): 101430. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2021.101430.

BARRIOS, J. M. "Staggeringly Problematic: A Primer on Staggered DiD for Accounting Researchers." 2021. Available at SSRN https://ssrn.com/abstract=3794859.

BEAVER, W.; C. EGER; S. RYAN; and M. WOLFSON. "Financial Reporting, Supplemental Disclosures, and Bank Share Prices." *Journal of Accounting Research* 27 (1989): 157–78. ISSN 00218456. https://doi.org/10.2307/2491230.

BERG, T.; M. REISINGER; and D. STREITZ. "Spillover Effects in Empirical Corporate Finance." *Journal of Financial Economics* 142 (2021): 1109–27. ISSN 0304-405X. https://doi.org/10.1016/j.jfineco.2021.04.039.

BERTRAND, M., and A. SCHOAR. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics* 118 (2003): 1169–208. ISSN 00335533. https://doi.org/10.1162/003355303322552775.

BLANKESPOOR, B.; E. DEHAAN; J. WERTZ; and C. ZHU. "Why Do Individual Investors Disregard Accounting Information? The Roles of Information Awareness and Acquisition Costs." *Journal of Accounting Research* 57 (2019): 53–84. https://doi.org/10.1111/1475-679X.12248.

BLANKESPOOR, E.; E. DEHAAN; and I. MARINOVIC. "Disclosure Processing Costs, Investors' Information Choice, and Equity Market Outcomes: A Review." *Journal of Accounting and Economics* 70 (2020): 101344. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2020.101344.

BLOOMFIELD, R.; K. RENNEKAMP; and B. STEENHOVEN. "No System Is Perfect: Understanding How Registration-Based Editorial Processes Affect Reproducibility and Investment in Research Quality." *Journal of Accounting Research* 56 (2018): 313–62. https://doi.org/10.1111/1475-679X.12208.

BORUSYAK, K.; X. JARAVEL; and J. SPIESS. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* (2024): rdae007. https://doi.org/10.1093/restud/rdae007.

BREUER, M. "How Does Financial-Reporting Regulation Affect Industry-Wide Resource Allocation?" *Journal of Accounting Research* 59 (2021): 59–110. https://doi.org/10.1111/1475-679X.12345.

BREUER, M. "Bartik Instruments: An Applied Introduction." *Journal of Financial Reporting* 7 (2022): 49–67. ISSN 2380-2154. https://doi.org/10.2308/JFR-2021-003.

BREUER, M. "Another Way Forward: Comments on Ohlson's Critique of Empirical Accounting Research." *Accounting, Economics, and Law: A Convivium* (2023). https://doi.org/10.1515/ael-2022-0093.

BREUER, M.; K. HOMBACH; and M. A. MÜLLER. "When You Talk, I Remain Silent: Spillover Effects of Peers' Mandatory Disclosures on Firms' Voluntary Disclosures." *The Accounting Review* 97 (2022): 155–86. ISSN 0001-4826. https://doi.org/10.2308/TAR-2019-0433.

BREUER, M.; C. LEUZ; and S. VANHAVERBEKE. "Reporting Regulation and Corporate Innovation." Working Paper 26291, National Bureau of Economic Research, 2021. https://doi.org/10.3386/w26291.

BREUER, M., and H. SCHÜTT. "Accounting for Uncertainty: An Application of Bayesian Methods to Accruals Models." *Review of Accounting Studies* 28 (2023): 726–68. https://doi.org/10.1007/s11142-021-09654-0.

BRÜGGEMANN, U.; J.-M. HITZ; and T. SELLHORN. "Intended and Unintended Consequences of Mandatory IFRS Adoption: A Review of Extant Evidence and Suggestions for Future Research." *European Accounting Review* 22 (2013): 1–37. https://doi.org/10.1080/09638180.2012.718487.

BRINCH, C. N.; M. MOGSTAD; and M. WISWALL. "Beyond LATE with a Discrete Instrument." *Journal of Political Economy* 125 (2017): 985–1039. https://doi.org/10.1086/692712.

CABANILLAS, O. B.; J. D. MICHLER; A. MICHUDA; and E. TJERNSTRÖM. "Fitting and Interpreting Correlated Random-coefficient Models Using Stata." *The Stata Journal* 18 (2018): 159–73. https://doi.org/10.1177/1536867X1801800109.

CAMERAN, M.; D. CAMPA; and J. R. FRANCIS. "The Relative Importance of Auditor Characteristics Versus Client Factors in Explaining Audit Quality." *Journal of Accounting, Auditing & Finance* 37 (2022): 751–76. https://doi.org/10.1177/0148558X20953059.

CAMERON, A. C., and D. L. MILLER. "A Practitioner's Guide to Cluster-Robust Inference." 2015. Available at http://cameron.econ.ucdavis.edu/research/Cameron_Miller_JHR_2015_February.pdf.

CARD, D. "Design-Based Research in Empirical Microeconomics." *American Economic Review* 112 (2022): 1773–81. https://doi.org/10.1257/aer.112.6.1773.

CENGIZ, D.; A. DUBE; A. LINDNER; and B. ZIPPERER. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics* 134 (2019): 1405–54. ISSN 0033-5533. https://doi.org/10.1093/qje/qjz014.

CHAMBERLAIN, G. "Chapter 22 Panel Data." in *Handbook of Econometrics*, volume 2. Elsevier: 1984: 1247–318. ISSN 1573-4412. https://doi.org/10.1016/S1573-4412(84)02014-6.

CHEN, W.; P. HRIBAR; and S. MELISSA. "Incorrect Inferences When Using Residuals as Dependent Variables." *Journal of Accounting Research* 56 (2018): 751–96. https://doi.org/10.1111/1475-679X.12195.

CHRISTENSEN, H. B.; L. HAIL; and C. LEUZ. "Capital-Market Effects of Securities Regulation: Prior Conditions, Implementation, and Enforcement." *The Review of Financial Studies* 29 (2016): 2885–924. ISSN 0893-9454. https://doi.org/10.1093/rfs/hhw055.

Cochrane, J. "A Brief Parable of Over-Differencing." 2012. Research Note. Available at https://static1.squarespace.com/static/5e6033a4ea02d801f37e15bb/t/5eda716edf351879908c4cad/1591374191294/overdifferencing.pdf.

Conley, T.; S. Goncalves; and C. Hansen. "Inference with Dependent Data in Accounting and Finance Applications." *Journal of Accounting Research* 56 (2018): 1139–203. https://doi.org/10.1111/1475-679X.12219.

Cornelissen, T.; C. Dustmann; A. Raute; and U. Schönberg. "From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions." *Labour Economics* 41 (2016): 47–60. ISSN 0927-5371. https://doi.org/10.1016/j.labeco.2016.06.004. SOLE/EALE conference issue 2015.

Correia, S. "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix." 2015. Research Note. Available at http://scorreia.com/research/singletons.pdf.

Correia, S. "A Feasible Estimator for Linear Models with Multi-Way Fixed Effects." Working Paper, 2016. Available at http://scorreia.com/research/hdfe.pdf.

Correia, S.; P. Guimarães; and T. Zylkin. "Fast Poisson Estimation with High-Dimensional Fixed Effects." *The Stata Journal* 20 (2020): 95–115. https://doi.org/10.1177/1536867X20909691.

Daske, H.; L. Hail; C. Leuz; and R. Verdi. "Mandatory IFRS Reporting Around the World: Early Evidence on the Economic Consequences." *Journal of Accounting Research* 46 (2008): 1085–142. https://doi.org/10.1111/j.1475-679X.2008.00306.x.

Daske, H.; L. Hail; C. Leuz; and R. Verdi. "Adopting a Label: Heterogeneity in the Economic Consequences Around IAS/IFRS Adoptions." *Journal of Accounting Research* 51 (2013): 495–547. https://doi.org/10.1111/1475-679X.12005.

de Chaisemartin, C., and X. D'Haultfœuille. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (2020): 2964–96. https://doi.org/10.1257/aer.20181169.

de Chaisemartin, C., and Z. Lei. "More Robust Estimators for Panel Bartik Designs, with an Application to the Effect of Chinese Imports on US Employment." 2023. Available at SSRN: https://ssrn.com/abstract=3802200.

De George, E. T.; X. Li; and L. Shivakumar. "A Review of the IFRS Adoption Literature." *Review of Accounting Studies* 21 (2016): 898–1004. https://doi.org/10.1007/s11142-016-9363-1.

DeFond, M. L., and J. Jiambalvo. "Debt Covenant Violation and Manipulation of Accruals." *Journal of Accounting and Economics* 17 (1994): 145–76. ISSN 0165-4101. https://doi.org/10.1016/0165-4101(94)90008-6.

deHaan, E. "Using and Interpreting Fixed Effects Models." 2021. Available at SSRN https://ssrn.com/abstract=3699777.

deHaan, E.; N. Li; and F. S. Zhou. "Financial Reporting and Employee Job Search." *Journal of Accounting Research* 61 (2023a): 571–617. https://doi.org/10.1111/1475-679X.12469.

deHaan, Ed.; J. R. Moon, J. E. Shipman; Q. T. Swanquist; and R. L. Whited. "Control Variables in Interactive Models." *Journal of Financial Reporting* 8 (2023b): 77–85. https://doi.org/10.2308/JFR-2021-023.

deHaan, E.; Y. Song; C. Xie; and C. Zhu. "Obfuscation in Mutual Funds." *Journal of Accounting and Economics* 72 (2021): 101429. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2021.101429.

Dekeyser, S.; A. Gaeremynck; W. R. Knechel; and M. Willekens. "Multimarket Contact and Mutual Forbearance in Audit Markets." *Journal of Accounting Research* 59 (2021): 1651–88. https://doi.org/10.1111/1475-679X.12406.

Engel, E.; R. M. Hayes; and X. Wang. "The Sarbanes-Oxley Act and Firms' Going-Private Decisions." *Journal of Accounting and Economics* 44 (2007): 116–45. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2006.07.002. Conference Issue on Corporate Governance: Financial Reporting, Internal Control, and Auditing.

Fama, E. F., and K. R. French. "New Lists: Fundamentals and Survival Rates." *Journal of Financial Economics* 73 (2004): 229–69. ISSN 0304-405X. https://doi.org/10.1016/j.jfineco.2003.04.001.

FAMA, E. F., and J. D. MACBETH. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (1973): 607–36. ISSN 00223808. https://doi.org/10.1086/260061.

FOSTER, G. "Intra-Industry Information Transfers Associated with Earnings Releases." *Journal of Accounting and Economics* 3 (1981): 201–32. ISSN 0165-4101. https://doi.org/10.1016/0165-4101(81)90003-3.

FRISCH, R., and F. V. WAUGH. "Partial Time Regressions as Compared with Individual Trends." *Econometrica* 1 (1933): 387–401. ISSN 00129682. https://doi.org/10.2307/1907330.

GARBER, S., and S. KLEPPER. "Extending the Classical Normal Errors-in-Variables Model." *Econometrica* 48 (1980): 1541–46. ISSN 00129682. https://doi.org/10.2307/1912823.

GASSEN, J., and D. VEENMAN. "Outliers and Robust Inference in Archival Accounting Research." 2021. Available at SSRN https://ssrn.com/abstract=3880942.

GIPPER, B.; C. LEUZ; and M. MAFFETT. "Public Oversight and Reporting Credibility: Evidence from the PCAOB Audit Inspection Regime." *The Review of Financial Studies* 33 (2019): 4532–79. ISSN 0893-9454. https://doi.org/10.1093/rfs/hhz149.

GLAESER, S., and W. R. GUAY. "Identification and Generalizability in Accounting Research: A Discussion of Christensen, Floyd, Liu, and Maffett (2017)." *Journal of Accounting and Economics* 64 (2017): 305–12. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2017.08.003.

GLAESER, S., and J. D. OMARTIAN. "Public Firm Presence, Financial Reporting, and the Decline of U.S. Manufacturing." *Journal of Accounting Research* 60 (2022): 1085–130. https://doi.org/10.1111/1475-679X.12411.

Goldsmith-Pinkham, P.; I. SORKIN; and H. SWIFT. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110 (2020): 2586–624. https://doi.org/10.1257/aer.20181047.

GOODMAN-BACON, A. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2021): 254–77. ISSN 0304-4076. https://doi.org/10.1016/j.jeconom.2021.03.014.

GORMLEY, T. A., and D. A. MATSA. "Common Errors: How to (and Not to) Control for Unobserved Heterogeneity." *The Review of Financial Studies* 27 (2013): 617–61. ISSN 0893-9454. https://doi.org/10.1093/rfs/hht047.

GRANJA, J. "Disclosure Regulation in the Commercial Banking Industry: Lessons from the National Banking Era." *Journal of Accounting Research* 56 (2018): 173–216. https://doi.org/10.1111/1475-679X.12193.

GREENE, W. H. *Econometric Analysis.* Upper Saddle River, New Jersey 07458: Prentice Hall, 2002.

GUAY, W.; D. SAMUELS; and D. TAYLOR. "Guiding through the Fog: Financial Statement Complexity and Voluntary Disclosure." *Journal of Accounting and Economics* 62 (2016): 234–69. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2016.09.001. Conference papers 2015.

HECKMAN, J. J.; S. URZUA; and E. VYTLACIL. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *The Review of Economics and Statistics* 88 (2006): 389–432. ISSN 00346535. http://www.jstor.org/stable/40043006.

HUBER, K. "Estimating General Equilibrium Spillovers of Large-Scale Shocks." *The Review of Financial Studies* 36 (2023): 1548–84. ISSN 0893-9454. https://doi.org/10.1093/rfs/hhac057.

IACUS, S. M.; G. KING; and G. PORRO. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20 (2012): 1–24. https://doi.org/10.1093/pan/mpr013.

IMAI, K., and I. S. KIM. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2019): 467–90. https://doi.org/10.1111/ajps.12417.

IMBENS, G. W. "Statistical Significance, p-Values, and the Reporting of Uncertainty." *Journal of Economic Perspectives* 35 (2021): 157–74. https://doi.org/10.1257/jep.35.3.157.

IMBENS, G. W., and D. B. RUBIN. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge: Cambridge University Press, 2015. https://doi.org/10.1017/CBO9781139025751.

JAYARAMAN, S., and S. P. KOTHARI. "Cross-Border Financing by the Industrial Sector Increases Competition in the Domestic Banking Sector." *The Accounting Review* 91 (2016): 535–58. ISSN 00014826. https://doi.org/10.2308/accr-51199.

Jennings, J.; J. M. Kim; J. Lee; and D. Taylor. "Measurement Error, Fixed Effects, and False Positives in Accounting Research." *Review of Accounting Studies* 29 (2024): 959–95. https://doi.org/10.1007/s11142-023-09754-z.

Jollineau, S. J., and R. M. Bowen. "A Practical Guide to Using Path Analysis: Mediation and Moderation in Accounting Research." *Journal of Financial Reporting* 8 (2023): 11–40. ISSN 2380-2154. https://doi.org/10.2308/JFR-2021-004.

Jones, J. J. "Earnings Management During Import Relief Investigations." *Journal of Accounting Research* 29 (1991): 193–228. ISSN 00218456. https://doi.org/10.2307/2491047.

Kallapur, S. "Beyond P<0.05: Scientific Inference in Accounting Research." *Studies in Accounting Research, American Accounting Association* 34 (2022). https://aaahqbookstore.org/catalog/book/beyond-p.

Kim, S. "Delays in Banks' Loan Loss Provisioning and Economic Downturns: Evidence from the U.S. Housing Market." *Journal of Accounting Research* 60 (2022): 711–54. https://doi.org/10.1111/1475-679X.12415.

Lancaster, T. "The Incidental Parameter Problem since 1948." *Journal of Econometrics* 95 (2000): 391–413. ISSN 0304-4076. https://doi.org/10.1016/S0304-4076(99)00044-5.

Lavine, M. "Frequentist, Bayes, or Other?" *The American Statistician* 73 (2019): 312–18. https://doi.org/10.1080/00031305.2018.1459317.

Leone, A. J.; M. Minutti-Meza; and C. E. Wasley. "Influential Observations and Inference in Accounting Research." *The Accounting Review* 94 (2019): 337–64. ISSN 0001-4826. https://doi.org/10.2308/accr-52396.

Leuz, C. "Was the Sarbanes-Oxley Act of 2002 Really this Costly? A Discussion of Evidence from Event Returns and Going-Private Decisions." *Journal of Accounting and Economics* 44 (2007): 146–65. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2007.06.001. Conference Issue on Corporate Governance: Financial Reporting, Internal Control, and Auditing.

Leuz, C. "Evidence-based Policymaking: Promise, Challenges and Opportunities for Accounting and Financial Markets Research." *Accounting and Business Research* 48 (2018): 582–608. https://doi.org/10.1080/00014788.2018.1470151.

Leuz, C. "Towards a Design-Based Approach to Accounting Research." *Journal of Accounting and Economics* 74 (2022): 101550. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2022.101550.

Leuz, C., and P. D. Wysocki. "The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research." *Journal of Accounting Research* 54 (2016): 525–622. https://doi.org/10.1111/1475-679X.12115.

Lev, B. "On the Usefulness of Earnings and Earnings Research: Lessons and Directions from Two Decades of Empirical Research." *Journal of Accounting Research* 27 (1989): 153–92. ISSN 00218456. https://doi.org/10.2307/2491070.

Liu, C.-C., and S. G. Ryan. "The Effect of Bank Loan Portfolio Composition on the Market Reaction to and Anticipation of Loan Loss Provisions." *Journal of Accounting Research* 33 (1995): 77–94. ISSN 00218456. https://doi.org/10.2307/2491293.

Lovell, M. C. "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis." *Journal of the American Statistical Association* 58 (1963): 993–1010. https://doi.org/10.1080/01621459.1963.10480682.

McMullin, J. L., and B. Schonberger. "Entropy-Balanced Accruals." *Review of Accounting Studies* 25 (2020): 84–119. https://doi.org/10.1007/s11142-019-09525-9.

McShane, B. B.; D. Gal; A. Gelman; C. Robert; and J. L. Tackett. "Abandon Statistical Significance." *The American Statistician* 73 (2019a): 235–45. https://doi.org/10.1080/00031305.2018.1527253.

McShane, B. B., and A. Gelman. "Selecting on Statistical Significance and Practical Importance Is Wrong." *Journal of Information Technology* 37 (2022): 312–15. https://doi.org/10.1177/02683962221086297.

McShane, B. B.; J. L. Tackett; U. Böckenholt; and A. Gelman. "Large-Scale Replication Projects in Contemporary Psychological Research." *The American Statistician* 73 (2019b): 99–105. https://doi.org/10.1080/00031305.2018.1505655.

MILLIMET, D. L., and M. F. BELLEMARE. "Fixed Effects and Causal Inference." IZA Discussion Paper Series No. 16202. 2013. Available at https://docs.iza.org/dp16202.pdf.

MOGSTAD, M.; A. SANTOS; and A. TORGOVITSKY. "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters." *Econometrica* 86 (2018): 1589–619. https://doi.org/10.3982/ECTA15463.

MOGSTAD, M., and A. TORGOVITSKY. "Identification and Extrapolation of Causal Effects with Instrumental Variables." *Annual Review of Economics* 10 (2018): 577–613. https://doi.org/10.1146/annurev-economics-101617-041813.

MUMMOLO, J., and E. PETERSON. "Improving the Interpretation of Fixed Effects Regression Results." *Political Science Research and Methods* 6 (2018): 829–35. https://doi.org/10.1017/psrm.2017.44.

MUNDLAK, Y. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46 (1978): 69–85. ISSN 00129682. https://doi.org/10.2307/1913646.

NEYMAN, J., and E. L. SCOTT. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16 (1948): 1–32. ISSN 00129682. https://doi.org/10.2307/1914288.

NICKELL, S. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (1981): 1417–26. ISSN 00129682. https://doi.org/10.2307/1911408.

NIKOLAEV, V., and L. VAN LENT. "The Endogeneity Bias in the Relation Between Cost-of-Debt Capital and Corporate Disclosure Policy." *European Accounting Review* 14 (2005): 677–724. https://doi.org/10.1080/09638180500204624.

OSTER, E. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2019): 187–204. https://doi.org/10.1080/07350015.2016.1227711.

PETERSEN, M. A. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *The Review of Financial Studies* 22 (2008): 435–80. ISSN 0893-9454. https://doi.org/10.1093/rfs/hhn053.

PLÜMPER, T., and V. E. TROEGER. "Not so Harmless After All: The Fixed-Effects Model." *Political Analysis* 27 (2019): 21–45. https://doi.org/10.1017/pan.2018.17.

ROBERTS, M. R., and T. M. WHITED. "Endogeneity in Empirical Corporate Finance." in *Handbook of the Economics of Finance*, volume 2, edited by G. M. Constantinides, M. Harris, and R. M. Stulz. Amsterdam, The Netherlands: Elsevier, 2013: 493–572. https://doi.org/10.1016/B978-0-44-453594-8.00007-0.

SANI, J.; N. SHROFF; and H. WHITE. "Spillover Effects of Mandatory Portfolio Disclosures on Corporate Investment." *Journal of Accounting and Economics* 76 (2023): 101641. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2023.101641.

SHROFF, N. "Real Effects of PCAOB International Inspections." *The Accounting Review* 95 (2020): 399–433. ISSN 0001-4826. https://doi.org/10.2308/accr-52635.

SRIVASTAVA, A. "Why Have Measures of Earnings Quality Changed over Time?" *Journal of Accounting and Economics* 57 (2014): 196–217. ISSN 0165-4101. https://doi.org/10.1016/j.jacceco.2014.04.001.

SUN, L., and S. ABRAHAM. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2021): 175–99. ISSN 0304-4076. https://doi.org/10.1016/j.jeconom.2020.09.006.

SURI, T. "Selection and Comparative Advantage in Technology Adoption." *Econometrica* 79 (2011): 159–209. https://doi.org/10.3982/ECTA7749.

WASSERSTEIN, R. L.; A. L. SCHIRM; and N. A. LAZAR. "Moving to a World Beyond "p<0.05"." *The American Statistician* 73 (2019): 1–19. https://doi.org/10.1080/00031305.2019.1583913.

WOOLDRIDGE, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2010.

ZHOU, X. "Understanding the Determinants of Managerial Ownership and the Link between Ownership and Performance: Comment." *Journal of Financial Economics* 62 (2001): 559–71. ISSN 0304-405X. https://doi.org/10.1016/S0304-405X(01)00085-X.