JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 54, No. 1, Feb. 2019, pp. 1–29
COPYRIGHT 2018, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109018000996

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

Panel-Data Estimation in Finance: Testable Assumptions and Parameter (In)Consistency

William D. Grieser and Charles J. Hadlock*

Abstract

We investigate the strict-exogeneity assumption, a necessary condition for estimator consistency in many finance panel-data applications. We outline tests for strict exogeneity in both traditional (non–instrumental variable (IV)) and IV settings. When we apply these tests in common traditional finance panel regressions, we find that the strict-exogeneity assumption is often strongly rejected, suggesting large inference errors. We test for strict exogeneity in specific finance panel-data IV settings and illustrate the potential for these tests to help confirm, or rule out, the validity of common panel-data IV estimators. We offer recommendations to address the strict-exogeneity issue in finance research.

I. Introduction

The use of panel data is common in finance research. An important benefit of this data structure is that it allows researchers to control for unobserved unit-level factors that may be correlated with the explanatory variables of interest. In many finance applications, the firm-year is the unit of observation, and the researcher selects an estimator that accounts for time-invariant firm heterogeneity. The most common estimator choice in this context is the fixed-effects estimator, followed by the first-difference estimator and then some other alternatives.

An important stream of recent research highlights errors that frequently appear when common estimators are applied in finance panel-data contexts. First, as Gormley and Matsa (2014) illustrate, researchers often implement the fixed-effects estimation procedure incorrectly. Second, as highlighted by Petersen (2009) and Thompson (2011), researchers frequently rely on standard errors that

*Grieser, w.grieser@tcu.edu, Texas Christian University Neeley School of Business; Hadlock (corresponding author), hadlock@msu.edu, Michigan State University Broad College of Business. We thank an anonymous referee, Bernard Black, Jennifer Conrad (the editor), Todd Gormley, Jeff Wooldridge, and seminar participants at the University of Iowa, the University of Nebraska, the 2015 London Business School (LBS) Symposium on Causal Inference, and the 2016 Financial Research Association Meetings for helpful comments and suggestions. All errors remain our own.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

do not adequately adjust for the underlying error covariance structure. These authors offer guidance on how to correctly estimate parameters of interest and the associated standard errors.

Although this recent literature makes many useful and important points, it does not address the preliminary issue of whether certain estimators are appropriate in a given setting in the sense that the selected estimator is expected to yield consistent estimates of the underlying parameters of interest. Most authors recognize that contemporaneous exogeneity of the explanatory variables is necessary for the fixed-effects (FE) estimator, or its cousin, the first-difference (FD) estimator, to consistently estimate the parameters of interest. As Roberts and Whited (2012) and others have highlighted, this contemporaneous-exogeneity assumption is untestable, at least when convincing instruments are not available. Largely unrecognized in the finance literature is the fact that the consistency of these common estimators requires reliance on a *much stronger* notion of the strict exogeneity of the explanatory variables. Importantly, a key element of this assumption *is testable*. Similarly, when instruments are available, many panel-data instrumental variable (IV) estimators rely on the assumption of the strict exogeneity of the instruments, a *testable* assumption that is also ubiquitously overlooked.

In this article, we argue that the strict-exogeneity assumption should be considered in all finance panel-data applications and offer guidance on how to do so. Our initial focus, and much of our attention, is for the case when a researcher is contemplating using traditional FE or FD estimators (referred to hereafter as the traditional case/setting) because these estimators are by far the most frequently used in finance panel-data contexts. However, given the increasing popularity of IV estimation, we extend some of our analysis and discussion of the strict-exogeneity issue to finance panel-data IV settings (referred to as the IV case/setting).

In the traditional case, the relevant strict-exogeneity assumption needed for consistent estimation requires there to be no correlation between the model error term and the model covariates at all leads and lags, conditional on the unit-level unobserved effect (see Wooldridge (2010)). This assumption will be violated in any setting where there is feedback from a dependent variable to future values of the independent variable, a prominent feature in many dynamic theoretical models.[1] It is also violated if the explanatory variables and the dependent variable are partially governed by common shocks but with different leads or lags in their responses to these shocks. Certainly, even a cursory look at the variables studied in the related literature suggests that this assumption may frequently be violated.

To investigate the traditional case, we select a set of canonical panel-data regression models from the empirical finance literature. We then consider the simple tests for strict exogeneity outlined by Wooldridge (2010). Because the properties of these tests have not been widely explored, we first present evidence using simulated data in which the data-generating process is known. These simulation results indicate that the Wooldridge (2010) tests have substantial power. When we then turn to actual data, we find that these tests reject the assumption of strict

[1]See, for example, Fischer, Heinkel, and Zechner (1989), Hermalin and Weisbach (1998), Hennessy and Whited (2005), and Strebulaev and Whited (2012).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

exogeneity in virtually all of the large-sample models we consider. This evidence implies that many of the common FE and FD estimates that appear in the related literature are unlikely to be consistent estimates of the parameters of interest, independent of whether the more frequently recognized assumption of contemporaneous exogeneity holds.

If FE and FD estimates do not converge to their true values in large samples, reported estimates may have little economic content. To consider the possible economic magnitude of inference errors, we explore differences between FE and FD estimators in our selected settings because these estimators should have identical probability limits, the true parameter of interest, when the strict-exogeneity assumption holds. In this analysis, we find that the differences between FE and FD estimators are often quite large, frequently differing by a factor of more than 50%, even after conditioning on the estimates being of the same sign. Moreover, there are cases in which these estimators are significant and of opposite sign. These pathological cases are on the order of 18 times more prevalent than would be expected by chance under conservative assumptions.

This evidence from comparing estimates suggests that the magnitude of the potential inconsistencies in FE and FD estimators when strict exogeneity is violated will often be quite substantial. Given our evidence on both the scope and the size of the problem, it would appear essential for finance researchers to consider the strict-exogeneity issue whenever they are considering using traditional FE or FD estimates. As we show, this is not current practice. A quick way to detect whether this may be an issue is to conduct a simple comparison of FE and FD estimates for the same model. If a violation is suspected, based on either this comparison or a priori reasoning, a formal test should be conducted. Unfortunately, as we show, if there is a violation, the performance of the FE and FD estimators cannot be ordinally ranked in many finance applications, even when the time dimension of the sample is long.

We emphasize that in traditional settings, tests for strict exogeneity do not eliminate the need to consider the issue of contemporaneous exogeneity. Thus, nonrejection of the null that strict exogeneity holds only suggests that the researcher does not have to account for this additional empirical concern. When the strict-exogeneity tests do reject, as appears to be common, the researcher must directly confront the issue of parameter inconsistency regardless of whether the untestable contemporaneous-exogeneity assumption holds.

If tests for strict exogeneity in the traditional case reject, a researcher in search of consistent estimates must turn to alternative estimators. One possibility is to use dynamic panel-data generalized method of moments (GMM) estimators that rely on lagged values as instruments (e.g., Arellano and Bond (1991), Blundell and Bond (1998)). We note that many of these estimators also rely on testable assumptions related to their suitability, notably, with no serial correlation in the error term. We briefly investigate and find that these tests often reject the underlying assumptions in the finance settings we identify, suggesting severe limitations to this approach in many contexts.[2]

[2]Consistent with our evidence, Dang, Kim, and Shin (2015) illustrate that these tests often lead to rejection of the underlying assumptions in the context of dynamic capital structure models. See also

In recent years, there has been a trend in finance toward estimating panel-data models using external instruments or related approaches that exploit external exogenous variation (e.g., natural experiments).[3] However, in IV settings, the strict-exogeneity issue does not disappear. The most common panel-data IV estimators (e.g., fixed-effects 2-stage least squares (FE-2SLS)) rely on the assumption of the strict exogeneity of the instrument. This assumption is stronger than the notion of contemporaneous exogeneity, which is the (generally) untestable assumption that finance researchers appear to have in mind in panel-data contexts when they justify the exclusion restriction. Loosely speaking, the strict exogeneity of the instrument requires that there be no partial correlation between the dependent variable and *future* values of the instrument, conditional on the unit-level unobserved effects and other covariates. To the best of our knowledge, a direct test for the strict exogeneity of an instrument has never been reported in any published empirical finance article, even though simple tests are available (e.g., Wooldridge (2016)).

Because there is no straightforward way to examine a large set of canonical panel finance models in IV settings given the unique data that are often used in these studies, we instead examine two examples motivated by recent research. When we test for strict exogeneity in these selected examples, in one case we find that the proposed instrument passes the strict-exogeneity test, whereas in the other case, strict exogeneity is rejected. We use these examples both to demonstrate the mechanics of how to conduct these tests and also to discuss the types of mechanisms through which apparently exogenous instruments (in the contemporaneous sense) may fail to be strictly exogenous. We advocate for the use of these tests in IV settings to either uncover a problematic instrument or, in the more optimistic scenario, to increase confidence in the validity of a panel-data IV analysis.

We hope that the tests we outline and the guidance we provide, both in the traditional case and the IV case, become standard practice as the field moves toward more convincing empirical modeling. Collectively, our evidence suggests that these tests will, in many cases, have substantial power to detect substantive inconsistencies in common estimators. Our evidence also suggests that inconsistent estimates arising from violations of strict exogeneity permeate the prior, but hopefully not the future, published finance literature. Although this may not be surprising to skeptics of the existing empirical literature, we establish this point independent of the usual philosophical debate regarding contemporaneous endogeneity.

The rest of the article is organized as follows: In Section II, we outline the basic econometric issues, survey the existing literature, and outline the tests for strict exogeneity, in both the traditional and IV settings. In Section III, we apply these tests to actual data in a large set of traditional-setting specifications selected from the recent finance literature. In Section IV, we consider strict-exogeneity tests in the IV setting. Section V concludes.

the comments by Roberts and Whited (2012) on the limitations of IV techniques that rely on lagged instruments in empirical finance applications.

[3]For some interesting recent successes in applying this strategy, see Matsa (2010) and Agrawal and Matsa (2013).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

II. Related Literature and Empirical Strategy

A. Outlining the Issue in the Traditional Case

Although the strict-exogeneity assumption is discussed in textbook treatments of panel data, this assumption is almost never explicitly acknowledged in finance panel-data applications. Given this lack of familiarity to finance audiences, we briefly outline the issue here, with a specific eye toward finance applications. Textbook treatments provide more of the technical details (e.g., Cameron and Trivedi (2005), Wooldridge (2010)). Our discussion here is focused on the traditional setting in which the researcher is considering FE or FD estimators. In a later subsection, we explicitly consider the IV setting.

We consider a simple regression model with a dependent variable y, a single independent variable of interest x, and an assumed model in which

$$y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it}, \tag{1}$$

where i denotes an arbitrary cross-sectional unit (from 1 to N), and t denotes an arbitrary time period (from 1 to T).[4] Because N is usually much larger than T, all asymptotics are for N approaching infinity. Following Wooldridge (2010), we will refer to the assumption $\mathrm{E}(\epsilon_{it} \mid x_{it}, \alpha_i) = 0$ as the contemporaneous-exogeneity assumption and $\mathrm{E}(\epsilon_{it} \mid x_{is}, \alpha_i) = 0$ for all t and s as the strict-exogeneity assumption. Assuming contemporaneous exogeneity holds and recognizing that lagged values of the explanatory variables can always be added to the model, we are concerned primarily with violations of strict exogeneity in which $\mathrm{E}(\epsilon_{it} | x_{is}, \alpha_i) \neq 0$ for $s > t$.

To see how this assumption may be violated, consider the case in which high realizations of the dependent variable at time t (e.g., firm performance) have a positive effect on *future values* of the explanatory variable (e.g., managerial ownership from exercising stock options after they vest). In this case, strict exogeneity would be violated because $\mathrm{E}(\epsilon_{it} x_{i(t+1)} \mid \alpha_i) > 0$; higher values of this period's performance are associated with higher levels of future ownership because of a dynamic feedback effect. An alternative mechanism that will also lead to a violation of the strict-exogeneity assumption is one in which the dependent variable at time t (e.g., a firm's stock return volatility) and the independent variable in a future period $t+n$ (e.g., managerial ownership) both respond to the same economic shock. An example of this type could be an industry shock that increases risk at time t but where the two variables respond with different lags (e.g., returns respond immediately because they are forward-looking, whereas ownership may respond with a lag because of adjustment costs). In this case, strict exogeneity would be violated because $\mathrm{E}(\epsilon_{it} x_{i(t+n)} \mid \alpha_i) \neq 0$.

To understand the resulting bias in the FE and FD estimators, consider the simple case in which we have 2 time periods (call them 1 and 2) so that the FE and FD estimators are numerically identical and the resulting parameter estimate for β is derived from a simple linear regression of changes in y (e.g., firm performance) on changes in x (e.g., ownership). Suppose also that the true β is 0, in which case there is no causal effect of x on y, but also assume that y has

[4] For ease of exposition, we omit additional control variables in this discussion. These can be included in the model without loss of generality.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

a positive feedback effect on future values of x so that $\mathrm{E}(\epsilon_{i1}x_{i2}\,|\,\alpha_i) > 0$. If we regress $(\Delta y = y_2 - y_1)$ on $(\Delta x = x_2 - x_1)$, the only systematic variation in the data will arise from high (low) y_1 values tending to feed back to high (low) x_2 values. This will result in an apparent negative correlation between Δy and Δx and will yield a spurious negative estimated β coefficient. Extensions of this argument apply to upward and downward inconsistencies in parameter estimates that depend on the sign of the actual coefficient (when $\beta \neq 0$) and the sign of the correlation between the model error term and future values of the explanatory variables (i.e., the sign on $\mathrm{E}(\epsilon_{i1}x_{i2}\,|\,\alpha_i)$). Although the direction of bias can be easy to sign in fairly simple settings, multivariate relationships and additional time periods make it difficult to sign and understand the resulting inconsistencies.

Given the potentially serious consequences of this issue for estimation, it is not surprising that Wooldridge (2010) and others emphasize the importance of testing for strict exogeneity before relying on FE or FD estimation procedures. However, as we discuss later, researchers in finance almost never test for strict exogeneity, raising the possibility of the widespread reporting of inconsistent estimates that capture something different from what the researcher intends.

B. Recent/Current Practice

To characterize current practice, we search through every issue of the *Journal of Finance*, the *Journal of Financial Economics*, and the *Review of Financial Studies* from 2006 to 2013 for the mention of the word combination "fixed effect" (or a synonym).[5] We quickly scan each flagged article to determine whether the paper features a panel-data empirical model that accounts for unit-level (e.g., firm, bank, person) unobserved effects, rather than solely time (e.g., year, quarter, etc.) effects. We place each paper into non–mutually exclusive categories based on whether the author(s) report i) traditional FE estimates, ii) traditional FD estimates, and/or iii) some version of a dynamic panel-data GMM estimate. We do not categorize models that rely on panel-data IV estimations with external instruments because our initial focus is on evaluating models in which external information is not exploited. There are far fewer published studies of this external type. Although they are increasingly common, they still appear to be in the distinct minority, as shown in the following discussion.

Our procedure flags 251 articles that report FE or FD estimates (222 and 47, respectively) and 27 that report dynamic panel-data GMM estimates. If we exclude models that include lagged dependent variables, the corresponding numbers are 240, 216, 44, and 11, respectively. These figures illustrate that FE is the most popular panel-data estimation procedure. In all of the articles that report solely FE estimates, only three mention the words "strict exogeneity," and of these, only a single article tests the strict-exogeneity assumption. Clearly, the field has not widely recognized this issue.

To update this evidence, in late 2017 we searched through the *Journal of Finance* for all articles that use the phrase "firm fixed effect," and we sorted the

[5]These journals are the three pure finance journals with the highest impact factors. We use some articles for early 2014, with the cutoff for each journal depending on our online library access at the time of the data collection.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

resulting articles in reverse chronological order. We read the most recent 15 articles and identified all tables/models that include the use of these effects. Within this set of articles, 12 rely on traditional FE/FD estimates, three report results from panel-data IV estimations, and none mentions the issue of strict exogeneity.[6]

C. Prior Work in Finance That Accounts for Violations of Strict Exogeneity

As is well known, any panel-data model with unit-level unobserved effects and a lagged dependent variable as an explanatory variable must violate the strict-exogeneity assumption (see Angrist and Pischke (2009)). A prominent empirical finance literature with this feature is the dynamic capital structure literature, wherein current leverage is often assumed to be partially governed by past leverage. Because conventional FE and FD estimators are inconsistent in this context, several recent articles in this area have exploited variants of the dynamic panel-data GMM approach.[7]

As we discussed earlier, the strict-exogeneity issue is almost entirely unacknowledged in panel-data finance models that do not include a lagged dependent variable. The one notable exception is the important work of Wintoki, Linck, and Netter (2012). These authors highlight the importance of the strict-exogeneity issue in a specific setting, namely, the effect of board structure on firm performance. They test for strict exogeneity in this context and reject the validity of this assumption, leading them to question prior work.

Our article is closely related to that of Wintoki et al. (2012). The distinguishing feature of our study is that we highlight that this issue applies to a large set of empirical models in finance and demonstrate that strict exogeneity is routinely rejected in these models. Thus, the concerns raised by Wintoki et al. (2012) in a specific context turn out to be only the tip of the iceberg. In addition, we i) explore test power for the first time, ii) demonstrate the possible magnitude of the resulting inconsistencies, iii) recommend a comparison of FE and FD estimates as a simple diagnostic, iv) illustrate that long panels often will not ameliorate the estimation challenges, and v) highlight that testing for the strict exogeneity of an external instrument is both feasible and advisable before using common panel-data IV estimators.

D. Testing for Strict Exogeneity in Traditional Settings

Wooldridge (2010) outlines two simple regression-based tests for the null hypothesis that the strict-exogeneity assumption is valid. Both tests rely on considering estimated coefficients on leading values of the explanatory variable because these should theoretically be 0 under the null. For a model with a single

[6] In articles that report traditional FE/FD estimates, there has been a secular trend toward selecting explanatory variables that are more plausibly exogenous, say, because the variation is driven by a natural experiment. The key is to select variables that are strictly, rather than just contemporaneously, exogenous. The tests we propose in the traditional setting (IV setting) allow one to test this assumption when the treatment variable enters the panel regressions directly (as an instrument).

[7] For an analysis of estimators in finance models with lagged dependent variables, see Flannery and Hankins (2013). For a skeptical view of these estimators in a dynamic capital structure context, see Iliev and Welch (2011).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

explanatory variable, the FE version of the test estimates the regression equation

$$y_{it} = \alpha_i + \beta x_{it} + \gamma x_{i(t+1)} + \epsilon_{it} \quad (2)$$

using the fixed-effects transformation, whereas the FD version of the test estimates the equation

$$\Delta y_{it} = \beta \Delta x_{it} + \rho x_{it} + \Delta \epsilon_{it}. \quad (3)$$

[8]

Strict exogeneity is rejected if the coefficients on γ and/or ρ differ significantly from 0. In the case of multiple explanatory variables, γ or ρ coefficients are estimated for each variable, and the tests are for whether the set of γ or ρ coefficients is jointly significant. Test statistics are calculated with standard errors clustered at the firm level to allow for arbitrary serial correlation and heteroscedasticity.

Before using these tests on actual data, we explore the properties of these tests using simulated data. In these simulations, the underlying equation of interest is

$$y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it}, \quad (4)$$

where the ϵ_{it} values are independent and identically distributed (IID) N(0,1). We allow for possible correlation of the explanatory variable in this equation with i) the unobserved firm intercept, ii) lagged values of the error term, and iii) the contemporaneous error term. We create these correlations by modeling

$$x_{it} = 0.40\alpha_i + \theta y_{i(t-1)} + \lambda \epsilon_{it} + u_{it}, \quad (5)$$

where the u_{it} values are IID N(0,1).[9] If θ and λ are both equal to 0, there is no endogeneity issue after accounting for the unobserved effect, and the FE and FD estimators should consistently estimate the underlying parameter β. Pooled ordinary least squares (OLS) will mistakenly confuse the 0.40 correlation between the unobserved effect α_i and the explanatory variable x_{it} as a reflection of a causal relation between y_{it} and x_{it}, which is, of course, the usual motivation for using traditional panel estimators. In a slight abuse of language, we will call θ the strict-exogeneity parameter and λ the contemporaneous-exogeneity parameter. When we say that strict (contemporaneous) exogeneity holds, we technically mean that $\theta = 0$ ($\lambda = 0$).[10] When we say that strict (contemporaneous) exogeneity does not hold, we set $\theta = 0.2$ ($\lambda = 0.2$).

[8] The FE and FD versions of the test are two different ways of testing the same underlying assumption. They have no bearing on selecting between the FE or FD estimators. Under the null hypothesis, both tests should lead to nonrejection.

[9] Values of α_i are independently drawn from a standard normal distribution, and the initial value of the explanatory variable x is governed by the equation $x_{i0} = 0.40\alpha_i + u_{i0}$.

[10] To be precise, when contemporaneous exogeneity does not hold, it is automatically the case that strict exogeneity is violated. Thus, our language and notation are imprecise, but hopefully informative, because we focus on cases in which the key element of the strict-exogeneity assumption that *is not subsumed by* the contemporaneous-exogeneity assumption is/is not violated. The parameter θ is intended to capture the primary case of interest in which strict exogeneity is violated because of a correlation between the error term and *future* values of the explanatory variable (i.e., $E(\epsilon_{it} \mid x_{is}, \alpha_i) \neq 0$ for $s > t$).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

TABLE 1

Coefficients and Test Statistics using Simulated Data

Each panel in Table 1 reports statistics using simulated data, where the simulations are created using the data structure outlined in the text. All simulated data are panel-data sets with a firm-specific effect and a single explanatory variable. For each model, we create 1,000 simulated data sets with 1,000 firms (i.e., $N = 1{,}000$) in each data set for the indicated number of T periods (2, 5, or 10). By construction, the correlation of the firm-specific effect and the explanatory variable is 0.40. The parameter representing contemporaneous (strict) endogeneity is set equal to 0.20 in models in which the assumption of contemporaneous (strict) exogeneity is violated. The figures in columns 1–3 are the mean of the estimated coefficients on the explanatory variable using the indicated estimation method (i.e., ordinary least squares (OLS), fixed effects (FE), or first difference (FD)) in a model in which the actual coefficient on the explanatory variable is 0. These means are calculated across the 1,000 estimations, one for each of the 1,000 simulated data sets. The asterisks indicating significance levels in these columns are based on the mean p-value for the indicated coefficient across the same set of estimations. The figures reported in columns 4 and 5 indicate the percentage (in decimal form) of the 1,000 simulations in which the FE and FD tests for strict exogeneity outlined in the text indicate a rejection of the null hypothesis of strict exogeneity holding at the 5% level or better. These tests require at least 3 time periods, and thus the $T = 2$ cells are empty for these columns. *** indicates significance at the 1% level.

	OLS Bias	FE Bias	FD Bias	% FE Strict-Exogeneity Tests Significant	% FD Strict-Exogeneity Tests Significant
	1	2	3	4	5
Panel A. Contemporaneous Exogeneity Holds, Strict Exogeneity Holds					
$T = 2$	0.400***	0.001	0.001		
$T = 5$	0.400***	0.000	0.001	0.040	0.006
$T = 10$	0.399***	−0.000	−0.000	0.053	0.002
Panel B. Contemporaneous Exogeneity Violated, Strict Exogeneity Holds					
$T = 2$	0.543***	0.194***	0.194***		
$T = 5$	0.542***	0.192***	0.192***	0.053	0.005
$T = 10$	0.543***	0.192***	0.192***	0.059	0.001
Panel C. Contemporaneous Exogeneity Holds, Strict Exogeneity Violated					
$T = 2$	0.458***	−0.095***	−0.095***		
$T = 5$	0.458***	−0.039***	−0.096***	1.00	1.00
$T = 10$	0.456***	−0.019***	−0.096***	1.00	1.00
Panel D. Contemporaneous Exogeneity Violated, Strict Exogeneity Violated					
$T = 2$	0.573***	0.094***	0.094***		
$T = 5$	0.572***	0.150***	0.096***	1.00	1.00
$T = 10$	0.572***	0.168***	0.096***	1.00	1.00

For a given choice of θ and λ, we create simulated data sets of 1,000 firms for a panel length of T years, where we experiment with different T values. For each such data set, we estimate β using i) pooled OLS, ii) FE estimation, and iii) FD estimation, and we calculate p-values for the FE and FD versions of the tests outlined previously. For each permutation of θ, λ, and T, we create 1,000 such data sets. In Table 1, we report mean values for the estimates calculated over all 1,000 data sets, as well as the proportion of these simulated data sets in which the associated strict-exogeneity tests reject at the 5% level. The true value of the underlying β coefficient is set equal to 0, so the mean values provide a sense of the underlying inconsistency of a given estimator.

Panel A of Table 1 reports results in the ideal situation in which the contemporaneous- and strict-exogeneity assumptions are both valid ($\theta = 0$, $\lambda = 0$). As expected, the OLS coefficient is highly biased, falsely attributing the mechanical 0.4 correlation between the unobserved effect and the explanatory variable to a causal role for the explanatory variable. The FE and FD estimates are both extremely close to the true value of 0. The FE test falsely rejects strict exogeneity at the 5% level in slightly less than 5% of the models, whereas the FD test falsely rejects much less frequently. This relative advantage of the FD test in

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

terms of type 1 error may reflect the particular data-generating process that we have implemented.

In Panel B of Table 1, we consider the troublesome situation in which contemporaneous exogeneity is violated, but strict exogeneity holds ($\theta = 0$, $\lambda = 0.2$). This is a case in which only an untestable assumption is violated. In this case, we see that all of the estimation procedures are badly biased. However, the tests for strict exogeneity do not reject in an abnormal number of cases (Panels A and B reject at very similar rates), illustrating that the strict-exogeneity tests are silent on the well-known problem of contemporaneous endogeneity. This finding highlights the fact that strict exogeneity can be a distinct issue from the usual contemporaneous-endogeneity concerns.

In Panel C of Table 1, we consider the case in which contemporaneous exogeneity holds, but strict exogeneity is violated ($\theta = 0.2$, $\lambda = 0$). This is the case in which an important testable assumption is violated, a possibility ignored by almost all prior authors. Here we see that the FE and FD tests reject in virtually all of the simulated data sets, demonstrating that these tests can have substantial power. As the panel illustrates, with 2 time periods, the average bias in the FE and FD estimators (which are identical only in this case) is on the order of -0.094. When T increases, the FD estimator continues to display this level of bias, which is expected because differencing treats all consecutive observation pairs the same way. In contrast, the bias in the FE estimator declines to -0.039 when $T = 5$ and -0.019 when $T = 10$, illustrating Nickell's (1981) result that the inconsistency of the FE estimator declines as T grows. This provides some hope that FE estimators may be reasonably informative in long panels when the strict-exogeneity assumption is violated. However, reliance on this result requires stability in the underlying fixed effects, behavior that we later show is typically not evident in finance panels.

Finally, in Panel D of Table 1, we consider the case in which both contemporaneous and strict exogeneity are violated. The figures here are informative. The FE and FD tests reject in essentially all cases (see columns 4 and 5), indicating that adding the contemporaneous-exogeneity violation does not substantively affect the ability of these tests to detect strict-exogeneity violations. The estimated coefficients remain badly biased. The bias in the FE estimator does not decline as T lengthens. In fact, as the panel lengthens, the performance of the FE estimator is substantially poorer than that of the FD estimator.

Collecting this evidence, the tests we explore would appear useful in several situations.[11] First, consider the case in which the researcher is confident in the contemporaneous-exogeneity assumption. In this case, the tests provide evidence related to whether the FE and FD estimators are suitable for the task at hand (see Panel C of Table 1). In particular, rejection of the strict-exogeneity assumption should usually rule out these estimators. Second, consider the case in which a researcher has less confidence in the contemporaneous-exogeneity assumption but is confident that violations in contemporaneous and strict exogeneity

[11] The findings in Table 1 are substantively unaltered if we set the exogeneity parameters to 0.3 or 0.1 rather than 0.2, with generally larger inconsistencies when these parameters are set at larger values.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

appear concurrently. In this case, the strict-exogeneity tests may detect estimator inconsistencies, independent of the question of whether the contemporaneous-exogeneity assumption holds (see Panel D).

If either of the cases just described is common, the tests we advocate using may detect many inappropriate estimator choices in finance research. At the same time, the usual caution must be applied. If the tests do not reject, this does not necessarily imply that one can proceed with traditional estimators. When only contemporaneous exogeneity is in question, the tests are uninformative, and one must turn elsewhere for guidance (see Panel B of Table 1).

E. Testing for Strict Exogeneity in the IV Setting

In the IV case, variation in an instrument is used to identify the causal effects of the endogenous instrumented explanatory variable on the dependent variable. Similar to the traditional case, common panel-data IV estimation methods require not only that the contemporaneous value of the instrument is exogenous but also that the instrument's exogeneity holds for all leads and lags. Again, because of the possibility of including lagged values in the model, if the contemporaneous exogeneity of the instrument holds, the key concern is whether the current values of the dependent variable are related to the *future values* of the instrument. This relation could arise either because of feedback effects or differential timing in the response of the dependent variable and the instrument to common shocks.

Using our earlier notation, and denoting by z a candidate instrument for variable x, strict exogeneity of the instrument requires $\mathrm{E}(\epsilon_{it}|z_{is},\alpha_i)=0$ for $s>t$, which is the analog of the requirement discussed in the traditional setting, with the instrument z replacing the endogenous variable x. If the strict exogeneity of the instrument is violated, common panel-data IV techniques, including FE-2SLS and first-difference 2SLS (FD-2SLS) will be inconsistent (Wooldridge (2010)). The intuition for this inconsistency is similar to the traditional case, as outlined earlier, but with the added subtlety that variation in the instrument is used to identify exogenous variation in the explanatory variable.

Because FE-2SLS and FD-2SLS both converge to the underlying true parameter value when strict exogeneity holds, comparing these two estimates to detect whether there may be an issue would appear prudent (and low cost) in all cases. Again, if a violation is suspected, a formal test is in order. Wooldridge (2016) suggests a test in the IV setting that is analogous to the corresponding FE test in the traditional case outlined previously. In particular, one can estimate the augmented model $y_{it}=\alpha_i+x_{it}\beta+z_{i(t+1)}\delta+\epsilon_{it}$ by FE-2SLS, using as instruments both z_{it} and $z_{i(t+1)}$ (additional control variables can also be added to the equation). The test for strict exogeneity of the instruments is based on a test of whether $\delta=0$. If the test rejects, this would indicate a relation between the error term and the future values of the instrument, ruling out the instrument for consistent estimation using the usual FE-2SLS or FD-2SLS procedures.

Because the mechanism through which an apparently exogenous instrument may fail the strict-exogeneity assumption can be subtle, we defer discussion until later when we consider two specific examples. Many recent panel-data IV analyses in finance use instruments that would appear to be exogenous, at least in the usual contemporaneous sense. Typically, the researcher devotes significant

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

attention to justifying the contemporaneous-exogeneity assumption (i.e., the exclusion restriction). The point that we emphasize is that many panel-data IV methods require a stronger assumption, and the key extra component of this assumption is testable. If instruments that at first glance would appear contemporaneously exogenous are sometimes later invalidated, checking for the stronger notion of strict exogeneity would appear prudent given that a test is available. (Again, passing the test does not eliminate the need to consider the usual contemporaneous-exogeneity issue.)

III. Testing for Strict Exogeneity in Common Finance Research Settings

A. Identifying a Set of Canonical Traditional-Setting Regressions

To explore these issues in traditional settings, we identify a set of regression models from the associated literature. To do this, we assign each panel-data study flagged in our earlier literature search into a topic category based on the main dependent variable(s) of interest. An article can be assigned to more than one category if it contains several dependent variables. The 5 largest categories have dependent variables of the following type: i) a firm's investment choice, ii) a firm's leverage/capital structure, iii) managerial compensation or ownership, iv) a firm's annual fundraising choice, and v) a measure of firm performance (stock returns, accounting performance, Tobin's Q, etc.). In total, 64% of the published panel-data studies we identify can be placed into one or more of these five categories.

For ease of exposition, we will refer to the selected dependent variables as DEPVAR1 through DEPVAR5. These variables capture, in order, measures of leverage, fundraising activities, investment spending, firm performance, and managerial incentives. For each of the 5 dependent-variable categories, we identify two common dependent-variable constructs and add the letters *A* and *B* to the end of the DEPVAR notation to indicate the two variables selected for that category, for example, book leverage (DEPVAR1A) and market leverage (DEPVAR1B).

We model each dependent variable as a function of the independent variables appearing in the associated model of Gormley and Matsa (2014) or, when a Gormley and Matsa (2014) model is not available, independent variables that appear in at least 20% of all associated published studies identified in our literature search. We restrict attention to variables that can be constructed using standard data sets (Compustat, Center for Research in Security Prices (CRSP), ExecuComp), and we apply some subjective judgment in collapsing similar variables together. The number of independent variables varies depending on the dependent variable, with a maximum of 6.

Rather than discuss each of the independent variables in detail, we report in Table 2 a listing of the independent variables used in modeling each dependent-variable category. Details regarding variable constructions/definitions are relegated to the Appendix. None of these choices should be controversial. We are simply trying to collect and characterize a large and varied literature in a succinct manner. At times, we refer to these as INDVAR1 through INDVAR6, where the mapping in Table 2 can be used to recover the underlying economic constructs.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

TABLE 2

Variables and Models Selected for Tests of Strict Exogeneity in Traditional Settings

In Table 2, all dependent variables are measured as of time t, where stock variables are measured as of the end of fiscal year t, and flow variables are measured over the course of year t. All variables are constructed from Compustat, Center for Research in Security Prices (CRSP), or ExecuComp data. Normalizations of flow (stock) dependent variables are divided by assets at the start (end) of the period. All variable definitions, constructions, and timing conventions are detailed in the Appendix. The dependent variables are selected based on our survey of all articles published in three elite finance journals over a recent time period in which the authors use unit-level unobserved effects. The independent variables are selected from corresponding models in Gormley and Matsa (2014) or, when no such model exists, the variables that appear most commonly in the associated literature based on our literature survey.

Dependent Variable	INDVAR1	INDVAR2	INDVAR3	INDVAR4	INDVAR5	INDVAR6
DEPVAR1A = BOOK_LEVERAGE	$TOBIN_Q_t$	LOG_SALES_t	ROA_t	Z_SCORE_t	$LOG_MARKET_CAP_t$	$TANGIBILITY_t$
DEPVAR1B = MARKET_LEVERAGE	$TOBIN_Q_t$	LOG_SALES_t	ROA_t	Z_SCORE_t	$LOG_MARKET_CAP_t$	$TANGIBILITY_t$
DEPVAR2A = DEBT_ISSUANCE	$TOBIN_Q_{t-1}$	LOG_SALES_t	ROA_t	Z_SCORE_{t-1}	$LOG_MARKET_CAP_{t-1}$	$TANGIBILITY_{t-1}$
DEPVAR2B = EQUITY_ISSUANCE	$TOBIN_Q_{t-1}$	LOG_SALES_t	ROA_t	Z_SCORE_{t-1}	$LOG_MARKET_CAP_{t-1}$	$TANGIBILITY_{t-1}$
DEPVAR3A = CAPEX	$TOBIN_Q_{t-1}$	LOG_ASSETS_{t-1}	Z_SCORE_{t-1}	$CASH_FLOW_t$	$CASH_{t-1}$	$TANGIBILITY_{t-1}$
DEPVAR3B = R&D	$TOBIN_Q_{t-1}$	LOG_ASSETS_{t-1}	Z_SCORE_{t-1}	$CASH_FLOW_t$	$CASH_{t-1}$	$TANGIBILITY_{t-1}$
DEPVAR4A = TOBIN_Q	$R\&D_t/ASSETS_{t-1}$	LOG_SALES_t	ROA_t			
DEPVAR4B = ROA	$TOBIN_Q_{t-1}$	LOG_SALES_t	$R\&D_t/ASSETS_{t-1}$			
DEPVAR5A = OWNERSHIP	$TOBIN_Q_t$	ROA_t	$STOCK_RETURNS_t$	LOG_ASSETS_t	$VOLATILITY_t$	
DEPVAR5B = CEO_COMPENSATION	$TOBIN_Q_{t-1}$	ROA_t	$STOCK_RETURNS_t$	LOG_ASSETS_{t-1}	$VOLATILITY_t$	

B. Testing for Strict Exogeneity in Identified Traditional-Setting Regressions

For each dependent variable, we select one or more of the identified explanatory variables and conduct the FE and FD traditional-setting strict-exogeneity tests outlined earlier. Each test is for a standard linear panel-data model in which the dependent variable is a linear function of the unit-level (i.e., firm) unobserved effects, year dummies, and the selected explanatory variable(s). As discussed earlier, these tests add theoretically irrelevant leading values of the independent variables to the regression and test whether the associated coefficients are 0, as should be the case if strict exogeneity holds. Thus, evidence of a nonzero coefficient, or set of coefficients, is taken as evidence against the strict-exogeneity assumption.

We conduct these tests using the entire universe of available Compustat data, excluding financials and utilities, from 1965 to 2012. All variables are winsorized at the 1% tails, and all nonnormalized variables are inflation adjusted throughout the article. We report p-values for these strict-exogeneity tests in Table 3. The dependent variable for each model is listed in the left column, and the other column headings indicate the test that is conducted, with FE1 (FD1), for example, indicating the fixed-effects (first-difference) version of the test with INDVAR1 as the sole independent variable (in addition to the year effects). The FEJ and FDJ test designations are for the pooled versions of these tests derived from models that include all of the selected explanatory variables jointly together.

As the figures in Table 3 indicate, the vast majority of the p-values are below 0.01, indicating that in most cases, strict exogeneity can be rejected with a high degree of confidence. In fact, of the 104 tests conducted on individual explanatory variables, we can reject the strict-exogeneity assumption at the 1% level in more

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

TABLE 3
Tests of Strict Exogeneity in Identified Traditional Settings

Each cell of Table 3 reports *p*-values pertaining to two tests of a null hypothesis of the validity of the strict-exogeneity assumption. The first (second) number in each cell is the *p*-value for the fixed-effects (FE) (first-difference (FD)) regression test outlined by Wooldridge (2010) in which theoretically irrelevant values of explanatory variables are included in the regression equation. All standard errors in the derivation of these *p*-values are clustered at the firm level and are thus robust to arbitrary heteroscedasticity and serial correlation. The rows indicate the dependent variables in the models that are tested, and the columns indicate the independent variables. The number following the FE/FD designation indicates the sole explanatory variable that is included in the regression equation using the mapping in Table 2 (in addition to year dummies). In these tests, the *p*-value is based on the *t*-statistic for the coefficient on the theoretically irrelevant term in the regression equation. The FEJ/FDJ test is for a model in which all explanatory variables are included together, and *p*-values, in this case, are for an *F*-test of the joint significance of the entire set of theoretically irrelevant coefficient terms. The sample includes all Compustat firms with available data from 1965 to 2012 with the exception of financials and utilities.

	Model Used for Test						
	FE1/FD1	FE2/FD2	FE3/FD3	FE4/FD4	FE5/FD5	FE6/FD6	FEJ/FDJ
DEPVAR1A = BOOK_LEVERAGE	0.000/0.001	0.000/0.000	0.006/0.000	0.000/0.000	0.000/0.002	0.000/0.977	0.000/0.000
DEPVAR1B = MARKET_LEVERAGE	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.294/0.177	0.000/0.000
DEPVAR2A = DEBT_ISSUANCE	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000
DEPVAR2B = EQUITY_ISSUANCE	0.000/0.000	0.000/0.000	0.000/0.759	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000
DEPVAR3A = CAPEX	0.000/0.000	0.000/0.000	0.000/0.000	0.047/0.000	0.000/0.000	0.000/0.000	0.000/0.000
DEPVAR3B = R&D	0.000/0.000	0.000/0.001	0.187/0.841	0.000/0.000	0.000/0.001	0.000/0.000	0.000/0.000
DEPVAR4A = TOBIN_Q	0.000/0.000	0.000/0.000	0.002/0.000				0.000/0.000
DEPVAR4B = ROA	0.882/0.000	0.000/0.008	0.000/0.001				0.000/0.000
DEPVAR5A = OWNERSHIP	0.000/0.000	0.039/0.012	0.000/0.001	0.496/0.000	0.005/0.000		0.002/0.000
DEPVAR5B = CEO_COMPENSATION	0.000/0.813	0.107/0.000	0.000/0.000	0.000/0.041	0.000/0.841		0.000/0.000

than 85% of the reported models. Moreover, in the joint tests that include all explanatory variables, the strict-exogeneity assumption is rejected at the 1% level in all models. Clearly, in the models we explore, violations of strict exogeneity are quite common.

To explore further, we conduct the preceding analysis for the entire sample of firms restricted to nonoverlapping 10-year subsamples, followed by an analysis for all years with subsamples restricted to single industries based on 1-digit Standard Industrial Classification (SIC) codes. Because this subsample analysis generates multiple *p*-values for each dependent variable–independent variable combination (one for each subsample), we tabulate the median *p*-value for the associated set of test statistics.

As we would expect given the smaller sample sizes, the *p*-values in both panels of Table 4 are generally higher than the corresponding values in Table 3. However, it is quite remarkable that the median *p*-values are below 0.05 for more than half of the reported models (58.17%), indicating that, even in these smaller subsamples, there is substantial evidence against a maintained assumption of strict exogeneity. As the final column in Table 4 indicates, in the joint tests that include all explanatory variables, the tests reject strict exogeneity at the 5% level or better in 35 out of 40 of the models. If we couple these observations with strong a priori theoretical reasons to believe that strict exogeneity will be violated, the general case for suspecting that most FE and FD estimates in these types of models are inconsistent is convincing.

One may be concerned that peculiarities in our modeling choices may be driving the results. To investigate, we have experimented with i) a more aggressive winsorization at the 5% tails rather than 1%, ii) trimming (dropping) the 1% and 5% tail observations rather than winsorizing, and iii) eliminating all winsorization of the data. We have also experimented with restricting the sample

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

TABLE 4
Subsample Tests of Strict Exogeneity

Each cell of Table 4 indicates the median *p*-values for tests of strict exogeneity, where each test is conducted over the indicated subsample. The first (second) number in each cell is the median *p*-value for the fixed-effects (FE) (first-difference (FD)) regression test outlined by Wooldridge (2010) in which theoretically irrelevant values of explanatory variables are included in the regression equation with the median calculated over the set of *p*-values derived from the indicated set of subsamples. All standard errors in the derivation of these *p*-values are clustered at the firm level and are thus robust to arbitrary heteroscedasticity and serial correlation. The rows indicate the dependent variables in the models that are tested, and the columns indicate the independent variables. The number following the FE/FD designation indicates the sole explanatory variable that is included in the regression equation using the mapping in Table 2 (in addition to year dummies). In these tests, the *p*-value is based on the *t*-statistic for the coefficient on the theoretically irrelevant term in the regression equation. The FEJ/FDJ test is for a model in which all explanatory variables are included together, and *p*-values, in this case, are for an *F*-test of the joint significance of the entire set of theoretically irrelevant coefficient terms.

	Model Used for Test						
	FE1/FD1	FE2/FD2	FE3/FD3	FE4/FD4	FE5/FD5	FE6/FD6	FEJ/FDJ
Panel A. 10-Year Subsamples							
DEPVAR1A = BOOK_LEVERAGE	0.005/0.069	0.100/0.000	0.287/0.000	0.000/0.232	0.000/0.165	0.345/0.077	0.000/0.000
DEPVAR1B = MARKET_LEVERAGE	0.074/0.001	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.403/0.000	0.000/0.000
DEPVAR2A = DEBT_ISSUANCE	0.000/0.000	0.000/0.231	0.008/0.000	0.000/0.001	0.001/0.249	0.001/0.002	0.000/0.000
DEPVAR2B = EQUITY_ISSUANCE	0.004/0.117	0.000/0.208	0.000/0.001	0.000/0.111	0.000/0.000	0.000/0.004	0.000/0.000
DEPVAR3A = CAPEX	0.009/0.389	0.000/0.025	0.008/0.000	0.000/0.000	0.000/0.000	0.000/0.000	0.000/0.000
DEPVAR3B = R&D	0.000/0.000	0.000/0.054	0.512/0.235	0.049/0.000	0.214/0.225	0.000/0.100	0.000/0.000
DEPVAR4A = TOBIN_Q	0.000/0.082	0.000/0.017	0.023/0.040				0.000/0.000
DEPVAR4B = ROA	0.145/0.019	0.004/0.177	0.077/0.378				0.000/0.040
DEPVAR5A = OWNERSHIP	0.026/0.000	0.272/0.174	0.051/0.016	0.789/0.017	0.111/0.110		0.177/0.370
DEPVAR5B = CEO_COMPENSATION	0.000/0.536	0.228/0.003	0.000/0.000	0.000/0.332	0.000/0.195		0.000/0.000
Panel B. 1-Digit Industry Subsamples							
DEPVAR1A = BOOK_LEVERAGE	0.056/0.170	0.153/0.000	0.113/0.000	0.001/0.173	0.001/0.172	0.049/0.105	0.000/0.002
DEPVAR1B = MARKET_LEVERAGE	0.191/0.001	0.000/0.000	0.001/0.000	0.000/0.001	0.000/0.000	0.289/0.134	0.000/0.000
DEPVAR2A = DEBT_ISSUANCE	0.000/0.000	0.000/0.039	0.221/0.000	0.001/0.014	0.001/0.107	0.208/0.013	0.000/0.000
DEPVAR2B = EQUITY_ISSUANCE	0.322/0.012	0.000/0.182	0.188/0.156	0.000/0.352	0.000/0.000	0.000/0.001	0.000/0.000
DEPVAR3A = CAPEX	0.411/0.372	0.000/0.034	0.542/0.000	0.000/0.424	0.000/0.000	0.000/0.000	0.000/0.000
DEPVAR3B = R&D	0.022/0.052	0.006/0.243	0.425/0.260	0.614/0.034	0.308/0.256	0.157/0.041	0.021/0.129
DEPVAR4A = TOBIN_Q	0.000/0.035	0.000/0.282	0.019/0.202				0.000/0.002
DEPVAR4B = ROA	0.141/0.271	0.006/0.120	0.628/0.279				0.028/0.033
DEPVAR5A = OWNERSHIP	0.138/0.029	0.416/0.346	0.056/0.272	0.514/0.269	0.299/0.304		0.192/0.328
DEPVAR5B = CEO_COMPENSATION	0.031/0.448	0.500/0.085	0.007/0.000	0.000/0.373	0.048/0.425		0.000/0.002

to the post-1980 and post-1990 time periods and with estimating the models on only the industries that were excluded from the initial sample (utilities and financials). Finally, in all cases in which we could identify a common alternative definition/construction of an independent variable, we have substituted this alternative into the estimated models. In all cases, the results with these model or sample alterations are substantively unchanged from what we report in Table 3. Thus, the evidence against strict exogeneity in the models we consider appears quite robust.

For those who have taken the estimates in most of the associated past literature at face value, our findings should be concerning because they raise serious questions about the validity of many reported estimates. On the other hand, for skeptics, our evidence confirms suspicions of the widespread presence of poorly identified coefficients independent of the usual vexing issue of contemporaneous endogeneity. Of course, when our tests do not reject, the usual philosophical debate on contemporaneous endogeneity must be addressed. Practically speaking, our evidence suggests that in many settings, this last introspective step is unnecessary.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

C. Magnitudes of Potential Inconsistencies in Traditional Settings

Our preceding evidence of widespread violations of the strict-exogeneity assumption in many traditional finance panel-data models is concerning. However, if the resulting inconsistencies in the FE or FD estimators are small, conclusions regarding the magnitude of a coefficient of interest may be at least approximately valid. Although it is difficult to make more precise statements without understanding the dynamic structure of the underlying variables, some potentially useful information can be inferred by comparing the FE and FD estimates. Large differences would suggest an inconsistency of substantial magnitude because both estimators should converge to the true parameter value when the maintained model assumptions hold.

To investigate, we compare the magnitudes and signs of the corresponding FE and FD coefficient estimates for all of the models and samples in Tables 3 and 4. For the entire set of coefficient pairs, we first calculate the percentage of pairs in which the FE and FD coefficient estimates are of opposite sign and also the percentage of cases in which the two coefficients both differ in sign and are significant at the 10% level or higher. In addition, we calculate the median ratio of the larger-magnitude coefficient to the smaller-magnitude coefficient in the subset of cases in which both coefficients have the same sign.

As we report in Panel A of Table 5, the figures reveal that FE and FD estimates are not infrequently of opposite sign. This is concerning because we would expect two informative estimators applied to the same data to agree in sign. It would be particularly concerning if these two estimators yield significant coefficients of opposite sign. If the true coefficient is 0, the likelihood of observing two coefficients of opposite sign at the 10% significance level is 0.5% under the extreme assumption of independence. In general, a positive correlation between the estimators and a nonzero underlying coefficient should lead to a rate substantially below this 0.5% level. As we report in Panel B in the estimated models, this outcome is not nearly as rare as would be expected, with rates in many cases far above the 0.5% threshold. If we pool across all coefficient pairs included in this panel, we arrive at a rate of 9.3%, a figure that is more than 18 times the expected rate under these conservative assumptions.

Turning to the subset of cases in which the FE and FD estimates are at least of the same sign, we report in Panel C of Table 5 that the median ratio of the larger-magnitude estimate to the smaller is frequently quite large. Pooling across all dependent variables, the median of these medians is 1.63. This evidence indicates that substantial differences in estimated coefficients are quite common, even after excluding the extreme cases in which the point estimates have opposite signs. Collecting the evidence, it is apparent that FE and FD estimates often deviate substantially from one another. This suggests that the magnitude of the error in economic inferences from using FE and/or FD estimates in traditional settings when strict exogeneity is violated could, in many cases, be quite substantial.

D. Hoping That a Long Panel Solves the Problem in Traditional Settings

Although the FE and FD estimators are both inconsistent when strict exogeneity fails, one might expect the degree of inconsistency of the FE estimator to be small in a long panel given Nickell's (1981) result that inconsistency of the

TABLE 5
Comparing FE and FD Estimates

In Table 5, we compare pairs of fixed-effects (FE) and first-difference (FD) estimates for the same explanatory variable for the set of subsamples included in Table 3 based on either period subsamples or industry subsamples. In models explaining the indicated dependent variable as a function of the indicated independent variable (plus year effects), we identify the fraction of all cases in which the FE and FD estimates differ in sign (in Panel A) or differ in sign and are both significant at the 10% level or higher (in Panel B). In Panel C we restrict attention to cases in which the FE and FD estimates on the independent variable have the same sign and report the median value for the ratio of the larger-magnitude (absolute value) FE or FD coefficient to the smaller-magnitude coefficient in the pair over the set of all subsamples.

	INDVAR1	INDVAR2	INDVAR3	INDVAR4	INDVAR5	INDVAR6
Panel A. Proportion of Cases in Which FD and FE Estimates Differ in Sign						
DEPVAR1A	0.000	0.278	0.000	0.000	0.000	0.000
DEPVAR1B	0.000	0.111	0.000	0.000	0.000	0.000
DEPVAR2A	0.000	0.056	0.056	0.000	0.000	0.222
DEPVAR2B	0.000	0.556	0.000	0.000	0.278	0.167
DEPVAR3A	0.000	0.000	0.056	0.111	0.000	0.389
DEPVAR3B	0.000	0.000	0.000	0.057	0.389	0.222
DEPVAR4A	0.167	0.000	0.056			
DEPVAR4B	0.167	0.000	0.000			
DEPVAR5A	0.252	0.583	0.255	0.083	0.417	
DEPVAR5B	0.167	0.000	0.000	0.389	0.167	
Panel B. Proportion of Cases in Which FD and FE Estimates Differ in Sign and Both Significant at 10%						
DEPVAR1A	0.000	0.056	0.000	0.000	0.000	0.000
DEPVAR1B	0.000	0.111	0.000	0.000	0.000	0.000
DEPVAR2A	0.000	0.056	0.000	0.000	0.000	0.111
DEPVAR2B	0.000	0.556	0.000	0.000	0.222	0.000
DEPVAR3A	0.000	0.000	0.056	0.004	0.000	0.389
DEPVAR3B	0.000	0.000	0.000	0.000	0.167	0.056
DEPVAR4A	0.056	0.000	0.000			
DEPVAR4B	0.000	0.000	0.000			
DEPVAR5A	0.000	0.250	0.167	0.083	0.167	
DEPVAR5B	0.111	0.000	0.000	0.389	0.000	
Panel C. Median Ratio of Larger-Magnitude FE/FD Coefficient to Smaller-Magnitude FE/FD Coefficient if Same Signs						
DEPVAR1A	1.496	2.164	1.315	1.128	1.142	1.161
DEPVAR1B	1.278	1.341	1.387	1.257	1.143	1.174
DEPVAR2A	1.268	1.985	1.718	1.756	1.243	2.236
DEPVAR2B	1.098	2.605	1.081	1.774	5.429	3.648
DEPVAR3A	1.120	2.773	1.928	1.439	1.417	1.839
DEPVAR3B	1.237	2.169	1.286	1.487	1.971	1.781
DEPVAR4A	4.198	1.266	2.211			
DEPVAR4B	1.750	1.343	1.113			
DEPVAR5A	4.372	2.647	4.593	3.970	3.550	
DEPVAR5B	1.651	1.616	1.365	1.925	3.813	

FE estimator is on the order of $1/T$. Our earlier evidence using simulated data illustrates the decline in the asymptotic bias of the FE estimator for large T when only strict exogeneity is violated. Thus, hope remains that FE coefficients may be relatively informative in some finance settings with long panels.

The derivation of the $1/T$ result requires the unobserved firm heterogeneity governing the dependent variable to be stable over the sample period. However, given the occasional sharp changes that often occur in firms' management, ownership, and asset structures via control activity and other shocks, the assumption of stable unit-level unobserved effects is questionable.[12] To investigate, we calculate correlations in estimated firm-specific intercepts derived from models estimated over different subperiods. In particular, for each dependent variable, we derive model estimates from the FE estimator using the entire set of associated

[12] Some authors explicitly account for this possibility by allowing for different subperiod unit-level unobserved effects (e.g., Hoechle, Schmid, Walter, and Yermack (2012), Chen and Chen (2012)).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

independent variables for nonoverlapping 10-year subperiods (5-year periods in the case of the incentive variables), starting with the most recent observation year and then rolling backward. We collect the firm-specific intercepts for firm i in period p using the equation

$$\hat{\alpha}_{ip} = \bar{y}_{ip} - \bar{x}_{ip}\hat{\beta}_p. \quad (6)$$

[13]

The resulting pairwise correlations are reported in Table 6 for each of the estimated models. As the figures indicate, for all models, there is strong evidence of a monotonic decline in the correlation of the firm-specific intercepts as the estimation time periods grow further apart. For the furthest lags, most of the correlations (7 of 10) are below 0.50, and in some cases, they are quite small (below 0.20). Certainly, this does not offer strong support for the notion that unit-level unobserved characteristics governing the dependent variable are stable over long time periods. When we make the corresponding calculations for 5-year estimation

TABLE 6

Correlation of Firm-Specific Intercepts for Different Time Periods

For each dependent variable in Table 6, we calculate the fixed-effects (FE) estimator for a regression with the modeled dependent variable indicated in the top-left cell of each set of figures regressed against all of the independent variables outlined in Table 2 (plus year effects) for 10-year nonoverlapping periods starting at the end of the sample and working backward. In the case of the OWNERSHIP and CEO_COMPENSATION variables, we use 5-year periods. For each model and time period, we estimate firm-specific intercepts and report in each cell the simple correlation of these estimated coefficients for the same firm across different subperiods, where P1 is the most recent period, and P5 (or P4) is the most distant time period.

Period	P1	P2	P3	P4	P5	Period	P1	P2	P3	P4	P5
BOOK_LEVERAGE						*MARKET_LEVERAGE*					
P1	1.000	0.684	0.338	0.201	0.075	P1	1.000	0.629	0.375	0.288	0.280
P2	0.684	1.000	0.567	0.308	0.066	P2	0.629	1.000	0.557	0.357	0.279
P3	0.338	0.567	1.000	0.607	0.294	P3	0.375	0.557	1.000	0.637	0.410
P4	0.201	0.308	0.607	1.000	0.629	P4	0.288	0.357	0.637	1.000	0.703
P5	0.075	0.066	0.294	0.629	1.000	P5	0.280	0.279	0.410	0.703	1.000
DEBT_ISSUANCE						*EQUITY_ISSUANCE*					
P1	1.000	0.391	0.178	0.190	0.195	P1	1.000	0.421	−0.027	−0.077	−0.047
P2	0.391	1.000	0.407	0.344	0.239	P2	0.421	1.000	0.020	−0.107	−0.087
P3	0.178	0.407	1.000	0.272	0.166	P3	−0.027	0.020	1.000	0.312	0.236
P4	0.190	0.344	0.272	1.000	0.106	P4	−0.077	−0.107	0.312	1.000	0.278
P5	0.195	0.239	0.166	0.106	1.000	P5	−0.047	−0.087	0.236	0.278	1.000
CAPEX						*R&D*					
P1	1.000	0.775	0.755	0.701	0.694	P1	1.000	0.759	0.667	0.522	0.518
P2	0.775	1.000	0.602	0.590	0.648	P2	0.759	1.000	0.845	0.670	0.636
P3	0.755	0.602	1.000	0.784	0.731	P3	0.667	0.845	1.000	0.803	0.712
P4	0.701	0.590	0.784	1.000	0.831	P4	0.522	0.670	0.803	1.000	0.855
P5	0.694	0.648	0.731	0.831	1.000	P5	0.518	0.636	0.712	0.855	1.000
TOBIN_Q						*ROA*					
P1	1.000	0.645	0.422	0.384	0.369	P1	1.000	0.739	0.678	0.565	0.577
P2	0.645	1.000	0.425	0.269	0.282	P2	0.739	1.000	0.674	0.600	0.656
P3	0.422	0.425	1.000	0.588	0.426	P3	0.678	0.674	1.000	0.616	0.590
P4	0.384	0.269	0.588	1.000	0.522	P4	0.565	0.600	0.616	1.000	0.539
OWNERSHIP						*CEO_COMPENSATION*					
P1	1.000	0.796	0.578	0.449		P1	1.000	0.645	0.519	0.455	
P2	0.796	1.000	0.649	0.514		P2	0.645	1.000	0.649	0.514	
P3	0.578	0.649	1.000	0.738		P3	0.519	0.649	1.000	0.738	
P4	0.449	0.514	0.738	1.000		P4	0.455	0.514	0.738	1.000	

[13]Results are similar if we estimate models that use single explanatory variables, one at a time, rather than models with all variables included together. Results are also substantively unchanged if we calculate firm-specific subperiod intercepts using sample-wide slope coefficients rather than subperiod slopes.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

periods, the same basic story emerges (figures not tabulated). Thus, it appears that as panels get longer, any assumption of constant firm-level heterogeneity becomes highly questionable.

To investigate the relative performance of FE and FD estimators when unit-level unobserved fixed effects are not stable, we augment our earlier simulations to incorporate the presence of stochastic shocks to these effects. To create data sets with this added feature, we allow a given unit-level fixed effect to experience a shock each year with probability p (i.e., SHOCK_{it} is distributed binomially $(1, p)$). When the fixed effect experiences a shock, the new fixed effect is assumed to equal the old effect plus an IID N(0,1) increment denoted by η_{it}. Thus, using the same notation as in our earlier simulations,

$$\alpha_{it} = \alpha_{i(t-1)} + \eta_{it} \times \text{SHOCK}_{it}. \tag{7}$$

All other features of the simulations are the same as in our earlier analysis.

We simulate 1,000 data sets, each with 1,000 firms, for varying panel lengths ($T = 2$, 5, 10, or 15) and different shock probabilities ($p = 0.1, 0.2$, or 0.3). In all simulations, we set the contemporaneous-endogeneity coefficient to 0 and the strict-exogeneity coefficient to 0.2. Thus, these data sets are the analog of Panel C of Table 1, but with the addition of stochastic shocks to the unit-level unobserved fixed effects. As in that earlier table, we assume the underlying relation between the dependent variable y and the independent variable x has a true coefficient of $\beta = 0$, and we use the mean of the coefficient estimates across simulated data sets as a measure of asymptotic bias/inconsistency. The resulting statistics are reported in Table 7.

TABLE 7

Estimated Bias Using Simulated Data with Time-Varying Firm-Specific Intercepts

Each cell in Table 7 reports the mean coefficient on a single explanatory variable for a set of 1,000 simulated data sets, each containing 1,000 firms, in a balanced panel of T years using the indicated estimator (i.e., fixed effects (FE) or first difference (FD)). The structure of the simulations parallels the simulations in Table 1 except the firm-specific intercept changes with a probability p in each period where p is indicated in the column headings. Additional details are provided in the text. By construction, the correlation of the firm-specific intercepts and the explanatory variable is 0.40. As in Panel C of Table 1, in all models, the parameter representing contemporaneous endogeneity is set equal to 0 (i.e., contemporaneous exogeneity holds), and the parameter corresponding to strict endogeneity is set equal to 0.20.

	Intercept Shock					
	$p=0.1$		$p=0.2$		$p=0.3$	
	FE Bias	FD Bias	FE Bias	FD Bias	FE Bias	FD Bias
	1	2	3	4	5	6
$T=2$	−0.072	−0.072	−0.046	−0.046	−0.023	−0.023
$T=5$	0.019	−0.070	0.072	−0.047	0.121	−0.024
$T=10$	0.088	−0.071	0.180	−0.047	0.261	−0.024
$T=15$	0.141	−0.071	0.265	−0.047	0.369	−0.023

As expected, the figures in Table 7 indicate a constant level of bias for the FD estimator across all panel lengths. Things vary much more widely for the FE estimator. As T gets larger, the FE estimator becomes more biased in a positive direction. For $T = 10$ or greater, the magnitude of the bias is always larger than when $T = 2$. When $T = 15$ and $p = 0.3$, the bias is a stunning 0.369. With the exception of the case where $T = 5$ and $p = 0.1$, the bias in the FE estimator is always

larger in magnitude than the bias in the corresponding FD estimator, although this may reflect the particular data-generating process we have chosen.

Summarizing this evidence, it appears that i) unit-level fixed effects are rarely stable over long periods in finance panels, and ii) when these effects are unstable, both the FE and FD estimators can perform poorly even in long panels, with the FE estimator potentially performing substantially worse. Thus, when strict exogeneity is violated in traditional settings in finance research, both FE and FD estimators should be viewed with much caution because the magnitudes of resulting inference errors can be large.

IV. Strict Exogeneity with External Information and Instruments

Recent empirical research in finance emphasizes convincing identification. Some of this research exploits an explanatory variable that plausibly varies exogenously, for instance, because of time-varying shocks (law changes, natural disasters, etc.), and proceeds to use the exogenous explanatory variable directly in a panel regression. Our examination of 15 recent *Journal of Finance* articles identified earlier suggests that this approach remains much more common than formal panel-data IV regressions. The traditional case tests we outlined earlier are relevant for these types of studies. The hope is that the variable that appears to be exogenous does, in fact, pass these strict-exogeneity tests. As a first pass, the researcher can compare FE and FD estimates for the same underlying economic model.

In these (mostly recent) studies in which the key variable of interest may plausibly be exogenous in both a contemporaneous and strict sense, researchers frequently include control variables similar to the ones in the canonical regressions discussed earlier. Because, as we showed earlier, these control variables are often likely not to be strictly exogenous, this practice can be problematic because inclusion of these controls can render all of the coefficients inconsistent, even the coefficient on the key variables of interest that may, in fact, vary exogenously. Thus, it would appear prudent for researchers exploiting a (hopefully) exogenously varying explanatory variable of interest to estimate models with and without a full set of controls if the control variables are not strictly exogenous.

A. Internal Instruments

If external variation is not available to the researcher, one possibility is to exploit dynamic panel-data GMM estimators that rely on lagged instruments. This approach has the potential to yield consistent estimates in the presence of violations of strict exogeneity because these estimators do not maintain this restrictive assumption. In the literature search described earlier, we identify 27 articles that use this approach (compared with 251 using traditional FE/FD).

Although dynamic panel-data GMM estimators may be useful and informative, it is worth emphasizing that many of these estimators also rely on testable assumptions, and these assumptions often are not tested in finance research. For example, in the finance papers, we identify that utilize the Arellano and Bond (1991) and Blundell and Bond (1998) dynamic panel-data GMM procedures, far

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

less than half actually test the serial correlation assumption that is necessary for these estimators to yield consistent estimates.

A comprehensive evaluation of dynamic panel-data GMM techniques is beyond the scope of our article. However, in untabulated results, we have conducted AR(2) tests for the maintained underlying assumption of no serial correlation for all of the regression models presented in Table 3, estimated using the Arellano and Bond (1991) and Blundell and Bond (1998) estimators. These tests reject at the 5% level in more than 60% of the estimated models, suggesting that these estimators frequently depend on assumptions that are not valid.

B. External Information and Instruments

In an increasing number of panel-data finance studies, a researcher identifies a plausibly exogenous external variable that is related to the explanatory variable of interest and thus may serve as an instrument. The most common estimators in these settings are FE-2SLS and FD-2SLS. As discussed earlier, these estimators rely on the strict exogeneity of the instrument, a testable assumption that is almost never acknowledged because most authors focus on the issue of contemporaneous exogeneity in their discussion of the exclusion restriction.[14]

To consider this issue in practice, we identify two external IV examples that are inspired by recent studies and explore the strict-exogeneity issue in these specific contexts. Our discussion of the economic issues is brief because our goal is to illustrate the econometric issues in plausible IV settings, rather than to provide a definitive conclusion on specific economic questions.

C. Cash Flows, Weather, and Investment

In an influential article, Perez-Gonzalez and Yun (2013) consider the role of weather shocks in the cash flows of utilities and the consequent effect on a firm's motivation to hedge. Because weather shocks would appear to be quite exogenous, we consider the direct role of weather on cash flows (a traditional-setting panel regression) and the indirect role of weather as an instrument for cash flows in explaining capital expenditure decisions (an IV-setting panel regression). Perez-Gonzalez and Yun (2013) discuss many of the underlying economic issues.

In this context, contemporaneous exogeneity in the relation between a firm's cash flows and the weather is the requirement that weather shocks i) directly affect contemporaneous cash flows from the sales of gas and electricity and ii) are uncorrelated with innovations in any omitted variables. Because local weather shocks should be largely unpredictable (after controlling for sample-wide year effects and firm fixed effects), it would seem unlikely that current sales or cash flows would be related to future weather, so strict exogeneity appears likely to be satisfied in this context. However, a test could still be informative because it is possible, for example, that a particularly cold late December often continues into early January, in which case one year's sales/cash flow (which includes December, assuming a December fiscal year-end) actually would somewhat predict next year's measure of weather, thus violating strict exogeneity. Fortunately, the strict-exogeneity tests

[14]For an illuminating and detailed discussion of the challenges associated with justifying the contemporaneous-exogeneity assumption in finance research, see Atanasov and Black (2016).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

can help address this possibility. Parallel comments apply to the strict exogeneity of weather as an instrument for cash flow in explaining investment spending.

To investigate, we obtain zip-code-level data on heating degree days (HDDs) and cooling degree days (CDDs) for 1980–2016 from data vendor Frontier Data (for details on HDDs and CDDs, see Perez-Gonzalez and Yun (2013)). We match these data to the zip code of the headquarters of each electric or gas utility (SIC code 49) listed on Compustat. Because gas (electric) firms tend to be more sensitive to heating (cooling) demand, we run preliminary regressions of CASH_FLOW against HDDs and CDDs for i) gas utilities as a group, ii) electric utilities as a group, and iii) all others (e.g., combined electric and gas). For each utility, we create a WEATHER variable by aggregating the zip-code-level HDD and CDD variables by the relative size of the estimated coefficients for the utility's group type. This procedure creates a single weather variable (WEATHER) that should be comparable across all utilities (see the Appendix for details). Because Perez-Gonzalez and Yun (2013) report that the sales of some firms with utility SIC codes do not respond to weather shocks, or respond in the opposite-to-expected direction, we first regress each firm's CASH_FLOW against our weather variable and select for further study the set of utilities with estimated coefficients above the sample median.

For these weather-sensitive utilities, we first consider traditional panel regressions in which a firm's cash flow depends on the contemporaneous weather variable, along with firm and year effects. To test for strict exogeneity, we augment these models by including the leading weather term associated with the FE and FD tests for strict exogeneity. As we report in columns 1 and 2 of Panel A in Table 8, firm CASH_FLOW appears highly sensitive to contemporaneous weather using either the FE or FD estimator. Importantly, the coefficient on the leading term testing for strict exogeneity is small and insignificant for both estimators. Thus, it appears that any concerns about cash flows incorporating expectations about the future weather are not confirmed by the data. Utilities sell more gas/electricity in cold winters and hot summers, and current sales/cash flows do not predict future weather. This example illustrates the ability of our tests to help confirm the consistency of FE and FD estimates in a traditional panel context.

In columns 3 and 4 of Panel A in Table 8, we present a traditional panel regression of capital expenditures as a function of a utility's cash flow. Similar to our earlier analysis for all Compustat firms, we see that the CASH_FLOW coefficients differ sharply for the FE and FD estimators and also that the FE and FD tests for strict exogeneity both reject at high significance levels. This evidence is not surprising because capital spending this year may be related to nonweather components of expected future cash flows (e.g., anticipated changes in regulation, pricing, etc.). This example illustrates the ability of our tests to rule out FE or FD estimators in a traditional panel regression without entering into a debate regarding contemporaneous exogeneity.

Given that our WEATHER variable appears causally related to cash flows, it is a candidate as an instrument for CASH_FLOW in a model of the investment–

TABLE 8

Strict Exogeneity and Instrumental Variables

All coefficient estimates in Table 8 are for linear regression models predicting the indicated dependent variable. All models include year effects and account for firm fixed effects. Standard errors clustered at the firm level are reported in parentheses under each coefficient estimate. FE (FD) indicates a traditional firm-fixed-effects (first-difference) estimation. FE-2SLS (FD-2SLS) indicates a fixed-effects (first-difference) 2-stage least squares estimation. The sample in Panel A is composed of weather-dependent utilities identified by the data screens described in the text. In Panel A, the CASH_FLOW and capital expenditures (CAPEX) variables are normalized by the start-of-year assets. The WEATHER variable is the weighted sum of heating and cooling degree days over the fiscal year using weights that pertain to the type of utility, as described in the text and the Appendix, normalized by the start-of-period assets. All coefficients for variables with the prefix *CONTEMP* are for the value of the explanatory variable that is contemporary to the dependent variable. All coefficients for variables with the prefix *LEAD* are for the theoretically irrelevant leading value of the explanatory variable in the respective tests for strict exogeneity, as outlined in the text (see the timing pertaining to parameters γ, ρ, and δ for FE, FD, and FE-2SLS tests, respectively, as outlined in Section II). In Panel A, contemporaneous weather is used as an instrument for CASH_FLOW in columns 5 and 6, whereas contemporaneous and leading (next year's) weather are used as instruments in column 7. The sample in Panel B is composed of all firms from unconcentrated industries with OWNERSHIP data available in ExecuComp. FIRM_RISK is the standard deviation in market-adjusted daily returns over the fiscal year, and INDUSTRY_RISK for each firm-year is the median value of FIRM_RISK for all firms in the same 4-digit industry in the same year with the exception of the firm itself. In Panel B, contemporaneous INDUSTRY_RISK is used as an instrument for FIRM_RISK in columns 5 and 6, whereas contemporaneous and leading (next year's) INDUSTRY_RISK weather are used as the instruments in column 7. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable						
	1	2	3	4	5	6	7
Panel A. Cash Flows, Weather, and Investment							
	CASH_FLOW	CASH_FLOW	CAPEX	CAPEX	CAPEX	CAPEX	CAPEX
CONTEMP_CASH_FLOW			0.108***	0.292*	0.383**	0.445***	0.298**
			(0.025)	(0.159)	(0.154)	(0.176)	(0.151)
LEAD_CASH_FLOW			−0.054**	0.044***			
			(0.027)	(0.015)			
CONTEMP_WEATHER	3.473***	2.467***					
	(0.551)	(0.381)					
LEAD_WEATHER	0.920	0.009					0.007
	(0.691)	(0.078)					(0.165)
R^2	0.098	0.076	0.062	0.027	0.051	0.030	0.050
No. of obs.	1,730	1,583	1,802	1,650	1,881	1,719	1,731
Estimation method	FE	FD	FE	FD	FE-2SLS	FD-2SLS	FE-2SLS
Panel B. Firm Risk, Industry Risk, and Ownership							
	FIRM_RISK	FIRM_RISK	OWNERSHIP	OWNERSHIP	OWNERSHIP	OWNERSHIP	OWNERSHIP
CONTEMP_FIRM_RISK			1.721**	0.699*	1.154***	2.312***	0.257
			(0.798)	(0.395)	(0.435)	(0.671)	(0.307)
LEAD_FIRM_RISK			0.218***	−0.888***			
			(0.080)	(0.231)			
CONTEMP_INDUSTRY_RISK	0.405***	0.488***					
	(0.022)	(0.143)					
LEAD_INDUSTRY_RISK							1.054***
							(0.339)
R^2	0.588	0.446	0.039	0.007	0.043	0.008	0.044
No. of obs.	12,895	13,277	11,733	12,111	12,895	13,277	11,397
Estimation method	FE	FD	FE	FD	FE-2SLS	FD-2SLS	FE-2SLS

cash flow relation.[15] Again, it remains possible that the instrument is not strictly exogenous with respect to investment because current investment spending could reflect anticipation regarding future weather. In columns 5 and 6 of Panel A in Table 8, we present FE-2SLS and FD-2SLS estimates of the relation between

[15]The models in columns 1 and 2 of Panel A in Table 8 demonstrate that the relevancy condition is satisfied. This evidence is unaltered if the leading weather variable term is dropped from these models.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

(instrumented) CASH_FLOW and capital investment. As we see, the two estimates are quite close to one another, lending some support to the strict-exogeneity assumption. The formal test for strict exogeneity in column 7 reveals a small and insignificant coefficient on the relevant leading weather variable term, supporting the strict-exogeneity assumption for the WEATHER instrument in predicting investment spending (i.e., nonrejection of the null that the assumption holds).[16]

This example illustrates a case in which an apparently contemporaneous exogenous variable passes the test for the stronger notion of strict exogeneity both i) as an explanatory variable in a traditional regression explaining cash flow and ii) as an instrument in an IV regression explaining capital spending as a function of cash flow. We deliberately selected an example from the recent literature in which we suspected strict exogeneity would hold, but even here there are potential avenues for an assumption violation. The tests we suggest should be helpful in these situations in (hopefully) ruling out some scenarios that may concern the researcher.

D. Firm Risk, Industry Risk, and Ownership

Gormley and Matsa (2014) discuss the possibility of using industry shocks to an explanatory variable of interest as an instrument in a panel context. They offer words of caution on this approach because it requires the industry variation to be exogenous to the dependent variable (i.e., the exclusion restriction must hold). Despite the Gormley and Matsa (2014) concerns, it would appear that in some cases, industry shocks may include useful exogenous variation to exploit. However, if the researcher is convinced that he or she has identified such a case, a test of strict exogeneity should still be considered because this part of the underlying assumptions is testable.

To consider an example of this type, we explore the relation between a firm's risk and its level of managerial ownership. This relation has attracted attention dating back to the work of Demsetz and Lehn (1985), with different theories predicting different signs. Directly regressing ownership on firm risk will have all of the usual endogeneity problems, and our earlier tests demonstrate that strict exogeneity is likely violated in this specific context.

In many cases, shocks to industry risk would appear to capture underlying exogenous changes in an industry that would have no direct effect on ownership, other than through the firm risk channel (i.e., the exclusion restriction may hold). This suggests that industry risk innovations could be a reasonable IV for innovations in firm risk in explaining ownership. Ideally, one would want to look more closely at the underlying forces behind changes in industry risk because some are likely more plausibly exogenous (in a contemporaneous sense) than others (e.g., changes in industry supply price uncertainty are likely more exogenous than changes in industry concentration).

For our illustrative purposes, we will assume that industry volatility innovations are contemporaneously exogenous and ask whether they satisfy the more

[16]The Table 7 models do not include a full set of controls because we are concerned that these controls may be endogenous. However, if we include the control variables used in the corresponding Table 2 models (capital expenditures explanatory variables for Panel A, ownership explanatory variables for Panel B), the results we report in Table 7 are substantively unchanged.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

stringent strict-exogeneity requirement as an IV for FIRM_RISK in explaining inside ownership (OWNERSHIP). Note that there are multiple economic mechanisms that could generate a violation of strict exogeneity in this context, even if contemporaneous exogeneity is satisfied. One possibility is that future changes in the industry-risk component of firm risk are somewhat predictable and that managers take these expected risk innovations into account in their current ownership-position decisions. Alternatively, shocks to the underlying economics of an industry, for example, changes in acquisition activity, may affect ownership decisions immediately and industry volatility in the future as the market gradually learns about the new postconsolidation market equilibrium.

To investigate, for each firm-year, we calculate INDUSTRY_RISK as the median level of the standard deviation of daily market-adjusted stock returns over the fiscal year, using all other firms in the same 4-digit (SIC) industry. OWNERSHIP is the total percentage of shares held by all executives associated with the firm as listed in ExecuComp. Because we want to rule out the possibility that ownership affects firm risk, which in turn affects industry risk, we restrict attention to unconcentrated industries where feedback effects from a single firm's risk choices on the risk levels of others in the industry are unlikely.[17]

In the first two columns of Panel B in Table 8, we find that INDUSTRY_RISK is strongly related to FIRM_RISK using both FE and FD estimators, with both coefficients in the 0.40–0.50 range. Not surprisingly, there is a strong industry component to risk innovations, and the relevancy condition is clearly satisfied. In columns 3 and 4 of this panel, we again show the problem with regressing OWNERSHIP against FIRM_RISK. Sidestepping the contemporaneous-endogeneity issue, both the FE and FD tests for strict exogeneity in these columns strongly reject. If we drop the leading FIRM_RISK term and compare the FE and FD coefficients in these models, the FE coefficient is almost 8 times the FD coefficient (coefficients untabulated). Clearly, as demonstrated earlier, directly regressing OWNERSHIP against FIRM_RISK is problematic.

Given that industry-risk innovations are strongly correlated with changes in firm risk, they may serve as useful instruments in a panel model of the dependence of OWNERSHIP on FIRM_RISK. We sidestep the thorny issue of contemporaneous endogeneity, but certainly, in a full-blown study of this economic question, one would want to investigate the source of the shocks to industry risk to evaluate this issue. However, if strict exogeneity is violated, the usual FE-2SLS and FD-2SLS estimators will be inconsistent, independent of the answer to this preliminary question. To investigate, we report FE-2SLS and FD-2SLS estimates of the relation between (instrumented) FIRM_RISK and OWNERSHIP. As we see in columns 5 and 6 of Panel B in Table 8, the two IV estimates on FIRM_ RISK differ substantially (2.312 vs. 1.154). Again, under strict exogeneity, these estimators should converge to the exact same value in large samples. The formal test for strict exogeneity of the instrument in column 7 confirms this suspicion, with

[17] We classify an industry as unconcentrated if the sales-based Herfindahl index for all Compustat firms in the 3-digit industry is below the sample median in a given year. Results are similar if we include all industries.

a highly significant coefficient on the leading INDUSTRY_RISK variable indicating a strong rejection of the strict-exogeneity assumption.

This example illustrates a case in which an IV variable that the researcher may believe satisfies the usual notion of contemporaneous exogeneity can nevertheless be ruled out as an IV for FE-2SLS or FD-2SLS estimation because of a clear failure of strict exogeneity. The test we suggest should be helpful in similar IV panel situations in which a researcher may be tempted to rely on estimates that are inconsistent. If the tests reject, this may in some cases prompt the researcher to investigate more carefully the mechanism governing variation in the proposed instrument. In the case of industry-risk shocks, rejection could lead to a careful consideration of different types of industry-volatility innovations, some of which may be exogenous in both a contemporaneous and strict sense. Hopefully, by relying on the test we outline, the economic content of reported estimates in these types of IV settings will be enhanced.

V. Conclusion

In this article, we discuss the requirement of strict exogeneity for consistency when using both traditional panel-data estimators (FE and FD) and also common panel-data IV estimators (FE-2SLS and FD-2SLS). Strict exogeneity is a much stronger assumption than the usual notion of contemporaneous exogeneity. Loosely speaking, the assumption requires there to be no feedback from the dependent variable to the future values of the independent variable/instrument, and it also requires that the future values of the independent variable/instrument do not respond to the same shocks as the contemporaneous values of the dependent variable. Importantly, these elements of the strict-exogeneity assumption are testable, yet these tests have largely been ignored in empirical finance research.

We select a set of canonical regressions from the finance literature and show that the strict-exogeneity assumption is routinely rejected in traditional (i.e., non-IV) panel settings. In fact, in some cases, the FE and FD estimators of the same model reveal significant coefficients that are opposite in sign. Our evidence indicates that many of the reported FE and FD estimates in the large associated empirical finance literature are inconsistent estimates of the parameters of interest, even if the researcher is willing to accept the usual (and often untestable) assumption of contemporaneous exogeneity. We show that the magnitude of the resulting inconsistencies can be large, that longer panel lengths are unlikely to solve the inconsistency problem, and that it is often not possible to ordinally rank the relative performance of the FE and FD estimators.

Our findings for traditional panel-data estimators, which remain the most commonly used tools in finance panel settings, clearly indicate that the strict-exogeneity assumption should *always* be considered when using these estimators. Given the current trend toward identifying explanatory variables that are more convincingly exogenous, it is quite possible that these selected variables will often pass the tests for strict exogeneity in future research, thus increasing confidence in results based on their variation. We identify one example in which a plausibly exogenous variable does pass the test, namely, the role of weather shocks on a utility firm's cash flows.

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

Although external instrument IV panel studies in finance cannot be neatly categorized given the unique nature of the external variation that is typically exploited, we do identify two examples motivated by recent research in which a plausibly contemporaneously exogenous instrument is available. In one case, the instrument passes the test for strict exogeneity, namely, the case in which weather is used as an instrument for cash flow in explaining capital expenditure decisions. However, when we consider industry innovations in risk as an instrument for firm risk in explaining managerial ownership, the strict-exogeneity assumption is soundly rejected. These examples illustrate the potential usefulness of tests of strict exogeneity in IV settings. These tests can, in the best case, help confirm the appropriateness of a common IV estimator and, in the less favorable case, rule out these estimators as appropriate tools for the job.

It is important to emphasize that none of our findings or tests eliminates the usual need to consider contemporaneous-exogeneity issues. However, if a researcher is satisfied that he or she has addressed the usual concerns, our evidence strongly suggests that the strict-exogeneity issue should *always* be considered because violations can lead to severe inference errors both in traditional and IV panel settings. As we illustrate, some violations may be subtle, so casually dismissing the issue would appear quite risky.

We conclude our discussion of the strict-exogeneity issue with three recommendations. First, as an extremely low-cost diagnostic tool, finance researchers should routinely compare FE and FD estimates, or in the IV case, FE-2SLS and FD-2SLS estimates, because significant differences are one of the predictable consequences of a violation of strict exogeneity. Second, if this comparison suggests large differences, or alternatively, if there is a suspected economic mechanism that could generate a correlation between the dependent variable and the future values of the explanatory variable/instrument, simple formal tests of the type we outline should be performed. Third, in the case when a strictly exogenous explanatory variable or instrument is identified, the researcher should consider estimating models both with and without a laundry list of control variables. Because typical control variables in finance studies will often not be strictly exogenous, their inclusion could render all of the coefficients inconsistent, thus invalidating inferences regarding the key coefficient of interest. We are hopeful that following these recommendations will lead to more robust and informative inferences concerning financial behavior.

Appendix. Variable Definitions

For all of the dependent-variable categories besides the investment variables (capital expenditures, research and development (R&D)), we select the independent variables primarily from the corresponding models of Gormley and Matsa (2014) (GM hereafter), with some small alterations. For the leverage and fundraising regressions, we use the same explanatory variables as in the leverage regression of GM, with the log of market equity added as an additional measure of firm size. For the incentives variables (compensation, ownership), we use the same explanatory variables as in the compensation regression of GM, but for reasons of parsimony, we exclude the lags of the performance variables. We also exclude explanatory variables in GM that are related to person-specific characteristics because the OWNERSHIP variable we examine captures total managerial ownership rather

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press

than ownership by a single executive. For the performance-dependent variables, we exclude the Delaware variable used by GM because this variable rarely changes in our panel. We also reverse the roles of TOBIN_Q and return on assets (ROA) when ROA is used as the performance dependent variable in place of TOBIN_Q. The explanatory variables in the investment-spending regressions are selected from the most common choices identified in our literature search, as outlined in the text.

In creating the variables that appear in our regressions, our basic timing principle is to regress stock variables (i.e., balance sheet items) against contemporaneous stock and flow variables (i.e., measured at the same point in time, or over the year that has just ended, all referred to as measured as of year t). In the case of flow dependent variables (i.e., income statement or other variables that occurred over the year that ends at time t), we select flow explanatory variables also measured over time t and stock variables measured at the start of the flow period, which is time $t-1$ (i.e., the end of the prior fiscal year). In addition, flow variables that are normalized are divided by the start of period values (i.e., flow variable ending at time t divided by stock variable at the start of year t/end of year $t-1$), whereas stock variables are normalized by contemporaneous values (i.e., value at time t divided by value at time t). CEO_COMPENSATION, STOCK_RETURNS, and VOLATILITY are treated as flow variables, and OWNERSHIP and the Z_SCORE are treated as stock variables. The normalization of the inputs into the Z-score follows the traditional Z-score timing convention. Finally, we winsorize all variables except dummy variables at the 1% and 99% tails, and all nonnormalized monetary variables are inflation adjusted.

In Table A1, we provide the definitions of all of our variables using Compustat/CRSP/ExecuComp variable names. The prefix "L." indicates a 1-year lag so, for example, L.at indicates the lagged value of the at (total assets) variable from Compustat. All variables are inflation adjusted to 1971 dollars before entering these formulas.

TABLE A1

Table A1 provides variable definitions for all variables used in our analysis.

Variable Name	Definition/Construction
BOOK_LEVERAGE	(dltt + dlc)/at
MARKET_LEVERAGE	(dltt + dlc)/(at + prccf × csho − ceq)
DEBT_ISSUANCE	(dltt + dlc − (L.dltt + L.dlc))/L.at
EQUITY_ISSUANCE	(ceq − re − (L.ceq − L.re)/L.at
CAPEX	capx/L.at
R&D	xrd/L.at, where xrd is set equal to 0 if missing
OWNERSHIP	The sum, over all executives, of the percentage of shares owned, excluding option grants outstanding (shrown_excl_opts_pct)
CEO_COMPENSATION	log(tdc1)
TOBIN_Q	(at + (prccf × csho) − ceq)/at
ROA	ni/L.at
LOG_SALES	log(sale)
MODIFIED_Z_SCORE	6.56 × wcap/at + 3.26 × re/at + 6.72 × (ebit/at) + 1.05(ceq/lt)
LOG_MARKET_CAP	log(prccf × csho)
TANGIBILITY	ppent/at
CASH_FLOW	(oiadp − xint − txpd)/L.at
CASH	(che)/at
STOCK_RETURNS	exp[sum over days in year[log(1+ mkt.-adj. daily ret)]] − 1
VOLATILITY	Standard deviation of daily market-adjusted returns over fiscal year

Creation of WEATHER Variable

For each utility, we calculate the total number of heating and cooling degree days (HDDs and CDDs, respectively) during the fiscal year. We run an initial regression of CASH_FLOW (normalized by assets) against year effects, firm fixed effects, and both HDDs and CDDs for gas utilities as a group, electric utilities as a group, and all others as a group (most gas and electric combined). Denoting the estimated heating

(cooling) coefficient for the group as β_H (β_C), we calculate an initial weather variable as $(\beta_H)/(\beta_H+\beta_C)\times \text{HDD}+(\beta_C)/(\beta_H+\beta_C)\times \text{CDD}$, where the relevant coefficients used for a given firm are selected from the utility group the firm belongs to, and the weather data is for the zip code of the firm's headquarters. The variable WEATHER is this initial variable scaled by the start-of-year book assets.

References

Agrawal, A., and D. A. Matsa. "Labor Unemployment Risk and Corporate Financing Decisions." *Journal of Financial Economics*, 108 (2013), 449–470.

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

Arellano, M., and S. Bond. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58 (1991), 277–297.

Atanasov, V., and B. Black. "The Trouble with Instruments: Re-Examining Shock-Based IV Designs." Working Paper, Northwestern University (2016).

Blundell, R., and S. Bond. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87 (1998), 115–143.

Cameron, A. C., and P. K. Trivedi. *Microeconometrics: Methods and Applications*. New York, NY: Cambridge University Press (2005).

Chen, H. J., and S. J. Chen. "Investment-Cash Flow Sensitivity Cannot Be a Good Measure of Financial Constraints: Evidence from the Time Series." *Journal of Financial Economics*, 105 (2012), 393–410.

Dang, V. A.; M. Kim; and Y. Shin. "In Search of Robust Methods for Dynamic Panel Data Models in Empirical Corporate Finance." *Journal of Banking and Finance*, 53 (2015), 84–98.

Demsetz, H., and K. Lehn. "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy*, 93 (1985), 1155–1177.

Fischer, E. O.; R. Heinkel; and J. Zechner. "Dynamic Capital Structure Choice: Theory and Tests." *Journal of Finance*, 44 (1989), 19–40.

Flannery, M. J., and K. W. Hankins. "Estimating Dynamic Panel Models in Corporate Finance." *Journal of Corporate Finance*, 19 (2013), 1–19.

Gormley, T. A., and D. A. Matsa. "Common Errors: How to (and Not to) Control for Unobserved Heterogeneity." *Review of Financial Studies*, 27 (2014), 617–661.

Hennessy, C. A., and T. M. Whited. "Debt Dynamics." *Journal of Finance*, 60 (2005), 1129–1165.

Hermalin, B. E., and M. S. Weisbach. "Endogenously Chosen Boards of Directors and Their Monitoring of the CEO." *American Economic Review*, 88 (1998), 96–118.

Hoechle, D.; M. Schmid; I. Walter; and D. Yermack. "How Much of the Diversification Discount Can Be Explained by Poor Corporate Governance?" *Journal of Financial Economics*, 103 (2012), 41–60.

Iliev, P., and I. Welch. "Reconciling Estimates of the Speed of Adjustment of Leverage Ratios." Working Paper, University of California at Los Angeles (2011).

Matsa, D. A. "Capital Structure as a Strategic Variable: Evidence from Collective Bargaining." *Journal of Finance*, 65 (2010), 1197–1232.

Nickell, S. A. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49 (1981), 1471–1426.

Perez-Gonzalez, F., and H. Yun. "Risk Management and Firm Value: Evidence from Weather Derivatives." *Journal of Finance*, 68 (2013), 2143–2176.

Petersen, M. A. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies*, 22 (2009), 435–480.

Roberts, M. R., and T. M. Whited. "Endogeneity in Empirical Corporate Finance." In *Handbook of the Economics of Finance, Vol. 2*, G. Constantinides, M. Harris, and R. Stulz, eds. Amsterdam, Netherlands: Elsevier, North-Holland (2012).

Strebulaev, I. A., and T. M. Whited. "Dynamic Models and Structural Estimation in Corporate Finance." *Foundations and Trends in Finance*, 6 (2012), 1–163.

Thompson, S. B. "Simple Formulas for Standard Errors That Cluster by Both Firm and Time." *Journal of Financial Economics*, 99 (2011), 1–10.

Wintoki, M. B.; J. S. Linck; and J. M. Netter. "Endogeneity and the Dynamics of Internal Corporate Governance." *Journal of Financial Economics*, 105 (2012), 581–606.

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press (2010).

Wooldridge, J. M. "Linear Panel Data Models." Lecture Notes from Summer Tutorial in Modern Applied Tools of Econometrics, Michigan State University (2016).

https://doi.org/10.1017/S0022109018000996 Published online by Cambridge University Press