The Review of Financial Studies

Methodological Variation in Empirical Corporate Finance

Todd Mitton
Brigham Young University

I document large variation in empirical methodology in corporate finance regressions in top finance journals. Although methodological variation allows for customization of empirical tests to fit specific theories, it can also enable excessive reporting of statistically significant results. For example, given discretion over 10 routine methodological decisions, a researcher could report that over 70% of randomly generated variables are statistically significant determinants of leverage at the 5% level. The methodological decisions that affect statistical significance the most are dependent variable selection, variable transformation, and outlier treatment. I discuss remedies that can mitigate the negative effects of methodological variation. (*JEL* C18, C52, G30)

Received December 21, 2019; editorial decision January 30, 2021 by Editor Itay Goldstein.

In empirical corporate finance, researchers make many routine methodological decisions that influence the magnitude and statistical significance of their findings. Ideally, researchers are guided by theory when making methodological decisions, but when theory does not dictate the use of one particular method, researchers often must choose from among multiple methods that are widely used and accepted in the literature. When researchers have discretion over methodological decisions, published papers may represent findings in which the methodology employed has been selected from among many possible methodologies in order to report statistically significant results. Selective publication of statistically significant results can occur at two levels during the research process: among researchers and among publishers.

At the first level, researchers may select statistically significant results by choosing methods that allow them to report desired significant findings—a practice sometimes referred to as "*p*-hacking." In his presidential address

I thank two referees, Itay Goldstein, Campbell Harvey, Jason Kotter, Yaron Levi, Thomas O'Connor, Ivo Welch, and seminar participants at Brigham Young University and the University of Southern California for helpful comments. I thank Ramin Baghai, Lucian Bebchuk, Efraim Benmelech, Adam Bonica, John Graham, Barry Hirsch, Gerard Hoberg, Jay Li, Stephen McKeon, David Macpherson, Lalitha Naveen, Renana Peres, Gordon Phillips, Alessio Saretto, Matthew Serfling, Jared Smith, Robert Tumarkin, Jing Wang, Philip Valta, Ekaterina Volkova, Jin Xu, and Ayako Yasuda for generously sharing data. I thank Greg Adams, Huong Bui, Toshinori Fukui, John Garrett, Michael Garrett, Duc Ho, Whitney Holman, Jonathan Jensen, Mark Johnson, Anthony Kent, Paige Nelson, Marshall Ringwood, Nick Steil, Alex Tuft, and Joseph Zhang for excellent research assistance. Send correspondence to Todd Mitton, tm@byu.edu.

The Review of Financial Studies 35 (2022) 527–575

doi:10.1093/rfs/hhab030 Advance Access publication March 16, 2021

to the American Finance Association, Harvey (2017) outlines the challenges presented by p-hacking in the financial economics literature. Many recent papers have drawn attention to the possibility of p-hacking in the asset pricing literature, with a focus on cross-sectional return anomalies and how they stand up to replication, out-of-sample testing, or multiple hypothesis testing.[1] This paper focuses on the corporate finance literature, where these issues remain relatively unexplored.[2]

At the second level, publishers may select statistically significant results by maintaining a bias toward publishing papers that report significant findings. Kim and Ji (2015), Harvey, Liu, and Zhu (2016), Harvey (2017), and Morey and Yadav (2018) have documented and discussed how finance journals display a bias toward publishing statistically significant results.[3] When publishers are biased, methodological variation can lead to the publication of studies that are misleading about the significance of results, even when no individual researcher engages in p-hacking. In other words, methodological variation is problematic even when all researchers are ethical, diligent, and transparent. If multiple researchers independently study the same hypothesis, each using a different methodology, then findings based on a methodology that produces significant results might be published while other equally valid insignificant findings remain unknown (Denton 1985; Gelman and Loken 2014). Indeed, researchers may not even submit papers with statistically insignificant results because they are less likely to be published or cited—an effect sometimes referred to as the "file drawer problem" (Rosenthal 1979). When readers of articles are unaware of alternative findings that are discarded during the research process due to either p-hacking or publication bias, their inferences about the importance and robustness of findings can be highly distorted.

In this paper, I assess the impact of methodological variation on research in corporate finance. First, I study to what degree different methodologies are employed and accepted in the literature. In the top three finance journals between 2000 and 2018, I find a total of 954 regressions in 604 articles in which the dependent variables are among the most common corporate finance outcomes studied: profitability, firm value, leverage, investment, payouts, or cash holdings. (All regressions of the same category in a given article are counted as one regression.) The incidence of these types of regressions in the top three journals increased greatly over the sample period, from a total of eight regressions in 2000 to a total of 89 in 2018. I study this sample of regressions and document the methods used in common decisions related to sample selection, variable transformation, and model specification.

1 See, e.g., Harvey, Liu, and Zhu (2016), McLean and Pontiff (2016), Yan and Zheng (2017), Hou, Xue, and Zhang (2020), Linnainmaa and Roberts (2018), Chordia, Goyal, and Saretto (2019), and Chen (2020).

2 Mulherin, Netter, and Poulsen (2018) discuss p-hacking as a key concern in corporate finance publishing.

3 See also De Long and Lang (1992), Brodeur et al. (2016), Andrews and Kasy (2019), and Chen and Zimmerman (2020).

The results of my study indicate a high degree of methodological variation in empirical corporate finance. One key source of variability is a lack of standardization in dependent variable selection. For example, researchers have used 61 unique measures of profitability as dependent variables, including 26 unique definitions of return on assets (ROA). I find similar variability for other categories of regressions, with leverage regressions having the largest number of unique dependent variables (96) and cash regressions having the fewest (9). I also find that correlations between alternative dependent variables are often not very high—the median correlation among the 10 most common in each category is 0.33—suggesting that dependent variable selection can often have a large impact on regression results.

Another source of methodological variation is control variable inclusion. For example, in value regressions (e.g., with Tobin's q as the dependent variable), firm size is the only control variable that is used consistently, as it appears in 84% of the value regressions in the sample. Even then there is considerable variation in the measure of firm size used, from among assets, sales, or others. A few other control variables—investment, leverage, and profitability—are included about half of the time, and other control variables are included sporadically. Overall, I find little consistency in control variable usage in any of the categories of regressions.

I also document a lack of consistency in other methodological decisions. For example, researchers include all industries in their regressions roughly half of the time and exclude certain industries (e.g., financial firms) roughly half of the time. Over the sample period as a whole, researchers retain outliers about as frequently as they winsorize outliers, and they use a variety of cutoffs when treating outliers. I also report methodological differences in lagging variables, converting continuous variables to dummies, logging variables, and defining industries. For some decisions, such as outlier treatment, a consensus on methodology appears to be building over time, but for most decisions I find no trend toward a consensus in the literature.

Taken together, the results of my study show that researchers have a wide variety of methodologies to choose from when performing empirical tests. On one hand, this methodological variation can be helpful to researchers, allowing them to tailor empirical tests more precisely to the theory being tested. For example, the richness of databases like Compustat allows researchers to explore intricate details of a firm's financial and operating performance. To the extent that researchers base decisions on theory, the observed methodological variation could be entirely appropriate. On the other hand, if researchers are not guided by theory when choosing among methods—if methods are selected randomly (or even strategically)—then methodological variation enables the selective reporting that results from p-hacking and publication bias. To understand the degree to which methodology is guided by theoretical considerations, I search the sample of 604 articles for explanations of why methodological decisions are made. I find that authors routinely leave key decisions unexplained. For example, authors explain their selection of a dependent variable only 22% of the time, they explain their method of outlier treatment only 6% of the time, and they explain why they convert continuous variables to dummy variables only 19% of the time. Although researchers might have unarticulated theoretical motivations in mind when making some decisions, the available evidence suggests that the majority of methodological decisions are made without theoretical guidance.

Next, I study how methodological variation affects the statistical significance of coefficients on hypothesized determinants of profitability, firm value, leverage, investment, payouts, or cash holdings. My procedure is to first regress one of the outcome variables on a hypothesized determinant using the most common methods. Then I change one binary methodological decision—while keeping all other decisions at the most common methodology—and repeat the regression. I do this for 14 different decisions, and I record how much the t-statistic for the hypothesized determinant changes with each change in methodology. By repeating this process for a large number of hypothesized determinants, I can estimate the average impact of decisions on the statistical significance of explanatory variables.

In my first set of tests, the "hypothesized" determinants are purely random normally distributed variables. For each category of regression, I randomly generate 1,000 explanatory variables and test the impact of each methodological decision on the t-statistic for each variable. I report the average change in the t-statistic across the 1,000 variables for each methodological decision. Note that the average change in the t-statistic depends upon how disruptive each methodological change is to the data underlying the regression. I show that, in theory, with randomly generated explanatory variables, the expected change in the t-statistic can be as high as 1.13 (for disruptive changes) or as low as 0.00 (for innocuous changes). For example, a t-statistic would be expected to change by about 1.13 on average if (instead of a typical change in methodology) the original explanatory variable were replaced with a completely new randomly generated variable. By contrast, a t-statistic would be expected to change by about 0.00 on average for a very minor change such as rounding the explanatory variable to the second decimal place. In my tests, the actual methodological decisions have varying impacts within this range. For example, outlier treatment is a very disruptive decision. In profitability regressions, the decision to winsorize or retain outliers changes the t-statistic by 1.11 on average, implying that outlier treatment is almost as disruptive to the regression as if an entirely new explanatory variable were generated. Dependent variable selection is also very disruptive; in profitability regressions, changing the dependent variable from the most common measure of ROA to the most common measure of return on equity (ROE) changes the t-statistic by 0.93 on average. Toward the other end of the scale, the decision to use two-digit SIC industry dummies or Fama-French industry dummies has only a small impact on t-statistics, 0.10 on average in profitability regressions. I repeat these tests using other types of randomly generated variables—lognormally distributed variables, dummy variables, and difference-in-differences variables—and find similar results.

I also repeat these tests using "hypothesized" determinants that are quasi-random; I create the explanatory variables with actual Compustat data, but by creating a ratio variable from randomly selected Compustat data items, as in Yan and Zheng (2017) and Chordia, Goyal, and Saretto (2020). In these tests, the upper limit of the expected change in the t-statistic is no longer 1.13 because of underlying correlations among the Compustat data items, and my tests demonstrate that changes in t-statistics are much greater than those for purely random explanatory variables. For example, in profitability regressions, winsorizing outliers changes the t-statistic by 12.86 on average and changing the dependent variable from ROA to ROE changes the t-statistic by 12.31 on average.

I also repeat these tests using actual hypothesized determinants from the literature, focusing on leverage as the dependent variable. I compile a set of 65 proposed determinants of leverage from previous studies and observe how methodological changes affect the statistical significance of these findings. The magnitude of the impact on t-statistics lies somewhere between the impact for purely random explanatory variables and quasi-random Compustat variables. For example, winsorizing outliers changes the t-statistic by 3.74 on average, and changing the dependent variable from book leverage to market leverage changes the t-statistic by 3.91 on average.

Next, I document to what degree methodological variation can enable statistically significant findings. I consider whether a researcher (or a set of researchers independently working on the same question) could demonstrate that a given explanatory variable is a statistically significant determinant of, say, profitability, if the researcher has discretion over a set of methodological decisions. I show first that, using only the most common methodology, purely random explanatory variables are significant only as often as would be expected by random chance: about 10% of the time at the 10% level of significance, 5% of the time at the 5% level, and 1% of the time at the 1% level. Then I incrementally allow methodological discretion. Allowing the researcher one binary methodological decision—to use the most common dependent variable or the second most common dependent variable—gives the researcher two methodological combinations to choose from, and with this freedom the researcher could report statistical significance of randomly generated variables 15% of the time at the 10% level, 7% of the time at the 5% level, and 2% of the time at the 1% level (across all categories of regressions). These percentages increase progressively as more methodological discretion is allowed. When the researcher has discretion over 10 binary methodological decisions, 94% of randomly generated variables can be found significant at the 10% level with at least one methodological combination, 73% can be found significant at the 5% level, and 23% can be found significant at the 1% level.

The high percentages of hypotheses that can be found significant illustrate the sensitivity of statistical significance to methodology; however, they should not be interpreted as probabilities that a given hypothesis could be credibly supported. Safeguards against reporting spurious findings include robustness checks, the editorial review process, and the prospect of having fragile results challenged by later papers. Additionally, researchers can increase confidence in their findings by using multiple approaches to testing a single hypothesis. Nevertheless, given that 10 binary decisions are but a small subset of the methods available to researchers, these results suggest that great caution is warranted when judging the statistical significance of any single finding.

Finally, I discuss remedies for the excessive reporting of statistical significance that can arise from methodological variation. Robustness checks are the most commonly used defense against fragile findings, and I show to what degree they constrain reporting of spurious significant results. My findings on which methodological decisions are most impactful provide guidance on where researchers and reviewers should direct their attention when assessing robustness. I also discuss possible drawbacks of robustness checks. If applied indiscriminately they can lead to false negative findings. Additionally, the practice of robustness testing often fails to recognize that robustness is usually a matter of degree. To illustrate this, I test the 65 proposed leverage determinants using 512 of the most common methodological combinations. I find that only one of the 65 proposed determinants is statistically significant (at the 10% level or higher) across all 512 specifications. On average, each determinant is significant in 43% of the specifications. These findings suggest that researchers should focus less on defending the robustness of a result and more on understanding why a result is robust in some specifications and not in others.

An alternative to standard robustness checks is to report results from a broad range of methodological possibilities simultaneously, an approach sometimes referred to as "specification checks." Relative to robustness checks, specification checks have certain advantages: they are more systematic, they demonstrate the effect of changing methods along multiple dimensions simultaneously, and they convey a great deal of information concisely, often in graphical form. To demonstrate these advantages, I present some examples from the capital structure literature that graphically illustrate differing patterns of robustness for different proposed determinants.

Another recommendation is for researchers to focus less on statistical significance and more on the economic significance of results. Aside from being a better indication of the importance of empirical findings, economic significance is less susceptible to specification searching. I also discuss several other remedies. Together, the suggested remedies can help mitigate the negative effects of methodological variation.

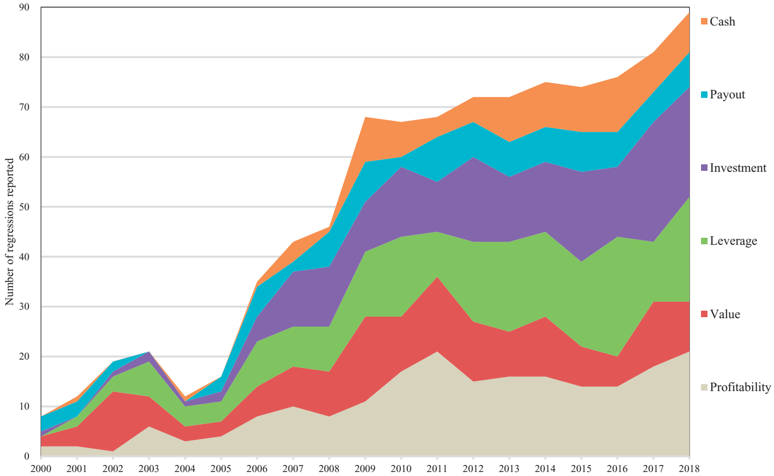

Figure 1
Corporate finance regressions in top journals
The number of corporate finance regressions (of the categories shown) reported in articles in the *Journal of Finance, Journal of Financial Economics*, and *Review of Financial Studies*, by year. Multiple regressions of the same category in an article are counted as one regression.

1. A Survey of Methodological Variation

To better understand current methodological practice in the corporate finance literature, I survey articles in top finance journals. Regressions in corporate finance study a wide variety of dependent variables, and I focus on six of the most common categories of regressions reported in the literature: those for which the dependent variable is profitability, firm value, leverage, investment, payouts, or cash holdings.

1.1 Sample of corporate finance regressions

I examine all regressions in the six common categories reported in the *Journal of Finance, Journal of Financial Economics*, and *Review of Financial Studies* from 2000 to 2018. Figure 1 shows the number of regressions reported in each of the six categories by year. In these statistics, a particular category of regression (e.g., a profitability regression) is counted only once in any given article, regardless of how many different specifications or robustness checks are reported in the article. The total number of regressions in the sample from all categories is 954. Figure 1 shows that these types of regressions became a much more integral part of the literature over this period. In 2000, only eight regressions were reported in all six categories combined. In subsequent years, the number of these regressions increased dramatically, with over 80 per year in both 2017 and 2018.

The total number of articles published per year in these journals increased over this period, from 181 in 2000 to 299 in 2018. However, even when scaling the number of regressions by the total number of articles, the incidence of regressions from all categories increased by more than a factor of seven, from 0.04 per article in 2000 to 0.30 per article in 2018.

1.2 Variability in dependent variables

Although each regression in the sample can be classified into a particular category of regression, the dependent variables used in each category vary widely. The availability of different measures is advantageous for research, because the dependent variable can be chosen to closely align with the theory being tested. For example, researchers perform leverage regressions with the common goal of understanding what factors contribute to a firm's usage of debt, but depending on the context, the dependent variable can be chosen from among total debt ratios, short-term debt ratios, long-term debt ratios, and others. However, flexibility in selecting numerators and denominators for a dependent variable could also lead to a proliferation of measures beyond what is necessary for matching the context of each theory tested. In this section, I document the occurrence of different dependent variables in my sample, and I report correlations between alternative dependent variables.

Table 1 reports statistics on dependent variable usage in the regressions in the sample. For each category of regression, the 10 most common dependent variables are listed, along with the number of occurrences for each and the corresponding percentage of total occurrences. Below the 10 measures, I report the number of other unique measures that are also used in the literature.[4] Some differences between dependent variables are not reflected in Table 1, including whether the dependent variable is industry adjusted or in first differences.

Panel A of Table 1 reports the findings for profitability regressions. The most commonly used dependent variable is EBITDA/total assets, but it is used in only 14% of the regressions. The other nine dependent variables listed are used in between 2% and 10% of the regressions. In total, 61 unique measures of profitability are used as dependent variables.

Panel B of Table 1 reports the findings for value regressions. Overall, 39 unique measures of firm value are used as dependent variables in the sample.[5] Among these 39 measures are 25 different measures of what is typically referred to as "Tobin's q," meaning some measure of the market value of assets scaled by some measure of the book value of assets.[6] The most commonly used measure of firm value (30% usage) is one of these 25 definitions of Tobin's q: total assets

[4] The total number of dependent variables reported in Table 1 exceeds the total number of regressions in the sample because articles often use multiple dependent variables in the same category.

[5] One of the 39 measures, excess value (i.e., of a diversified firm relative to single-segment comparables), is not included in the other statistics in panel B because of its specialized purpose in the valuation of diversified firms.

[6] By definition, the denominator of Tobin's q should be the replacement cost of assets. Although a handful of earlier papers use some measure of replacement cost in the denominator, I find no studies that attempt to calculate replacement cost subsequent to Gaspar and Massa (2007).

Table 1
Current practice in empirical corporate finance: Dependent variables

	Count	%Total		Count	%Total
A. Profitability			*D. Investment*		
EBITDA/TA	37	14%	CAPX/TA	78	32%
Net income/TA	25	10%	CAPX/Net PPE	31	13%
Operating income/TA	22	8%	R&D/TA	30	12%
Operating income before depreciation/TA	19	7%	CAPX	17	7%
Net income/BE	15	6%	R&D	11	4%
EBIT/TA	14	5%	(CAPX+R&D)/TA	9	4%
Net income before extraordinary items/TA	13	5%	CAPX/Sales	8	3%
EBITDA/Sales	8	3%	R&D/Sales	7	3%
Operating income/Sales	7	3%	Net PPE	7	3%
Net income/Sales	5	2%	(Change in Net PPE)/TA	3	1%
51 other unique measures	95	37%	39 other unique measures	46	19%
Total	260	100%	Total	247	100%
B. Firm value			*E. Payouts*		
(TA–BE+ME)/TA	47	30%	Dividends/TA	21	15%
(TA–BE–DT+ME)/TA	18	11%	Dividends/ME	15	11%
ME/BE	16	10%	Dividend payer dummy	13	9%
(ME+PS+Total debt)/TA	9	6%	Dividends/Net income	13	9%
(ME+Total debt)/TA	8	5%	Repurchases	9	6%
(ME+Total liabilities)/TA	6	4%	Dividends	7	5%
ME/TA	5	3%	Dividends/Sales	6	4%
Price/EPS	5	3%	Repurchases/TA	6	4%
ME/(BE+DT+ITC–PS)	2	1%	(Dividends+Repurchases)/TA	6	4%
(ME+Total liabilities+PS)/TA	2	1%	Repurchases/ME	6	4%
29 other unique measures	39	25%	23 other unique measures	40	28%
Total	157	100%	Total	142	100%
C. Leverage			*F. Cash holdings*		
Total debt/TA	79	27%	(Cash+Equivalents)/TA	57	62%
Total debt/(Total debt+ME)	28	9%	(Cash+Equivalents)/Noncash assets	13	14%
Long-term debt/TA	22	7%	Cash/TA	10	11%
Total debt/(TA–BE+ME)	9	3%	Cash+Equivalents	4	4%
Total debt	11	4%	(Cash+Equivalents)/Sales	3	3%
Total liabilities/TA	6	2%	Cash	2	2%
Total debt/(Total debt+ME+PS–DT–ITC)	5	2%	(Cash+Equivalents)/Net PPE	1	1%
Short-term debt/TA	5	2%	Cash/Noncash assets	1	1%
(Total debt–Cash)/TA	5	2%	(Cash–Total debt)/TA	1	1%
Total debt/(Total debt+BE)	5	2%			
86 other unique measures	123	41%			
Total	298	100%	Total	92	100%

The table reports the number of occurrences (and the percentage of the total) of dependent variables used in corporate finance regressions (of the categories shown) in the *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* between 2000 and 2018. Also reported for each category of regression are the number of additional unique measures used as dependent variables during the time period. Abbreviations in the table include TA for total assets, BE for book value of equity, ME for market value of equity, PS for preferred stock, DT for deferred taxes, and ITC for investment tax credits.

less book equity plus market equity all scaled by total assets. Another four of the 39 measures, including the third most popular measure, would typically be referred to as "market-to-book ratios," meaning market equity scaled by book equity or a variant.[7]

Panel C of Table 1 reports statistics for leverage regressions. The most common dependent variable in leverage regressions (27% usage) is total debt/total assets, where total debt refers to the sum of long-term debt and debt in current liabilities. In total, 96 unique measures of leverage are used as dependent variables.[8] The 96 measures are not all entirely interchangeable, because some studies focus on particular aspects of leverage, as noted earlier. But even in subcategories of leverage researchers have used multiple measures, including 52 different measures of total debt, five different measures of short-term debt, and eight different measures of long-term debt.

Panel D of Table 1 reports statistics for investment regressions. The dependent variable in investment regressions is usually a measure of capital expenditures or R&D (or a combination of the two) scaled by some measure of firm size, although unscaled measures are also used frequently. The most common dependent variable is capital expenditures/total assets (32% usage). Overall, I find 49 unique measures of investment used as dependent variables. Although physical investment and R&D are both categorized under investment (and are added together in some dependent variables), they measure two different types of investment and would often not be interchangeable. Researchers have many alternative measures to choose from for either type of investment. Among the 49 unique measures are 30 that focus on physical investment, 10 that focus on R&D, and nine that use a combination of physical investment and R&D.

Panel E of Table 1 reports statistics for payout regressions. Dependent variables in payout regressions measure dividends or repurchases (or both), often scaled by firm size, but sometimes unscaled. The most common dependent variable is dividends/total assets, but it is used in only 15% of the regressions. Not all measures categorized under payouts are interchangeable, because dividends and repurchases measure different aspects of the payout decision. But many alternative measures have been used for either dividends or repurchases. Among 33 unique measures of payouts in total, 15 focus on dividends, nine focus on repurchases, and nine use a combination of repurchases and dividends.

Panel F of Table 1 reports statistics for cash regressions, showing that cash and equivalents/total assets is the only dependent variable with majority status, as it is used in 62% of the cash regressions. Overall, cash regressions have the

---

[7] This nomenclature is not standardized—some authors refer to what is typically called Tobin's q as a market-to-book ratio, and others refer to what is typically called a market-to-book ratio as Tobin's q.

[8] The total of 96 unique measures includes nine different measures of debt maturity that are not included in the other statistics in panel C due to the different purpose of maturity as a dependent variable.

most consistent dependent variable usage; I find only nine different measures among the regressions in the sample.

1.2.1 Correlation of alternative dependent variables. Table 1 demonstrates the wide variety of dependent variables that are used, but if the alternatives are highly correlated, then the choice of dependent variable may not significantly alter regression results. Correlations among the most commonly used dependent variables are reported in Table 2. The table shows that, although some dependent variables are highly correlated, in general the correlations between alternative dependent variables are not particularly high, with a median (average) across all correlations in Table 2 of 0.33 (0.38).[9] Among the three most commonly used dependent variables in each category, the average correlations are 0.95 for profitability, 0.39 for firm value, 0.55 for leverage, 0.18 for investment, 0.63 for payouts, and 0.73 for cash holdings. Correlations among subcategories of measures tend to be higher than the overall correlation in a category, but not much higher. For example, the average correlation among all investment measures in panel D is 0.17, while the average among measures of physical investment is 0.21, and the average among measures of R&D is 0.26. The average correlation among all payout measures in panel E is 0.26, while the average among measures of dividends is 0.46, and the average among measures of repurchases is 0.15.

Even if the correlation between two alternative dependent variables is 1.00, substituting one for the other may not always give the same regression results, because observations can be missing in a database for one of the measures and not for the other. For example, EBIT/total assets and operating income/total assets have a correlation of 1.00, and they are almost always the same (within rounding error) in the Compustat data.[10] Nevertheless, in Compustat between 1963 and 2018, over 11,000 firm-year observations are missing EBIT but not operating income.

1.3 Variability in control variables

The regressions in the sample also exhibit a great deal of variation in control variable inclusion. Panel A of Table 3 reports the usage rates of the 10 most commonly used control variables among the 954 regressions in the sample. Panel A shows that firm size is by far the most commonly used control variable, as it appears in 79% of the regressions, and is the most common control in all six categories of regressions. Beyond firm size, there is little consistency in control variable usage. Reflecting the sometimes circular nature of empirical corporate finance, the next four most common control variables

[9] The correlations in Table 2 are reported after winsorization of the data at the 1st/99th percentiles. Without winsorization the median (average) correlation is 0.01 (0.19).

[10] Some sources consider EBIT and operating income to be identical, but depending on how nonoperating income is treated, they can be different.

Table 2
Correlations of commonly used dependent variables

	A. Profitability	1	2	3	4	5	6	7	8	9
1	EBITDA/TA									
2	Net income/TA	0.920								
3	Operating income/TA	0.994	0.927							
4	Operating income before dep./TA	1.000	0.921	0.994						
5	Net income/BE	−0.024	−0.038	−0.022	−0.027					
6	EBIT/TA	0.994	0.928	1.000	0.995	−0.026				
7	Net income before ex. items/TA	0.935	0.984	0.943	0.935	−0.039	0.943			
8	EBITDA/Sales	0.509	0.450	0.499	0.508	0.052	0.497	0.455		
9	Operating income/Sales	0.504	0.449	0.499	0.503	0.055	0.496	0.455	0.996	
10	Net income/Sales	0.501	0.506	0.496	0.500	0.056	0.494	0.500	0.937	0.940
	B. Firm value	1	2	3	4	5	6	7	8	9
1	(TA–BE+ME)/TA									
2	(TA–BE–DT+ME)/TA	0.999								
3	ME/BE	0.095	0.077							
4	(ME+PS+TD)/TA	0.986	0.983	0.130						
5	(ME+TD)/TA	0.983	0.980	0.135	0.998					
6	(ME+Total liabilities)/TA	0.999	0.997	0.098	0.985	0.985				
7	ME/TA	0.947	0.942	0.175	0.973	0.976	0.950			
8	Price/EPS	−0.095	−0.098	0.025	−0.095	−0.094	−0.094	−0.090		
9	ME/(BE+DT+ITC–PS)	0.084	0.071	0.876	0.118	0.122	0.088	0.161	0.026	
10	(ME+Total liabilities+PS)/TA	1.000	0.998	0.095	0.986	0.983	0.999	0.948	−0.095	0.084
	C. Leverage	1	2	3	4	5	6	7	8	9
1	TD/TA									
2	TD/(TD+ME)	0.483								
3	Long-term debt/TA	0.595	0.560							
4	TD/(TA–BE+ME)	0.555	0.897	0.690						
5	TD	0.066	0.194	0.130	0.143					
6	Total liabilities/TA	0.766	0.227	0.243	0.174	0.022				
7	TD/(TD+ME+PS–DT–ITC)	0.504	0.990	0.612	0.947	0.178	0.212			
8	Short-term debt/TA	0.758	0.259	−0.000	0.260	0.004	0.705	0.254		
9	(TD–Cash)/TA	0.907	0.576	0.609	0.627	0.099	0.661	0.596	0.658	
10	TD/(TD+BE)	0.217	0.508	0.459	0.510	0.144	−0.065	0.488	−0.003	0.285

(*Continued*)

Table 2
Continued

	D. Investment	1	2	3	4	5	6	7	8	9
1	CAPX/TA									
2	CAPX/Net PPE	0.423								
3	R&D/TA	−0.056	0.184							
4	CAPX	0.112	−0.056	−0.068						
5	R&D	−0.035	0.008	0.108	0.400					
6	(CAPX+R&D)/TA	0.539	0.407	0.794	0.005	0.061				
7	CAPX/Sales	0.517	0.257	0.070	0.054	−0.033	0.386			
8	R&D/Sales	−0.041	0.125	0.629	−0.048	0.045	0.486	0.281		
9	Net PPE	0.045	−0.126	−0.073	0.922	0.326	−0.041	0.019	−0.049	
10	(Change in Net PPE)/TA	0.398	0.187	−0.078	0.067	−0.002	0.151	0.236	−0.034	0.046
	E. Payouts	1	2	3	4	5	6	7	8	9
1	Dividends/TA									
2	Dividends/ME	0.689								
3	Dividend payer dummy	0.544	0.668							
4	Dividends/Net income	0.537	0.549	0.487						
5	Repurchases	0.086	0.041	0.162	0.052					
6	Dividends	0.268	0.239	0.293	0.244	0.492				
7	Dividends/Sales	0.679	0.609	0.388	0.474	0.029	0.206			
8	Repurchases/TA	0.085	0.002	0.054	0.003	0.401	0.067	0.035		
9	(Dividends+Repurchases)/TA	0.677	0.414	0.341	0.327	0.310	0.203	0.462	0.747	
10	Repurchases/ME	−0.001	0.004	−0.000	0.001	0.021	−0.000	−0.001	0.019	0.018
	F. Cash holdings	1	2	3	4	5	6	7	8	
1	(Cash+Equivalents)/TA									
2	(Cash+Equiv.)/Noncash assets	0.749								
3	Cash/TA	0.820	0.607							
4	Cash+Equivalents	0.022	−0.021	−0.023						
5	(Cash+Equivalents)/Sales	0.487	0.563	0.368	0.024					
6	Cash	0.002	−0.033	0.005	0.914	0.006				
7	(Cash+Equivalents)/Net PPE	0.500	0.576	0.411	0.076	0.354	0.038			
8	Cash/Noncash assets	0.658	0.720	0.844	−0.035	0.400	−0.023	0.441		
9	(Cash–TD)/TA	0.399	0.284	0.399	0.025	0.193	0.033	0.216	0.334	

The table reports correlation coefficients of the dependent variables most commonly used in corporate finance regressions (of the categories shown) in the *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* between 2000 and 2018. Data come from the Compustat database for the years 1963 to 2018. All variables are winsorized at the 1st/99th percentiles. Abbreviations in the table include TA for total assets, BE for book value of equity, ME for market value of equity, PS for preferred stock, DT for deferred taxes, TD for total debt, TL for total liabilities, and ITC for investment tax credits.

Table 3
Current practice in empirical corporate finance: Control variables

A. Most-common control variables

	Profitability	Value	Leverage	Investment	Payout	Cash	ALL
Firm size	81%	84%	87%	64%	80%	82%	79%
Profitability	21%	51%	73%	58%	65%	64%	53%
Value	33%	7%	66%	59%	54%	52%	45%
Leverage	38%	54%	20%	35%	45%	46%	38%
Investment	18%	54%	25%	13%	20%	39%	27%
Asset tangibility	9%	13%	52%	13%	12%	20%	21%
Firm age	23%	23%	11%	14%	14%	13%	17%
Growth	6%	17%	8%	18%	17%	10%	12%
Volatility	9%	15%	12%	4%	14%	13%	10%
Dividends	6%	11%	10%	4%	12%	33%	10%

B. Proxies for firm size

	Profitability	Value	Leverage	Investment	Payout	Cash	ALL
log(Total assets)	41%	55%	44%	35%	40%	53%	44%
log(Sales)	11%	11%	22%	8%	7%	14%	13%
log(Market value)	17%	7%	6%	7%	15%	5%	9%
Above measures, unlogged	4%	3%	3%	2%	5%	1%	3%
Other	3%	3%	5%	6%	6%	5%	5%
Multiple size controls	5%	4%	6%	6%	5%	5%	5%
None	19%	16%	13%	36%	20%	18%	21%

The table reports statistics on the usage of control variables in corporate finance regressions (of the categories shown) in the *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* between 2000 and 2018. Panel A reports usage rates of the 10 most commonly used control variables. Panel B reports usage rates of different measures of firm size as control variables.

are also common dependent variables: profitability (53% usage), firm value (45%), leverage (38%), and investment (27%).[11] The lack of consistency of control variable usage is compounded by the fact that many different proxies are used for each type of control variable, a source of variation not reported in panel A.

Because firm size is by far the most used control variable, panel B of Table 3 further delineates which measures of size are used. Panel B demonstrates the lack of standardization in size control usage. The most widely used measure of size is log(total assets), which is used in 44% of the regressions and is used most frequently in each of the six categories. The next two most prevalent size controls are log(sales) and log(market value), at 13% and 9%, respectively.

1.4 Variability in other methodological decisions

Researchers routinely face a number of other methodological decisions when testing hypotheses. The appropriate method for each decision ideally depends on the theory underlying the test or the nature of the data, but at times theory might be silent on a decision, or multiple theoretically acceptable alternatives

[11] Sometimes a category of regression also includes a control variable of the same category. The most common reason for this is the inclusion of a lagged value of the dependent variable as an explanatory variable. In other cases, the control is an industry average or the control represents a different aspect of the dependent variable (e.g., a control for capital expenditures when the dependent variable is R&D).

might be available. In Table 4, I report statistics on the alternatives chosen for many common decisions in my sample of articles.

Panel A of Table 4 reports statistics on industry inclusion. All industries are included in 46% of all regressions and 34% of those from 2016 to 2018. All other regressions exclude at least one industry from the sample, most commonly financial firms.

Panel B of Table 4 deals with the functional form of the key explanatory variable; that is, of the variable representing the primary hypothesis in each regression. Across all regressions, 55% of these variables are continuous and not logged, while 6% are continuous and logged. The remaining key explanatory variables are in dummy variable form, either naturally occurring dummy variables (25%) or dummy variables created from continuous variables (14%).

Panel C of Table 4 reports statistics on whether the explanatory variable is lagged by one period in the regression. Across all regressions, the contemporaneous explanatory variable is used 62% of the time, the explanatory variable is lagged 26% of the time, and both results are reported 4% of the time.

The next two panels in Table 4 deal with decisions on outlier treatment. Panel D reports the frequency with which researchers retain, winsorize, or trim outliers. The most common decision is to winsorize outliers, which is done in 48% of cases overall, and 62% of cases from 2016 to 2018. Across the entire sample, retaining outliers occurs almost as frequently as winsorizing, at 43%, and trimming outliers is less common, at 9%. Panel E reports the percentiles at which outliers are treated, conditional on outliers being either winsorized or trimmed. Cutoffs at the 1st and 99th percentiles are the most common, at 75% usage on average, but many other cutoffs, ranging from 0.5th/99.5th to 10th/90th, are used as well.[12]

Panel F of Table 4 reports statistics on the functional form of the dependent variable. Logging the dependent variable is uncommon in most categories, but in 25% of value regressions and 18% of cash regressions the dependent variable is logged. Payout regressions are unique in that the dependent variable is frequently a dummy variable.

Panel G of Table 4 addresses how ratio dependent variables are constructed when the ratio has a flow variable in the numerator and a stock variable in the denominator. The question for flow/stock variables is what denominator to use: the end-of-year measure, the beginning-of-year measure, or an average of the two. This is an issue primarily for profitability, investment, and payout regressions, although it occasionally arises in the other categories. Panel G shows that the end-of-year denominator is used most frequently (62% overall), with the beginning-of-year denominator used somewhat frequently (33%), and an averaged denominator used less frequently (4%).

---

12 Adams et al. (2019) discuss how outliers are treated in the finance literature and offer guidance for dealing with outliers.

Table 4
Current practice in empirical corporate finance: Other methodological decisions

	Profitability	Value	Leverage	Investment	Payout	Cash	ALL	ALL (2016–18)
A. Industry inclusion								
All	60%	52%	38%	39%	46%	30%	46%	34%
All except financial and utility	16%	28%	33%	34%	37%	46%	30%	34%
All except financial	13%	15%	17%	16%	14%	15%	15%	14%
Manufacturing only	3%	2%	6%	7%	1%	6%	4%	7%
Other	8%	4%	6%	5%	2%	3%	5%	11%
B. Key explanatory variable form								
Continuous—not logged	54%	59%	54%	56%	58%	51%	55%	51%
Dummy—naturally occurring	23%	20%	26%	25%	24%	32%	25%	30%
Dummy—created from continuous	15%	15%	13%	14%	16%	14%	14%	13%
Continuous—logged	8%	6%	7%	5%	2%	4%	6%	5%
C. Lag on explanatory variable								
Contemporaneous	58%	63%	65%	61%	67%	66%	62%	61%
Lagged	32%	19%	24%	30%	23%	16%	26%	30%
Both	4%	3%	3%	5%	3%	5%	4%	2%
Unclear	7%	15%	7%	4%	7%	13%	8%	7%
D. Outlier treatment								
Winsorize	49%	38%	48%	50%	49%	56%	48%	62%
Retain	43%	49%	43%	40%	47%	36%	43%	34%
Trim	8%	13%	9%	9%	4%	8%	9%	4%
E. Outlier cutoffs								
1st/99tth	74%	65%	75%	78%	82%	82%	75%	79%
5th/95th	9%	15%	5%	7%	6%	7%	8%	9%
0.5th/99.5th	5%	4%	8%	6%	2%	7%	6%	5%
2.5th/97.5th	3%	1%	2%	3%	4%	4%	3%	1%
2nd/98th	3%	2%	1%	0%	2%	0%	1%	1%
3rd/97th	1%	0%	1%	1%	0%	0%	1%	1%
10th/90th	1%	1%	0%	1%	0%	0%	1%	0%
Other/Not specified	4%	12%	8%	4%	4%	0%	6%	4%
F. Dependent variable form								
Continuous—not logged	97%	75%	91%	90%	73%	82%	87%	89%
Continuous—logged	2%	25%	8%	9%	4%	18%	10%	10%
Dummy	0%	0%	1%	2%	22%	0%	3%	0%
G. Denominator on flow/stock dependent variables								
End of year	70%	89%	40%	50%	76%	55%	62%	59%
Beginning of year	21%	11%	56%	49%	24%	41%	33%	37%
Averaged	9%	0%	4%	1%	0%	5%	4%	3%
H. Industry dummy definition								
2-digit SIC	24%	23%	29%	23%	19%	29%	25%	17%
Fama-French	12%	19%	21%	16%	14%	12%	16%	20%
3-digit SIC	8%	5%	13%	12%	12%	10%	10%	12%
1-digit SIC	3%	4%	4%	3%	7%	2%	4%	2%
4-digit SIC	2%	5%	5%	2%	5%	2%	4%	4%
NAICS	1%	0%	4%	2%	2%	0%	2%	2%
Other	8%	8%	6%	7%	7%	14%	8%	10%
Not specified	41%	38%	18%	34%	33%	31%	32%	34%

The table reports the percentage of articles using various methodological alternatives in corporate finance regressions (of the categories shown) in the *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* between 2000 and 2018.

Finally, panel H of Table 4 addresses the issue of what industry definitions are employed when regressions use industry fixed effects. About 46% of the papers in the sample (51% from 2016 to 2018) use industry fixed effects in at least some specifications. Across the entire sample, the most commonly used industry definitions are two-digit SIC, and the second most commonly used industry definitions are Fama-French, with Fama-French definitions slightly higher in popularity between 2016 and 2018.[13] However, the exact rates of usage are uncertain because 32% of papers using industry fixed effects are unclear about the industry definition used.

My survey of the literature does not cover all possible methodological decisions. Other important issues include instrumental variable selection (Harvey 2017), the use of interaction terms (Christensen and Miguel 2018) or subsamples (Gelman and Loken 2014), the time period to include in the sample, and the choice of estimation method (Harvey 2017).

1.5 Motivation for methodological decisions

The wide variety of methodologies documented in Tables 1, 3, and 4 raises the question of the motivation of researchers in choosing one methodological alternative over others. Methodological variation is helpful to the extent that researchers use different methods to accommodate specific theories. Different situations can imply different optimal procedures for outlier treatment, transforming variables, and other decisions. However, if researchers are not guided by theory, then methodological variation may simply add noise to the research process. To obtain some understanding of how researchers make these decisions, I study the sample of articles to determine whether authors provide explanations for their methodological decisions. I report whether authors provide a specific explanation for the decision, say that their decision follows prior literature, or provide no explanation for their decision. I do not attempt to evaluate the validity of the reasons stated.

Table 5 shows that for many methodological decisions, a large majority of articles provide no reason for the decision. For example, row 1 shows that when selecting a proxy for the dependent variable, authors state a reason for their selection 10% of the time, say that they follow prior literature 13% of the time, and provide no reason 78% of the time. Row 2 shows that in 63% of the cases in which firm size is included as a control variable, no explanation is given for doing so, and row 3 shows that in 92% of the cases, no explanation is given for the particular size proxy chosen.[14] Notably, row 7 shows that researchers convert continuous key explanatory variables to dummy variables in 131 cases, yet they state no reason for doing so 81% of the time, despite the fact that

13 Among papers using Fama-French definitions, 48 industries is the most common level of aggregation.

14 Kurshev and Strebulaev (2015) note, "Firm size has become such a routine control variable in empirical corporate finance studies that it receives little or no discussion in most research papers, even though it is not uncommon among the most significant variables." See also Lev and Sunder (1979).

Table 5
Current practice in empirical corporate finance: Explanations of methodological decisions

	Methodological decision	Related table	Number of occurrences	Specific reason stated	Follows prior literature	No reason stated
(1)	Why is the particular proxy for the dependent variable used?	1	954	10%	13%	78%
(2)	Why is firm size included as a control variable?	3A	755	18%	19%	63%
(3)	Why is the particular proxy for firm size used?	3B	755	1%	7%	92%
(4)	Why are financial firms excluded?	4A	429	28%	16%	56%
(5)	Why are utilities excluded?	4A	290	27%	18%	54%
(6)	Why is the key explanatory variable logged?	4B	53	26%	11%	62%
(7)	Why is a continuous key explanatory variable converted to a dummy variable?	4B	131	15%	4%	81%
(8)	Why is the explanatory variable lagged?	4C	263	50%	12%	38%
(9)	Why is the particular method of outlier treatment used?	4D	542	1%	4%	94%
(10)	Why are the outlier cutoffs chosen?	4E	542	1%	4%	95%
(11)	Why is the dependent variable logged?	4F	100	29%	18%	53%
(12)	Why is the beginning-of-year denominator used on flow/stock variables?	4G	150	7%	17%	77%
(13)	Why is the particular industry dummy definition chosen?	4H	275	3%	3%	94%

The table reports statistics on explanations given by authors for methodological decisions. Data come from 604 articles (including 954 regressions) in the *Journal of Finance, Journal of Financial Economics*, and *Review of Financial Studies* between 2000 and 2018.

doing so discards valuable information. Rows 10 and 11 show that researchers explain their decisions for outlier treatment and outlier cutoffs less than 10% of the time.

Thus, the available information suggests that a lack of theoretical guidance leaves a great deal of latitude for methodological choices in corporate finance. Of course, it is possible that authors omit explanations not for a lack of theoretical basis, but simply to avoid explaining what they view as routine or unimportant, or to reduce the length of a paper. On the other hand, even when authors provide theoretical explanations, it does not necessarily mean that theory restricts them to only one possible method. In any case, it should be emphasized that the absence of theoretical explanations does not imply unethical research practices. It does imply that research in corporate finance often requires choices among equally acceptable and defensible methodological alternatives. This methodological flexibility creates a challenge for inference, even among highly principled researchers.

1.6 Trends in methodological practice

Figure 2 shows trends in methodological decisions over time. I do not report trends for all of the decisions covered in Tables 1, 3, and 4, because many of the patterns are not particularly notable. Panel A shows usage rates for the

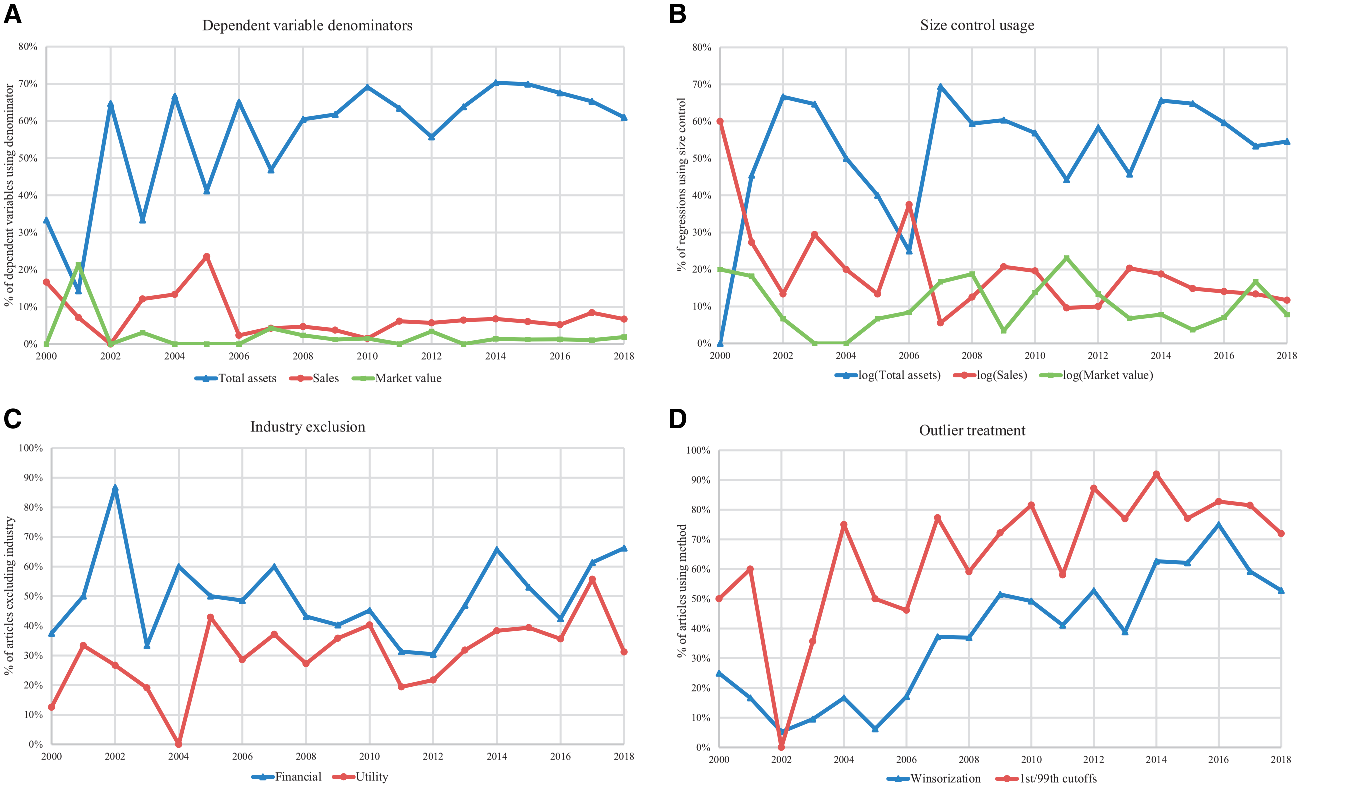

Figure 2
Trends in empirical corporate finance methods
The statistics are based on 954 regressions in 604 articles in the top three finance journals from 2000 to 2018. Panel A reports the percentage of dependent variables using various scaling variables (when a scaling variable is used). Panel B reports the percentage of regressions using various size controls. Panel C reports the percentage of regressions that exclude data from particular industries. Panel D reports the percentage usage of methods for outlier treatment.

most commonly used denominators (total assets, sales, market value) across all dependent variables in the sample (for regressions that have ratio dependent variables). It shows that total assets is the most commonly used denominator, especially in the last decade of the sample. Panel B reports size control usage conditional on a regression controlling for firm size. It shows that over time the literature has become more consistent in its use of log(total assets) as a size control. Panel C shows trends in the percentage of studies that exclude financial firms or utilities. These percentages have remained relatively stable over time; the literature has not reached a consensus on exclusion of these industries. Panel D shows that the practice of winsorizing variables has become more consistent over time. While the percentage of papers winsorizing data was usually below 20% in the earlier years of the sample, the percentage trended upward over the years to as high as 75% in 2016. Panel D also shows increasing consistency regarding outlier cutoffs. In more recent years, over 70% of studies that treat outliers (either by winsorizing or trimming) do so at the 1st and 99th percentiles.[15]

1.7 Most common methodology

Based on the data reported in Tables 1, 3, and 4, I specify the "most common methodology" as consisting of the following points: Use the most common dependent variable. Include all control variables that are used in a majority of regressions. Use log(total assets) as the size control. Include all industries in the sample. Do not log the explanatory variable. Use the contemporaneous explanatory variable. Winsorize non-indicator variables at the 1st/99th percentiles. Do not log the dependent variable. Use the end-of-year denominator on flow/stock dependent variables. Finally, when controlling for industry, use two-digit SIC dummies. These decisions serve as the baseline methodology for the tests in the next section.

2. Methodology and Statistical Significance

In this section I evaluate the impact of methodological variation on the statistical significance of coefficients from corporate finance regressions.

2.1 Data and summary statistics

I use data from the Compustat database for the dependent variables and control variables in my analysis. Compustat data are used in 77% of the articles in my sample. To avoid back-filling bias, I exclude observations prior to 1963, and I require that a firm appear in the dataset for two years before including it in the sample. The resulting dataset includes over 400,000 firm-year observations

---

15 Because Figure 2 is influenced by the composition of the sample, I also plot the trends for each category of regression separately (not reported). The trends are more erratic due to smaller sample sizes, but with a few exceptions, they follow the patterns shown in the sample as a whole.

between 1963 and 2018, although the number of available observations varies for different variables. Definitions of the Compustat variables are in Appendix Table A1, and summary statistics are in Appendix Table A2.

2.2 Changes in t-statistics

To assess the impact on statistical significance from using alternative methods, I perform panel regressions of the following form:

$$y_{ijt} = \alpha + \beta x_{ijt} + \mathbf{z}'_{ijt}\boldsymbol{\phi} + \gamma_j + \delta_t + \varepsilon_{ijt}, \tag{1}$$

where y_{ijt} is one of the dependent variables for firm i of industry j in year t, x_{ijt} is a hypothesized determinant of y, and $\mathbf{z}_{ijt}$ is a set of firm-level control variables. The term γ_j represents a set of industry fixed effects, and the term δ_t represents a set of year fixed effects. In some specifications, the industry fixed effects are replaced by firm fixed effects, γ_i. The coefficient of interest in the regression is β. In particular, we want to observe how the t-statistic on β changes when changing methods. The standard errors are clustered at the firm level. By testing a large number of hypothesized explanatory variables, I can measure the average effect of different methodological changes on t-statistics.

2.2.1 Purely random explanatory variables. I begin by testing randomly generated explanatory variables. Before proceeding with the tests, it is important to understand how much t-statistics should be expected to change when changing methods. To see this, define the random variable U as the t-statistic on β using the most common methodology, and the random variable V as the t-statistic on β using the same explanatory variable but an alternative method. With randomly generated explanatory variables, U and V are each asymptotically distributed as the standard normal. Then define the random variable $W = V - U$; that is, W is the difference in the t-statistic on β when using an alternative method compared to the most common methodology (on the same explanatory variable). The distribution of the changes in t-statistics is given by:

$$W \backsim N(\mu_V - \mu_U, \sigma_U^2 + \sigma_V^2 - 2\sigma_{UV}). \tag{2}$$

In the limit, with U and V each having the standard normal distribution, we have:

$$W \backsim N(0, 2 - 2\sigma_{UV}), \tag{3}$$

so that the distribution of changes in t-statistics depends upon the covariance between U and V.

Here it is helpful to consider two extreme possibilities. One extreme possibility is that U and V are identical, so that $\sigma_{UV} = 1$, giving W a variance of zero. In this case there is no difference in t-statistics when using the alternative method as compared to the most common methodology, which would occur when the methodological change is innocuous, causing little or no change to the data underlying the regression. As a hypothetical example, if the alternative method consisted of rounding the explanatory variable to the second decimal place, then we would expect a change in the t-statistic of approximately zero when changing the method.

A second extreme possibility is that U and V are independent, so that $\sigma_{UV}=0$, giving W a variance of two. In this case there is no relation between the t-statistics using the alternative method as compared to the most common methodology, which would occur if the methodological change were very disruptive to the data underlying the regression. As a hypothetical example, if the alternative method consisted of replacing the explanatory variable with a new randomly generated variable, then the t-statistic from the alternative method would be an independent draw from the distribution of t-statistics. In this case, the expected change in the t-statistic can be found by the mean absolute deviation (MAD) of the distribution of changes in the t-statistic. Because W is normally distributed, the MAD of W when $\sigma_{UV}=0$ is given by (e.g., Geary 1935):

$$MAD_W=\sqrt{\frac{2}{\pi}}\sigma_W=\sqrt{\frac{2}{\pi}(2)}=1.13. \tag{4}$$

In summary, if an alternative method is completely benign, the expected change in the t-statistic from changing the methodology is approximately zero. Alternative methods that disrupt the underlying data will have expected absolute changes in the t-statistic that are greater than zero and as high as 1.13. When the alternative method is so disruptive to the underlying data that the alternative regression is essentially independent of the original regression, then the expected absolute change in the t-statistic will be closer to 1.13. Thus, by empirically calculating average changes in t-statistics from making a particular methodological change (across many hypothesized explanatory variables), we can infer how disruptive the methodological change is by observing where the average absolute change in the t-statistic lies between zero and 1.13.[16]

To proceed with these tests, I randomly generate normally distributed explanatory variables by randomly selecting a mean between 10 and 100 and a standard deviation between 1 and 10. To assess the level of impact of each methodological decision, I first perform a baseline regression for each randomly generated explanatory variable, using the most common methodology. Then I perform 14 iterations of the same regression, in each iteration switching one and only one of 14 binary methodological decisions. By observing the t-statistic on the randomly generated explanatory variable in each iteration of the regression, I can assess the impact of each methodological decision on the statistical

16 MAD_W can be greater than 1.13 when $\sigma_{UV}<0$, but 1.13 is an appropriate upper benchmark for these purposes as it occurs when the two sets of t-statistics are unrelated.

significance of the key explanatory variable. I repeat this procedure for 1,000 randomly generated explanatory variables for each category of regression.

For purposes of comparison, I also report results for two extreme changes to the data, based on the earlier discussion. In the first, which I refer to as the "disruptive benchmark," I replace the explanatory variable with a new randomly generated explanatory variable. In the disruptive benchmark, the new explanatory variable would be expected to have no relation to the original explanatory variable, so that $\sigma_{UV} = 0$ and the expected change in the t-statistic would be approximately 1.13, as discussed earlier. In the second, which I refer to as the "innocuous benchmark," I round the explanatory variable to the second decimal place. In the innocuous benchmark, the regressions change very little, so that $\sigma_{UV} = 1$ and the expected change in the t-statistic would be approximately zero, as discussed earlier.

Summary statistics for the t-statistics generated from all of the regressions performed in the simulations (for the two benchmarks and the 14 methodological iterations) are reported in panel A of Appendix Table A3. The statistics confirm that the t-statistics are distributed as the standard normal, with mean t-statistics always close to zero and standard deviations of t-statistics always close to one. These statistics also confirm that the randomly generated variables are significant as often as would be expected, about 10% of the time at the 10% level, 5% of the time at the 5% level, and 1% of the time at the 1% level (though somewhat less often with firm fixed effects).

Table 6 reports the average changes in t-statistics when changing methodology, with panel A reporting results for industry fixed effects and panel B for firm fixed effects. Each row of data in the table corresponds to one methodological decision. Each number in the row reports the average absolute change in the t-statistic when changing the indicated decision. The results in columns 3 through 9 report results for the randomly generated normally distributed explanatory variables. As an example of interpreting the numbers in the table, in panel A, the number in row 2 of column 3 indicates that in profitability regressions, on average, across 1,000 randomly generated normally distributed explanatory variables, the t-statistic on the explanatory variable changes by 1.11 in absolute value when switching from winsorizing outliers (the most common method) to retaining outliers. Column 9 reports averages across all six categories of regressions (6,000 explanatory variables total). The decisions are listed in the table in order from most impactful to least impactful, based on column 9.

Row 1 of panel A of Table 6 shows that the disruptive benchmark performs as expected—when switching to a completely new explanatory variable, the average absolute change in the t-statistic is close to 1.13 in all categories of regressions. Row 2 shows that the methodological decision with the most impact on t-statistics is the decision of whether to winsorize outliers. The impact of outlier treatment is almost as large as the disruptive benchmark for all categories of regressions except for cash regressions. Rows 3 and 5 show

Table 6
Changes in t-statistics when changing methods: Randomly generated explanatory variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Industry fixed effects												
			Normal							Log	Dummy	DD
	Most common method	Alternative method	Profit.	Value	Lev.	Invest.	Payout	Cash	ALL	ALL	ALL	ALL
(1)	*Disruptive benchmark: Randomly generate a new explanatory variable*		*1.17*	*1.17*	*1.13*	*1.14*	*1.09*	*1.16*	*1.14*	*1.12*	*1.13*	*1.13*
(2)	Winsorize outliers at 1st/99th	Retain outliers	1.11	1.10	1.09	1.01	1.10	0.17	0.93	0.96	0.99	0.88
(3)	Most common dependent variable	Next most common with different denominator	0.93	0.80	0.67	0.80	0.60	0.60	0.73	0.75	0.76	0.70
(4)	Contemporaneous explanatory variable	Lagged explanatory variable	0.69	0.82	0.66	0.81	0.64	0.64	0.71	0.71	0.70	0.13
(5)	Most common dependent variable	Next most common with different numerator	0.37	0.09	0.71	1.14	1.12	0.58	0.67	0.65	0.67	0.59
(6)	Level dependent variable	Logged dependent variable		0.64	0.76			0.59	0.66	0.68	0.66	0.59
(7)	Winsorize outliers at 1st/99th	Trim outliers at 1st/99th	0.71	0.73	0.62	0.49	0.61	0.45	0.60	0.63	0.56	0.45
(8)	Continuous explanatory variable	Convert the explanatory variable to a dummy	0.51	0.48	0.50	0.50	0.48	0.49	0.49	0.59		
(9)	Winsorize outliers at 1st/99th	Winsorize outliers at 5th/95th	0.61	0.75	0.50	0.31	0.45	0.24	0.47	0.50	0.47	0.40
(10)	End-year denominator on dependent variable	Begin-year denominator on dependent variable	0.69			0.51	0.22		0.47	0.47	0.46	0.24
(11)	Most common size control	Second most common size control	0.36	0.47	0.39	0.32	0.08	0.34	0.33	0.33	0.33	0.29
(12)	Include all industries	Exclude financial firms	0.24	0.24	0.23	0.18	0.38	0.29	0.26	0.27	0.27	0.30
(13)	All control variables with majority usage	Add next most common control variable	0.34	0.01	0.14	0.05	0.08	0.23	0.14	0.15	0.14	0.13
(14)	2-digit SIC dummies	Fama-French dummies	0.10	0.06	0.11	0.13	0.14	0.22	0.13	0.12	0.13	0.17
(15)	Level explanatory variable	Logged explanatory variable	0.09	0.08	0.08	0.09	0.08	0.09	0.09	0.24		
(16)	*Innocuous benchmark: Round the explanatory variable to the second decimal place*		*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*

(*Continued*)

Table 6
Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
B. Firm fixed effects												
			Normal							Log	Dummy	DD
	Most common method	Alternative method	Profit.	Value	Lev.	Invest.	Payout	Cash	ALL	ALL	ALL	ALL
(1)	Winsorize outliers at 1st/99th	Retain outliers	1.01	1.11	1.07	1.02	1.05	0.23	0.91	0.93	0.91	0.83
(2)	Contemporaneous explanatory variable	Lagged explanatory variable	0.88	0.92	0.80	0.91	0.82	0.82	0.86	1.05	0.86	0.21
(3)	Level dependent variable	Logged dependent variable		0.80	0.90			0.83	0.84	0.85	0.85	0.94
(4)	Most common dependent variable	Next most common with different denominator	0.92	0.85	0.76	0.66	0.61	0.67	0.74	0.74	0.73	0.68
(5)	Most common dependent variable	Next most common with different numerator	0.46	0.09	0.72	1.03	1.06	0.62	0.66	0.66	0.66	0.59
(6)	Winsorize outliers at 1st/99th	Trim outliers at 1st/99th	0.67	0.71	0.61	0.48	0.58	0.44	0.58	0.61	0.54	0.36
(7)	End-year denominator on dependent variable	Begin-year denominator on dependent variable	0.83			0.53	0.27		0.54	1.10	0.54	0.28
(8)	Winsorize outliers at 1st/99th	Winsorize outliers at 5th/95th	0.73	0.77	0.53	0.32	0.44	0.20	0.50	0.50	0.48	0.38
(9)	Continuous explanatory variable	Convert the explanatory variable to a dummy	0.49	0.47	0.47	0.46	0.47	0.48	0.47	1.00		
(10)	Most common size control	Second most common size control	0.38	0.50	0.41	0.32	0.08	0.38	0.34	0.35	0.35	0.29
(11)	Include all industries	Exclude financial firms	0.22	0.20	0.22	0.18	0.30	0.27	0.23	0.23	0.23	0.29
(12)	All control variables with majority usage	Add next most common control variable	0.26	0.00	0.12	0.05	0.06	0.15	0.11	0.11	0.11	0.11
(13)	Level explanatory variable	Logged explanatory variable	0.08	0.08	0.08	0.08	0.09	0.08	0.08	0.23		

The table reports the average absolute change in the t-statistic for randomly generated explanatory variables when changing methodological decisions. Results are presented for randomly generated normally distributed explanatory variables, with summary results reported for three other types of randomly generated explanatory variables: lognormal, dummy, and difference-in-differences (DD). For each type of explanatory variable and each category of regression, 1,000 randomly generated variables are tested. The reported number is the average absolute change in the t-statistic across the 1,000 explanatory variables when the methodology switches between the most common method and the alternative method indicated, holding all other decisions at the most common methodology. Columns 9 through 12 average results across 6,000 explanatory variables (1,000 for each category). Methodological decisions are ordered from those that induce the largest changes to the smallest changes, based on column 9. For comparative purposes, results are also reported for a "disruptive benchmark" (randomly generating an entirely new explanatory variable), and an "innocuous benchmark" (rounding the explanatory variable to the second decimal place). All data other than the randomly generated explanatory variables come from the Compustat database for the years 1963 to 2018. Each regression includes year fixed effects and either industry fixed effects (panel A) or firm fixed effects (panel B). Standard errors are clustered at the firm level.

that dependent variable choice also has a large impact on t-statistics, inducing a change of 0.73 on average when switching to the next most common dependent variable with a different denominator, and 0.67 when switching to the next most common dependent variable with a different numerator (across all categories of regressions). Other decisions having a large impact include lagging the explanatory variable (0.71 on average), logging the dependent variable (0.66), and trimming outliers (0.60). The next three decisions listed have a moderately high impact on t-statistics, including converting the explanatory variable to a dummy (0.49), changing outlier cutoffs (0.47), and lagging the denominator of flow/stock dependent variables (0.47). Clearly, given a typical threshold for significance of $|t| > 1.96$, many of these decisions have a substantive impact on whether coefficients are considered significant or not. The other decisions—switching the size control, excluding financial firms, adding a control variable, changing industry definitions, logging the explanatory variable—generally do not have as great of an impact on t-statistics, although there are occasionally specifications for which the impact is larger. Finally, row 16 of panel A shows that the innocuous benchmark performs as expected, with an average change of 0.00 when rounding the explanatory variable.

In columns 10 through 12 of Table 6, I report additional tests using different types of explanatory variables. For brevity, I report results only for all categories of regressions combined. Column 10 reports results for lognormally distributed variables, which I create by randomly selecting a mean between 0 and 1 and a standard deviation between 0 and 1. Column 11 reports results for dummy variables, which I create as an indicator that equals one if the observation is above a randomly selected cutoff on a uniform random variable on the interval (0,1) and zero otherwise. Column 12 reports results for difference-in-differences (DD) variables, which I create with a procedure following Bertrand, Duflo, and Mullainathan (2004).[17] With a few exceptions, the results for the other types of variables are similar to those for the normally distributed variables. Similarly, the results for firm fixed effects (panel B) are not markedly different from those using industry fixed effects.[18]

2.2.2 Quasi-random explanatory variables. As an alternative to purely random explanatory variables, I create explanatory variables by randomly combining data items from Compustat. The procedure for creating these explanatory variables follows from Yan and Zheng (2017) and Chordia, Goyal, and Saretto (2020), who use random Compustat items to test hypothetical trading strategies. I create ratio explanatory variables by randomly selecting a numerator from one of 173 data items from Chordia, Goyal, and Saretto (2020)

[17] Brodeur, Cook, and Heyes (2020b) document the extent of p-hacking and publication bias in articles using causal identification techniques such as DD.

[18] Cells left blank are those for which a particular specification does not apply.

and then randomly selecting a denominator from one of 15 scaling variables from Yan and Zheng (2017). For each category of regression, I do not allow numerators that have obvious correlations with the dependent variable. For example, I do not allow measures of dividends (Compustat codes dv, dvc, dvp, $dvpa$, dvt) in payout regressions, and I do not allow measures of cash (ch, che, $chech$) in cash regressions. I create 1,000 such explanatory variables for each category of regression, and I repeat the procedures for testing the impact on t-statistics for each alternative method.

In contrast to the purely random explanatory variables, the quasi-random variables are created from actual firm-level data, so we would expect frequent correlations between explanatory variables and dependent variables, even though the Compustat data items are selected randomly. Because of these underlying correlations, the distribution of t-statistics is no longer the standard normal, and the expected change in t-statistics is no longer bounded by 1.13. Panel B of Appendix Table A3 reports summary statistics for the t-statistics from all regressions performed in these tests. The statistics confirm that the mean t-statistic varies from zero and that the standard deviation of t-statistics is much higher than one. Additionally, panel B shows that, across all regressions performed, over 70% of the coefficients on the quasi-random variables are statistically significant at the 10% level.

The results of these tests are reported in Table 7. For brevity, I report results only for firm fixed effects, given that Table 6 showed results are similar with either firm or industry fixed effects. The decisions are again listed from most impactful to least impactful. Table 7 shows that the average change in t-statistics for the various methodological changes is much higher when using Compustat explanatory variables. For example, row 1 shows that the decision to winsorize or retain outliers is the most impactful (as in Table 6), but the average absolute change in the t-statistic is 9.20, almost 10 times the size of the effect for purely random explanatory variables. As a point of comparison, the average absolute changes reported in row 1 are about the same magnitude as the average (absolute value) t-statistics across all regressions in the simulation (see Appendix Table A3). Across all types of regressions, with the exception of the decision on excluding financial firms (row 13), the impact of every decision is at least three times the size of the effect for purely random variables, and often much greater. Table 7 illustrates that the impact on t-statistics from changes in methods can be much larger when the explanatory and dependent variables have underlying correlations, as opposed to when they are constructed to be independent, as in Table 6.

2.2.3 Actual hypothesized explanatory variables. I also test the effect of methodological changes on t-statistics for coefficients on actual variables from the existing literature. In this analysis I focus on leverage regressions, compiling a set of 65 variables that have been proposed as determinants of leverage in other articles. These determinants are taken from a large set of studies, and

Table 7
Changes in t-statistics when changing methods: Random combinations of Compustat items

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Most common method	Alternative method	Profit.	Value	Lev.	Invest.	Payout	Cash	ALL
(1)	Winsorize outliers at 1st/99th	Retain outliers	12.86	8.15	10.86	6.30	5.29	11.75	9.20
(2)	Continuous explanatory variable	Convert the explanatory variable to a dummy	7.81	10.39	9.93	7.31	5.41	11.32	8.70
(3)	Level dependent variable	Logged dependent variable		7.90	8.64			7.49	8.01
(4)	Most common dependent variable	Next most common with different denominator	12.31	8.17	9.19	7.78	4.49	5.59	7.92
(5)	Winsorize outliers at 1st/99th	Winsorize outliers at 5th/95th	9.82	6.56	10.51	3.76	3.23	6.36	6.71
(6)	Contemporaneous explanatory variable	Lagged explanatory variable	6.46	4.45	3.37	6.14	1.72	6.09	4.71
(8)	End-year denominator on dependent variable	Begin-year denominator on dependent variable	6.73			4.83	1.21		4.26
(9)	Most common dependent variable	Next most common with different numerator	3.35	0.31	6.33	6.00	3.84	3.79	3.94
(10)	Winsorize outliers at 1st/99th	Trim outliers at 1st/99th	3.37	3.31	3.28	1.78	1.60	2.40	2.62
(11)	Most common size control	Second most common size control	4.90	2.50	1.18	0.94	1.94	2.93	2.40
(12)	All control variables with majority usage	Add next most common control variable	5.56	0.02	0.44	0.53	1.24	2.79	1.76
(13)	Include all industries	Exclude financial firms	0.57	0.47	0.56	0.41	0.58	0.53	0.52

The table reports the average absolute change in the t-statistic for explanatory variables when changing methodological decisions. The explanatory variables are randomly created from Compustat data items. For each category of regression, 1,000 randomly generated variables are tested. The reported number in each cell is the average absolute change in the t-statistic across all explanatory variables when the methodology switches between the most common method and the alternative method indicated, holding all other decisions at the most common methodology. Column 9 averages results across 6,000 explanatory variables (1,000 for each category). Methodological decisions are ordered from those that induce the largest changes to the smallest changes, based on column 9. All data come from the Compustat database for the years 1963 to 2018. Each regression includes year fixed effects and firm fixed effects. Standard errors are clustered at the firm level.

for brevity I do not discuss every determinant. A complete list of determinants with associated references is available in Table 10, and most of the variables are discussed in more detail in Titman and Wessels (1988), Frank and Goyal (2009), and Fukui, Mitton, and Schonlau (2020).[19] Of the 65 variables, 49 are continuous variables and 16 are dummy variables. I test the variables in the same manner as the tests in Tables 6 and 7, with firm fixed effects. Panel C of Appendix Table A3 reports summary statistics of the t-statistics generated from these tests.

The results of these tests are reported in Table 8. Column 3 reports average absolute changes in t-statistics for continuous variables, while results for dummy variables are reported in column 4, and combined results are reported in column 5. Row 1 shows that the most impactful methodological decision in these tests is the decision to use the most common dependent variable (total debt/total assets) or the next most common variable with a different denominator (total debt/market value of assets). Across all 65 leverage determinants tested, the decision of whether to use book leverage or market leverage changes the t-statistic on the coefficient for the determinant by 3.91, on average. The magnitude of this average change is comparable to the average (absolute value) t-statistics across all regressions in the simulation, which is 4.80 for continuous explanatory variables and 2.64 for dummy explanatory variables (see Appendix Table A3). Rows 2 and 3 show that decisions on winsorizing outliers and converting explanatory variables to dummy variables also have effects well above 3.00. In general, the impact of methodological decisions in these tests is greater than those for purely random explanatory variables (Table 6), but smaller than those for quasi-random variables created from Compustat items (Table 7).

2.3 Cumulative effects of methodological flexibility

I now evaluate the cumulative effect of multiple methodological decisions on the probability that a random hypothesis can be found statistically significant. With no methodological flexibility, a random hypothesis should be found significant about 10% of the time at the 10% level of significance. But if a random hypothesis is tested multiple times with multiple methodologies, then the probability that at least one of those methodologies will produce a significant coefficient rises above 10%, with the probability increasing more when the methodological variations are more disruptive to the original regression. I consider a setting in which researchers are given progressively more discretion over the methods that they choose, and ask whether a given level of methodological discretion would allow a randomly generated

[19] Many researchers graciously shared data that are used in this analysis. I have benefited from data used in many of the papers listed in Table 10 as well as data used in Hirsch and Macpherson (2003), Lovett, Peres, and Shachar (2014), Benmelech and Frydman (2015), Bonica (2016), Cain, McKeon, and Solomon (2017), Volkova (2018), and Harford, Schonlau, and Stanfield (2019).

Table 8
Changes in t-statistics when changing methods: Proposed determinants of leverage

	(1)	(2)	(3)	(4)	(5)
	Most common method	Alternative method	Continuous variables	Dummy variables	ALL
(1)	Most common dependent variable	Next most common with different denominator	4.68	1.56	3.91
(2)	Winsorize outliers at 1st/99th	Retain outliers	4.30	2.02	3.74
(3)	Continuous explanatory variable	Convert the explanatory variable to a dummy	3.72		3.72
(4)	Winsorize outliers at 1st/99th	Winsorize outliers at 5th/95th	2.24	0.59	1.83
(5)	Most common dependent variable	Next most common with different numerator	2.05	0.92	1.77
(6)	Contemporaneous explanatory variable	Lagged explanatory variable	1.61	0.81	1.41
(7)	Level dependent variable	Logged dependent variable	1.45	1.13	1.37
(8)	Level explanatory variable	Logged explanatory variable	1.10		1.10
(9)	Winsorize outliers at 1st/99th	Trim outliers at 1st/99th	1.20	0.32	0.99
(10)	Most common size control	Second most common size control	0.61	0.86	0.67
(11)	Include all industries	Exclude financial firms	0.28	0.25	0.27
(12)	All control variables with majority usage	Add next most common control variable	0.30	0.17	0.26

The table reports the average absolute change in the t-statistic for explanatory variables when changing methodological decisions. The explanatory variables are proposed determinants of leverage from the literature, with 49 continuous variables tested in column 3 and 16 dummy variables tested in column 4. For a list of the variables tested, see Table 10. The reported number in each cell is the average absolute change in the t-statistic across all explanatory variables when the methodology switches between the most common method and the alternative method indicated, holding all other decisions at the most common methodology. The explanatory variables come from a variety of sources, and all other data come from the Compustat database for the years 1963 to 2018. Each regression includes year fixed effects and firm fixed effects. Standard errors are clustered at the firm level.

hypothesis to be found statistically significant with at least one methodological combination.[20] I randomly generate normally distributed explanatory variables and perform regressions as in Equation 1 with firm fixed effects.[21] I begin with the case of no methodological discretion, in which only the most common methodology is used. Then I cumulatively allow discretion over additional binary decisions up to a maximum of 10 decisions. I repeat this procedure for 1,000 randomly generated variables for each category of regression, and report the percentage of the 1,000 variables that can be found statistically significant with at least one methodological combination under the different levels of methodological flexibility.

[20] See Simmons, Nelson, and Simonsohn (2011) for a similar experiment from the psychology literature.

[21] A similar analysis could be conducted with industry fixed effects or using different types of randomly generated explanatory variables. However, given that Table 6 reports similar impacts from all of these specifications, I limit my analysis in Table 9 to normally distributed explanatory variables and firm fixed effects for brevity. Also, I do not repeat this analysis with quasi-random Compustat variables or actual leverage determinants because most of those variables (84% of the quasi-random variables and 51% of the leverage determinants) are statistically significant using the most common methodology.

Table 9 reports the results of these simulations, with each panel of the table corresponding to a different category of regression, as noted. The percentages in the table indicate the percentage of randomly generated hypotheses for which statistical significance can be found with at least one combination of allowed methods. For example, in the profitability regressions in panel A, the first column reports the percentage of random hypotheses for which statistical significance can be found when no alternative methods are allowed; that is, when only the most common methodology is used. Panel A shows that when only one method is used, random hypotheses are found to have a significant relation with profitability 8% of the time at the 10% level, 4% of the time at the 5% level, and 1% of the time at the 1% level. These percentages are reasonably close to the percentages that would be expected due to random chance, and follow a similar pattern in all six panels of the table.

In the next column to the right in Table 9, one alternative method is allowed—researchers can use the second most common dependent variable instead of the most common dependent variable. In profitability regressions, this allows 15% of random hypotheses to be found significant at the 10% level, 7% at the 5% level, and 1% at the 1% level. In the next column to the right, two alternative methods are allowed—researchers can use the second most common dependent variable, and they can add the next most common control variable. (Note that the allowed methods are cumulative moving from left to right in the table.) In profitability regressions, giving discretion over two decisions allows 19% of random hypotheses to be found significant at the 10% level, with 8% at the 5% level, and 2% at the 1% level.

The columns in Table 9 proceed in this fashion until the final column in the table in which researchers have discretion over 10 binary methodological decisions. In this case, for each randomly generated hypothesis, I search for a combination of all 10 binary decisions that allows the hypothesis to be found statistically significant. In the case of profitability regressions, this amount of flexibility allows 100% of random hypotheses to be found significant at the 10% level, with 90% at the 5% level, and 36% at the 1% level. Results are similar for the other categories of regressions. In all six panels, at least 90% of hypotheses can be found significant at the 10% level, with significance at the 5% level ranging from 60% (for payouts) to 90% (for profitability), and significance at the 1% level ranging from 17% (for value) to 36% (for profitability).

It is important to recognize that the coefficients from the individual regressions performed in these simulations are still not significant any more than 10% of the time at the 10% level. The summary statistics of the t-statistics from the simulations, reported in panel D of Appendix Table A3, demonstrate this. But since the methodological variation allows for many regressions to be run for each hypothesis (up to 1,024 in the case of 10 decisions), each hypothesis can be significant more than 10% of the time at the 10% level. The extent to which the percentages in Table 9 are greater than 10% is driven by how disruptive the decisions are to the underlying data.

Table 9
Cumulative effect of discretion over methods

Number of alternative methods allowed	0	1	2	3	4	5	6	7	8	9	10
Additional alternative method allowed	None (Most common method only)	Second most common dependent variable	Add next most common control variable	Second most common size control	Exclude financial firms	Log explanatory variable	Convert explanatory variable to dummy	Lag explanatory variable	Retain outliers	Outlier cutoffs 5/95	Dependent variable: Log OR Lag denominator
A. Profitability											
10% level	8%	15%	19%	26%	32%	36%	46%	71%	86%	92%	100%
5% level	4%	7%	8%	12%	16%	18%	25%	46%	60%	70%	90%
1% level	1%	1%	2%	2%	3%	3%	5%	12%	14%	21%	36%
B. Firm value											
10% level	10%	16%	16%	23%	26%	28%	38%	61%	86%	90%	94%
5% level	5%	7%	7%	11%	13%	15%	22%	39%	56%	63%	70%
1% level	1%	1%	1%	2%	3%	3%	4%	8%	10%	13%	17%
C. Leverage											
10% level	9%	16%	18%	25%	30%	33%	44%	65%	83%	86%	91%
5% level	4%	8%	9%	12%	16%	18%	25%	38%	55%	59%	70%
1% level	1%	1%	2%	2%	4%	5%	6%	10%	13%	15%	23%
D. Investment											
10% level	9%	14%	16%	20%	25%	27%	37%	58%	81%	84%	96%
5% level	4%	7%	8%	10%	12%	14%	19%	35%	47%	52%	74%
1% level	1%	2%	2%	3%	3%	3%	5%	8%	10%	11%	20%

(*Continued*)

Table 9
Continued

Number of alternative methods allowed	0	1	2	3	4	5	6	7	8	9	10
Additional alternative method allowed	None (Most common method only)	Second most common dependent variable	Add next most common control variable	Second most common size control	Exclude financial firms	Log explanatory variable	Convert explanatory variable to dummy	Lag explanatory variable	Retain outliers	Outlier cutoffs 5/95	Dependent variable: Log OR Lag denominator
E. Payouts											
10% level	10%	16%	16%	17%	22%	24%	34%	56%	81%	85%	91%
5% level	5%	8%	9%	9%	13%	14%	18%	33%	45%	52%	60%
1% level	1%	2%	2%	2%	4%	4%	5%	10%	10%	14%	18%
F. Cash holdings											
10% level	9%	14%	16%	22%	30%	33%	44%	67%	86%	89%	93%
5% level	4%	7%	8%	11%	15%	17%	25%	43%	56%	61%	70%
1% level	1%	2%	2%	3%	4%	4%	6%	11%	15%	17%	24%
G. ALL											
10% level	9%	15%	17%	22%	27%	30%	41%	63%	84%	88%	94%
5% level	4%	7%	8%	11%	14%	16%	22%	39%	53%	60%	73%
1% level	1%	2%	2%	2%	3%	4%	5%	10%	12%	15%	23%

The table reports the percentage of randomly generated explanatory variables that can be found to have statistically significant effects on dependent variables with at least one methodological combination, given discretion over binary methodological decisions. In each successive column an additional alternative method is allowed (as noted in the column heading), so that each column reports results given cumulative discretion over the number of binary decisions noted in the column heading. Results are presented for randomly generated normally distributed variables. For each category of regression (noted in the panel headings), 1,000 randomly generated variables are tested. All data other than the randomly generated explanatory variables come from the Compustat database for the years 1963 to 2018. Each regression includes year fixed effects and firm fixed effects. Standard errors are clustered at the firm level.

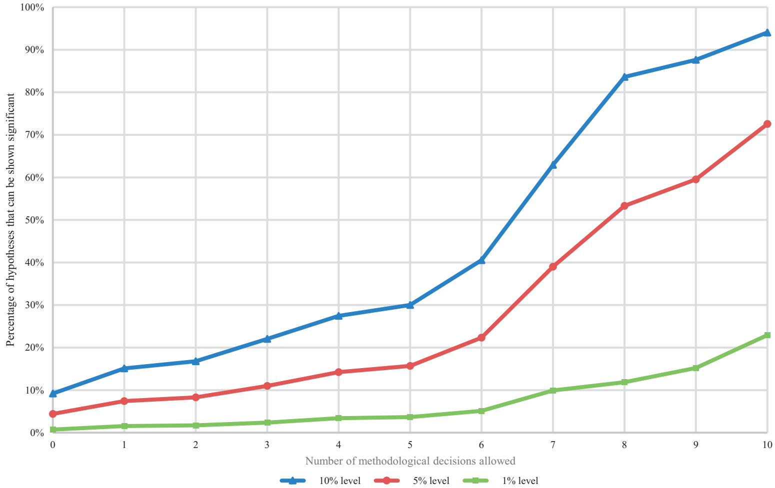

Figure 3
Random hypotheses that are significant with at least one methodology
The percentage of randomly generated hypotheses that can be shown to be statistically significant determinants of corporate finance outcomes with at least one methodological combination, given discretion over binary methodological decisions. Based on Compustat data from 1963 to 2018. The figure combines results for all categories of regressions (profitability, value, leverage, investment, payout, cash) as reported in Table 9. For explanations of the decisions allowed at each step, see Table 9.

Figure 3 displays the results of Table 9 graphically, with the results for all six categories of regressions combined. The left end of the figure shows that with no methodological discretion, roughly 10% of random hypotheses are significant at the 10% level, with about 5% at the 5% level, and about 1% at the 1% level. These percentages increase from left to right as greater methodological discretion is allowed, with steeper increases corresponding to points where more-impactful decisions are allowed.

As mentioned earlier, the results in Table 9 demonstrate that reporting statistical significance can be enabled by methodological variation, but the percentages do not represent probabilities that a randomly selected hypothesis could be credibly supported. Robustness testing, multiple approaches to testing, requests for additional tests by referees and editors, and the possibility of replication by other researchers all help ensure that spurious findings are not reported. In the next section, I discuss robustness checks and other remedies that can mitigate the negative effects of methodological variation.

3. Proposed Remedies

3.1 Robustness checks

Robustness checks are commonly used to ensure that reported results are not dependent upon a particular methodology. When a given hypothesis is required to be statistically significant across multiple specifications, it reduces the probability of a hypothesis being presented as a statistically significant result. Figure 4 shows to what degree robustness checks reduce the probability of finding statistically significant hypotheses. I use data from the same simulations as in Table 9, with Compustat data from 1963 to 2018, with 10 methodological decisions allowed, and with 1,000 randomly generated explanatory variables for each category of regression (with 6,000 total represented in the figure). The first set of points on the left of Figure 4 shows what percentage of random hypotheses can be shown statistically significant with at least one methodological combination when researchers have discretion over all 10 methodological decisions. As the points move from left to right, researchers still have discretion over the 10 decisions, but they are also required to show that the results are robust to changes in increasing numbers of the 10 decisions.

Three key points are illustrated in Figure 4. First, when a modest number of robustness checks is required, it remains relatively easy to find a combination of methodologies that allows for reporting significant results (the assumption in the figure is that researchers choose which robustness checks to report). Second, as the number of required robustness checks increases, the probability of finding significant results decreases at an increasing rate. Third, the number of hypotheses that can be shown significant eventually trends toward zero (below the baseline of 10%/5%/1%) as increasing numbers of robustness checks are required. The figure stops at 10 robustness checks—at which point random hypotheses can be shown significant 3% of the time at the 10% level, 1% of the time at the 5% level, and 0% of the time at the 1% level—but increased numbers of robustness checks would further reduce the probability of finding significant hypotheses.

Figure 4 illustrates one drawback of robustness checks: while they clearly reduce the probability of false positive findings, if required excessively and indiscriminately they can also lead to false negative findings.[22] Only rarely can a hypothesis survive every reasonable robustness check. Indeed, Harvey (2019) warns of "reverse p-hacking," or the potential to find specifications that contradict any hypothesis if one tries hard enough. To illustrate typical levels of robustness for variables in the corporate finance literature, I return to the set of 65 determinants of leverage considered in Table 8. I test each of the determinants in firm fixed effects and industry fixed effects specifications, using all combinations of methods as listed in Table 9, excluding the two methods that do not apply with dummy explanatory variables. I record the t-statistics for each proposed determinant across all 512 possible specifications, and report the percentage of specifications for which the coefficient is of the expected sign and significant at the 10% level or higher. In these tests I am not replicating the individual methodologies used in each of the papers that propose these determinants of leverage; I am imposing the same combinations of methods

[22] Harvey and Liu (2020) note that the relative costs of Type I and Type II error depend upon the issue being studied, and they propose a method for measuring and making efficient trade-offs of Type I and Type II error.

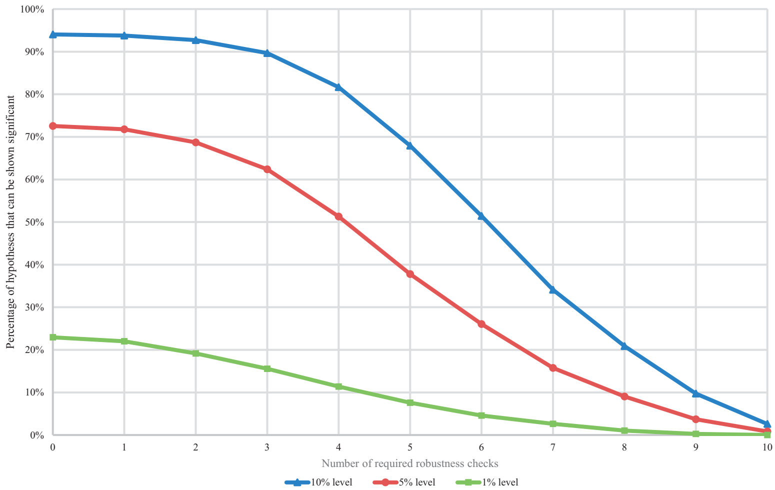

Figure 4
Random hypotheses that survive robustness checks
The percentage of randomly generated hypotheses that can be shown significant with at least one combination of 10 binary methodological decisions when increasing numbers of robustness checks are required. Based on Compustat data from 1963 to 2018. The figure combines results for all categories of regressions (profitability, value, leverage, investment, payout, cash).

on all 65 variables. The intent of these tests is not to support or discredit any particular finding, because I am painting a broad picture of overall robustness, ignoring nuances in theories or identification that could be important when testing an individual variable.

Table 10 lists the 65 proposed determinants of leverage, a reference for each variable, the expected sign on each variable, and the percentage of specifications that are robust for each variable. In some cases, the variable tested is exactly the same as the one tested in the referenced paper, and in other cases it is an updated version of the variable or a similar variable from a different dataset. Table 10 shows that only one proposed determinant is robust across all specifications—a dummy variable indicating whether a firm has a bond rating. Only six other determinants are robust in 90% or more of the specifications. The average percentage of robust specifications across all 65 determinants is 43%, so the typical proposed determinant has hundreds (or thousands, if more methodological combinations were considered) of robust specifications even while it has hundreds (or thousands) of non-robust specifications. Table 10 suggests that completely robust explanatory variables are few and far between, and that researchers should spend more time evaluating why variables are robust in some situations and not in others, and less time defending the robustness of a variable.

A second drawback of robustness checks is that, despite the fact that they have been used regularly for decades, they have not resolved the problem of

Table 10
Robustness of proposed determinants of leverage

Proposed leverage determinant	Reference	Expected sign	Specifications robust at 10% level	Proposed leverage determinant	Reference	Expected sign	Specifications robust at 10% level
Bond rating	Faulkender and Petersen (2006)	+	100%	Corruption	Smith (2016)	+	44%
Stock illiquidity	Fang, Now, and Tice (2009)	+	99%	Unem. insurance generosity	Agrawal and Matsa (2013)	+	44%
Trade secret protection	Klasa et al. (2018)	+	98%	CEO cash/stock compensation	Carlson and Lazrak (2010)	+	43%
Analyst coverage	Derrien and Kecskés (2013)	–	97%	CEO tenure	Berger, Ofek, and Yermack (1997)	–	41%
Ratings conservatism	Baghai, Servaes, and Tamayo (2014)	–	95%	Outside directors	Berger, Ofek, and Yermack (1997)	+	39%
Dedicated customer relationships	Banerjee, Dasgupta, and Kim (2008)	–	95%	CDS trading	Saretto and Tookes (2013)	+	38%
Share retainer	Sen and Tumarkin (2015)	+	94%	Dedicated supplier relationships	Banerjee, Dasgupta, and Kim (2008)	–	36%
Product diversification	Shleifer and Vishny (1992)	+	88%	Brand perception	Larkin (2013)	+	26%
Dividends	Frank and Goyal (2009)	–	84%	Union coverage	Matsa (2010)	+	25%
State tax rate increases	Heider and Ljunqvist (2015)	+	74%	Pilot CEO	Cain and McKeon (2016)	+	24%
Historical M/B ratio	Baker and Wurgler (2002)	–	71%	Pension liabilities	Shivdasani and Stefanescu (2010)	–	18%
Covenant strength	Denis and Wang (2014)	–	69%	Tax shelters	Graham and Tucker (2006)	–	17%
Advertising expenses	Frank and Goyal (2009)	–	66%	Sin stock	Hong and Kacperczyk (2009)	+	15%
Analyst disagreement	Dittmar and Thakor (2007)	+	66%	Options traded	Gao (2010)	+	14%
CEO delta	Chava and Purnanandam (2010)	–	65%	Male CEO	Huang and Kisgen (2013)	+	11%
Product market fluidity	Hoberg, Phillips, and Prabhala (2014)	–	64%	Supply uncertainty	Massa, Yasuda, and Zhang (2013)	–	9%
Location in industry cluster	Almazan et al. (2010)	–	63%	CDS-referenced customers	Li and Tang (2016)	–	8%
Industry concentration	Hoberg and Phillips (2016)	+	62%	CEO age	Bertrand and Schoar (2003)	–	7%
Firm age	Frank and Goyal (2009)	+	62%	Supplier R&D intensity	Kale and Shahrur (2007)	–	7%
Product similarity	Hoberg and Phillips (2016)	–	61%	Takeover defenses	Bebchuk, Cohen, and Ferrell (2009)	–	7%
Import penetration	Xu (2012)	–	60%	Customer R&D intensity	Kale and Shahrur (2007)	–	5%
CEO inside debt holdings	Cassell et al. (2012)	–	59%	Board size	Berger, Ofek, and Yermack (1997)	–	4%
Investment tax credits	Titman and Wessels (1988)	–	59%	Marginal tax rate	Graham, Lang, and Shackelford (2004)	+	3%
R&D expenses	Titman and Wessels (1988)	–	59%	Depreciation tax shields	DeAngelo and Masulis (1980)	–	2%
Regulated firm	Frank and Goyal (2009)	+	59%	Earnings volatility	Titman and Wessels (1988)	–	2%
Takeover susceptibility	Garvey and Hanka (1999)	+	58%	Military CEO	Malmendier, Tate, and Yan (2011)	+	1%
Selling expenses	Titman and Wessels (1988)	–	58%	Tournament incentives	Kini and Williams (2012)	+	0%
Growth	Titman and Wessels (1988)	–	56%	CEO vega	Brockman, Martin, and Unlu (2010)	+	0%
Employee treatment index	Bae, Kang, and Wang (2011)	–	52%	Large blockholder	Berger, Ofek, and Yermack (1997)	+	0%
Import tariffs	Frésard and Valta (2016)	+	51%	Board co-option	Coles, Daniel, and Naveen (2014)	–	0%
Lobbying expenditures	Kostovetsky (2015)	+	49%	Campaign contributions	Claessens, Feijen, and Laeven (2008)	+	0%
Equipment manufacturers	Titman and Wessels (1988)	–	44%	Tax loss carryforwards	Frank and Goyal (2009)	–	0%
Labor protection laws	Serfling (2016)	–	44%				

The table presents the percentage of specifications that are robust when testing variables proposed as determinants of leverage in the literature. The specifications tested are all combinations of the methods listed in Table 9, except for logging the explanatory variable and converting the explanatory variable to a dummy (because 16 of the proposed determinants are dummy variables). Firm fixed effects and industry fixed effects specifications are both tested. For each proposed determinant, the expected sign on the coefficient for that variable is reported, along with the percentage of specifications for which the coefficient is of the expected sign and significant at the 10% level. The explanatory variables come from a variety of sources; the dependent variables and control variables come from Compustat. Standard errors are clustered at the firm level.

selective reporting of statistically significant results. Christensen and Miguel (2018) note,

> The greater use of extra robustness checks in applied economics is designed to limit the extent of specification search, … but it is unclear how effective these changes are in reducing bias in practice. … [T]he analysis of 641 articles from three top economics journals in recent years presented in Brodeur et al. (2016) still shows a disturbing two-humped distribution of p-values, with relatively few p-values between 0.10 and 0.25 and far more just below 0.05.

Kim and Ji (2015) and Morey and Yadav (2018) report similar findings for finance journals.

A third concern with robustness checks is that they are not very systematic. A researcher cannot test every possible methodological combination, so the set of robustness checks that the researcher reports could, either intentionally or unintentionally, present a skewed picture of the robustness of a variable. The review process helps ensure that researchers do not miss performing important robustness checks. Too often, however, articles discuss only the robustness checks that confirm the reported result.

3.2 Specification checks

Another type of approach that has been proposed to address methodological variation is to report the entire range of possible outcomes resulting from different combinations of methods.[23] Brodeur, Cook, and Heyes (2020b) refer to this type of analysis as a "specification check," and their procedure consists of performing regressions with all possible combinations of methods, reporting in graphical form the distribution of results obtained.

As an example of how these tests can be applied to a corporate finance setting, I perform specification checks for some of the determinants of capital structure listed in Table 10. I select three examples that illustrate different patterns of robustness that can be revealed by specification checks. For each of the three proposed determinants, I plot histograms of the t-statistics from all of the specifications tested in Table 10. These histograms are reported in Figure 5. The first variable tested is stock illiquidity, which is hypothesized to be positively associated with leverage, as in Fang, Now, and Tice (2009). Panel A shows that stock illiquidity is a highly robust determinant of leverage. The t-statistics from all specifications are positive, and only a small number (about 1%) are below the cutoff of 1.96. The remainder of the specifications have t-statistics ranging from 1.96 to over 40. The next variable tested is analyst disagreement, which is hypothesized

[23] See Leamer (1983), Leamer and Leonard (1983), Yan and Zheng (2017), Simonsohn, Simmons, and Nelson (2020), and Brodeur, Cook, and Heyes (2020a).

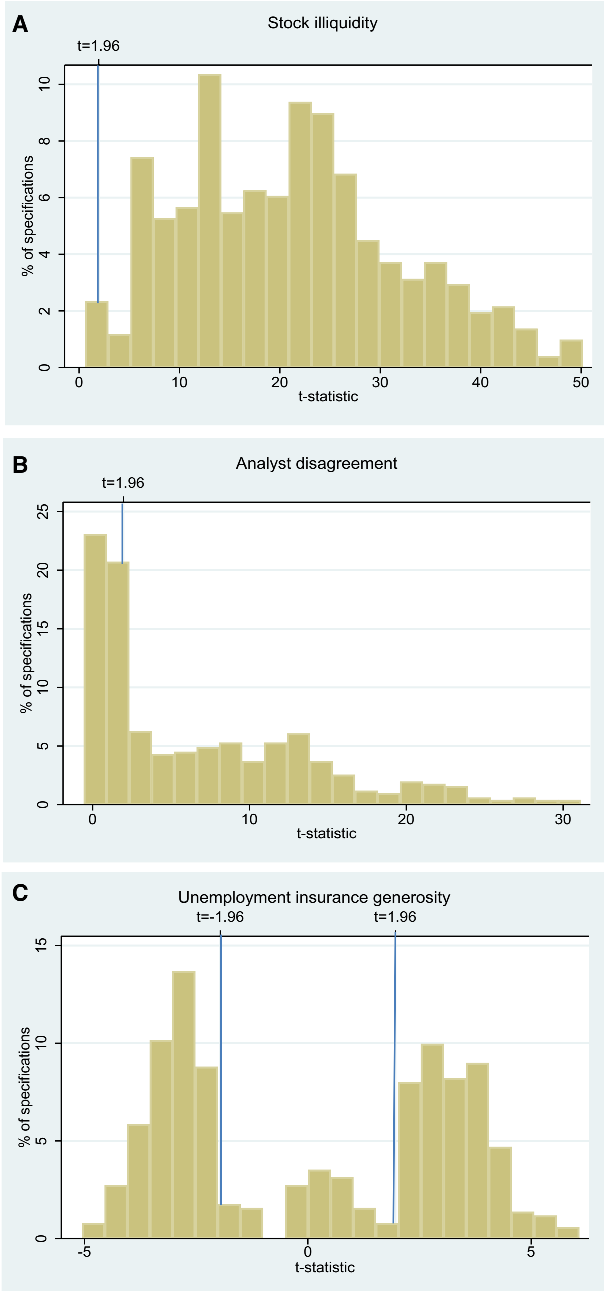

Figure 5
The distribution of t-statistics for proposed determinants of leverage
For each determinant, 512 methodological combinations are tested. In all three panels the explanatory variable is hypothesized to be a positive determinant of leverage. Vertical lines denote cutoffs for the 5% level of statistical significance.

to be positively associated with leverage, as in Dittmar and Thakor (2007). Panel B shows that analyst disagreement is a somewhat robust determinant of leverage. Some of the specifications result in negative t-statistics, and a fairly large percentage of t-statistics (over 30%) fall below the cutoff of 1.96. Still, hundreds of specifications have t-statistics ranging from 1.96 to over 30. The next variable tested is unemployment insurance generosity, which is hypothesized to be positively associated with leverage, as in Agrawal and Matsa (2013). Panel C shows that tests of unemployment insurance generosity result in a bimodal distribution of t-statistics. Most of the specifications result in t-statistics that are significant at the 5% level, but the percentage of specifications with positive significance is about the same as the percentage of specifications with negative significance. With a variable like this, a researcher could arrive at opposite conclusions about the direction of an effect depending on the methodological decisions made during the analysis.

An advantage of specification checks is that they graphically show the impact on statistical significance from changing many different methods at the same time, whereas robustness checks typically report results from changing one method at a time. Specification checks can be expanded to test as many different methodological combinations as desired. A drawback of specification checks is that the set of decisions to test is still determined at the discretion of the researcher, who could include only the methods that support a hypothesis, either intentionally or unintentionally. Nevertheless, in comparison to robustness checks, specification checks are a more systematic and comprehensive way to evaluate the validity of a result.

3.3 Economic significance

One important way for researchers to deal with the negative effects of methodological variation is to place less emphasis on statistical significance and greater emphasis on economic significance (De Long and Lang 1992; McCloskey and Ziliak 1996). Although it is well known that statistical significance does not measure practical importance, discussions of economic significance often take a back seat to discussions of statistical significance, perhaps because statistical significance is easier to measure, or because standardized thresholds for statistical significance make it easier to demonstrate significant results. Yet economic significance is ultimately the more relevant measure; we want to know not only whether an effect is statistically detectable, but how much of an impact it has on the real world. Additionally, Mitton (2021) shows that measures of economic significance are more immune to methodological variation than measures of statistical significance. So focusing on economic significance not only emphasizes the more relevant measure, but also weakens incentives for publication bias and p-hacking.

3.4 Other remedies

A few other recommendations for mitigating the negative effects of methodological variation have been suggested in the literature (see especially

Harvey (2017) and Ioannidis (2008)). First, researchers should transparently report all tests conducted over the course of their study, not just those that support the hypothesis being tested. Second, researchers should outline a research framework (including decisions on methodology) before looking at the data. Third, researchers should make public the data (when possible) and code used in generating their results (see Harvey (2019)).

A final recommendation is for researchers to make an effort to adjust for multiple testing. A test with multiple acceptable methodologies should ideally be treated as a multiple test, with statistical corrections like those required when dealing with multiple hypotheses or multiple comparisons (Gelman and Loken 2014). Harvey, Liu, and Saretto (2020) discuss alternative approaches to adjust for multiple testing in finance research. Unfortunately, although it is clear that thresholds for significance need to be more strict, knowing how to adjust thresholds for multiple methodologies is especially difficult due to ambiguity in how to account for almost endless permutations of methods. Indeed, partly for this reason, Simmons, Nelson, and Simonsohn (2011) argue that such adjustments are impractical.

4. Conclusion

With the abundance of research being produced in empirical corporate finance, it is essential to be able to determine which findings are truly important. The results of this paper show that statistical significance is not sufficient to establish that an empirical result is important. Aside from the fact that statistical significance does not measure practical importance, my tests show that statistically significant coefficients can often be produced by varying the empirical methodology on different dimensions. Thus, although the wide variety of methodologies gives researchers the flexibility to tailor empirical tests to closely match theories being tested, it can also enable p-hacking and publication bias. It is uncertain to what extent p-hacking occurs in the profession, but publication bias is well documented, and that alone is sufficient for methodological flexibility to create problems for statistical inference.

My analysis points to several remedies to mitigate the negative effects of methodological variation. Researchers should employ robustness checks, while striking a balance of applying them thoroughly but not excessively. Researchers should recognize that very few findings are robust on all dimensions, and spend more time understanding why results are fragile on certain dimensions, rather than simply defending the robustness of a result. Researchers should use specification checks as a more comprehensive and systematic way of evaluating the stability of a result. Additionally, researchers should focus more on the economic significance of results instead of statistical significance. By following these recommendations, as well as the others discussed, researchers can help avoid the distorted inferences that can arise from methodological variation.

Appendix

Table A.1
Variable definitions

Variable	Definition
Return on assets (1)	Annual earnings before interest, taxes, depreciation, and amortization (EBITDA) divided by total assets (AT)
Return on assets (2)	Net income (NI) divided by total assets (AT)
Return on equity	Annual net income (NI) divided by total common equity (CEQ)
Tobin's q (1)	Total assets (AT) less total common equity (CEQ) plus market equity, all divided by total assets
Tobin's q (2)	Total assets (AT) less total common equity (CEQ) plus market equity less deferred taxes (TXDB), all divided by total assets
Market-to-book	Market equity divided by total common equity (CEQ)
Book leverage	Long-term debt (DLTT) plus debt in current liabilities (DLC), all divided by total assets (AT)
Long-term debt ratio	Long-term debt (DLTT) divided by total assets (AT)
Market leverage	Long-term debt (DLTT) plus debt in current liabilities (DLC), all divided by the sum of long-term debt, debt in current liabilities, and market equity
Capital expenditures/assets	Total annual capital expenditures (CAPX) divided by total assets (AT)
R&D/assets	Total annual research and development (XRD) divided by end-of-year total assets (AT), times 100
Capital expenditures/capital	Total annual capital expenditures (CAPX) divided by net property, plant, and equipment (PPENT)
Dividends/assets	Annual common/ordinary dividends (DVC) divided by total assets (AT)
Repurchases/assets	Annual repurchases (PRSTKC) divided by total assets (AT)
Dividend yield	Annual common/ordinary dividends (DVC) divided by market equity
Cash/assets (1)	Cash and short-term investments (CHE) divided by total assets (AT)
Cash/assets (2)	Cash (CH) divided by total assets (AT)
Cash/net assets	Cash and short-term investments (CHE) divided by net assets (AT-CHE)
Total assets	Total assets (AT) in $ millions
Sales	Total annual sales (SALE) in $ millions
Market equity	End-of-calendar-year stock price (PRCC_C) times end-of-year shares outstanding (CSHO), in $ millions
Asset tangibility	Net property, plant, and equipment (PPENT) divided by total assets (AT)
Firm age	One plus the current year minus the first year for which Compustat has data for the firm

The table reports definitions of Compustat variables used in the empirical analysis. Compustat mnemonics are indicated in parentheses. Nonpositive values of AT, CSHO, PRCC_C, and SALE are deleted; negative values of CAPX, CH, CHE, DLTT, DLC, DVC, PPENT, PRSTKC, TXDB, and XRD are deleted. Missing values of XRD are replaced with zero. Balance sheet items are end-of-year values in all cases.

Table A.2
Summary statistics

Category	Variable	Mean	Minimum	Median	Maximum	Std. Dev.	N
Profitability	Return on assets (1)	−0.05	−4.85	0.09	0.43	0.64	394,538
	Return on assets (2)	−0.16	−6.37	0.02	0.36	0.80	402,707
	Return on equity	−0.12	−6.11	0.09	1.02	0.86	364,235
Value	Tobin's q (1)	2.58	0.47	1.25	46.32	5.48	348,857
	Tobin's q (2)	2.71	0.45	1.26	51.60	6.09	324,315
	Market-to-book	3.00	0.18	1.57	38.45	5.07	324,481
Leverage	Book leverage	0.30	0.00	0.22	3.01	0.40	399,463
	Long-term debt ratio	0.19	0.00	0.12	1.11	0.22	401,668
	Market leverage	0.28	0.00	0.22	0.95	0.27	348,954
Investment	Capital expenditures/assets	0.06	0.00	0.04	0.45	0.08	377,106
	R&D/assets	0.04	0.00	0.00	0.78	0.11	403,735
	Capital expenditures/capital	0.26	0.00	0.20	1.09	0.23	356,673
Payout	Dividends/assets	0.01	0.00	0.00	0.15	0.02	400,257
	Repurchases/assets	0.01	0.00	0.00	0.19	0.03	341,777
	Dividend yield	0.02	0.00	0.00	0.15	0.03	352,373
Cash	Cash/assets (1)	0.16	0.00	0.07	0.95	0.21	400,665
	Cash/assets (2)	0.11	0.00	0.04	0.90	0.17	355,575
	Cash/net assets	0.51	0.00	0.07	14.07	1.75	399,586
Firm size	log(Total assets)	4.77	−1.92	4.71	11.21	2.69	403,768
	log(Sales)	4.44	−2.88	4.47	10.36	2.63	383,987
	log(Market value)	4.51	−0.92	4.41	10.51	2.40	390,174
Other	Asset tangibility	0.30	0.00	0.22	0.94	0.27	396,762
	Firm age	13.17	2.00	9.00	55.00	11.39	456,444

The table reports summary statistics of variables used in the empirical analysis. Variable definitions are provided in Appendix Table A1. All data come from the Compustat database and are for the years 1963 to 2018. Statistics reported are calculated after winsorization of the data at the 1st/99th percentiles.

Table A.3
Summary statistics of t-statistics generated in simulations

Category of dependent variable	Fixed effects	Mean	Mean absolute value	Std. Dev.	Min	Median	Max	% Sig. at 10% level	% Sig. at 5% level	% Sig. at 1% level	N
A. Randomly generated normally distributed explanatory variables (data used in Table 6)											
Profitability	Industry	−0.04	0.83	1.02	−3.8	−0.0	3.2	10.5%	5.3%	1.0%	16,000
	Firm	0.00	0.75	0.94	−3.8	0.0	3.7	8.0%	3.5%	0.6%	15,000
Value	Industry	−0.02	0.82	1.02	−3.6	−0.0	3.6	10.7%	5.3%	1.0%	16,000
	Firm	0.00	0.76	0.95	−3.7	−0.0	3.3	8.3%	3.8%	0.5%	15,000
Leverage	Industry	0.01	0.79	1.00	−3.8	0.0	4.2	10.1%	5.0%	0.9%	16,000
	Firm	−0.00	0.76	0.96	−3.8	−0.0	3.4	8.5%	3.9%	0.7%	15,000
Investment	Industry	−0.01	0.80	1.01	−3.6	0.0	3.9	10.7%	5.4%	1.0%	16,000
	Firm	0.00	0.77	0.96	−3.5	−0.0	3.4	8.6%	3.9%	0.8%	15,000
Payout	Industry	−0.03	0.80	1.01	−3.7	−0.0	4.0	9.6%	4.5%	1.2%	16,000
	Firm	0.02	0.76	0.94	−3.8	0.0	3.9	7.5%	3.2%	0.5%	15,000
Cash	Industry	−0.03	0.80	1.00	−3.6	−0.0	3.4	9.8%	4.9%	1.0%	16,000
	Firm	−0.03	0.78	0.97	−3.9	−0.0	3.5	9.1%	4.7%	0.8%	15,000
B. Quasi-random Compustat explanatory variables (data used in Table 7)											
Profitability	Firm	0.75	9.81	15.75	−173.3	0.0	148.5	83.1%	80.4%	75.0%	11,881
Value	Firm	0.38	8.55	13.05	−115.0	−0.4	70.2	79.4%	76.2%	71.2%	11,914
Leverage	Firm	−3.48	11.71	18.39	−154.1	−0.9	164.6	83.2%	80.2%	75.1%	11,905
Investment	Firm	−1.36	6.46	9.45	−64.7	−1.0	63.3	75.6%	71.8%	64.8%	11,862
Payout	Firm	0.19	5.26	7.49	−39.8	0.1	36.2	72.6%	68.6%	61.7%	11,872
Cash	Firm	−2.15	11.82	17.27	−100.5	−1.0	130.2	85.2%	82.4%	77.6%	11,897
C. Actual proposed determinants of leverage (data used in Table 8)											
Leverage (continuous)	Firm	−1.21	4.80	7.49	−38.5	−0.5	43.7	56.1%	52.2%	48.9%	636
Leverage (dummy)	Firm	1.57	2.64	5.26	−7.8	0.3	23.6	48.3%	43.8%	33.5%	176
D. Randomly generated normally distributed explanatory variables (data used in Table 9)											
Profitability	Firm	−0.00	0.78	0.95	−4.2	−0.0	4.7	7.0%	3.0%	0.5%	1,024,000
Value	Firm	0.02	0.76	0.93	−4.1	0.0	3.9	7.2%	3.0%	0.4%	1,024,000
Leverage	Firm	0.01	0.76	0.96	−4.6	0.0	4.1	8.4%	3.9%	0.7%	1,024,000
Investment	Firm	0.01	0.79	0.95	−4.2	0.0	4.1	6.4%	2.6%	0.4%	1,024,000
Payout	Firm	−0.00	0.80	0.96	−3.9	−0.0	4.1	6.4%	2.6%	0.4%	1,024,000
Cash	Firm	−0.01	0.77	0.95	−4.0	−0.0	3.8	8.1%	3.8%	0.6%	1,024,000

The table reports summary statistics of t-statistics for coefficients on explanatory variables in simulated regressions. Results are presented for randomly generated normally distributed explanatory variables in panels A and D, for quasi-random Compustat variables in panel B, and for actual proposed determinants of leverage in panel C. For each row in panel A, 1,000 explanatory variables are tested, with 16 different methodological combinations for each (15 with firm fixed effects). For each row in panel B, 1,000 explanatory variables are tested, with 12 different methodological combinations for each (the t-statistics occasionally cannot be calculated). In panel C, 49 explanatory variables (with 13 methodological combinations) are tested in the first row and 16 explanatory variables (with 11 methodological combinations) are tested in the second row. For each row in panel D, 1,000 explanatory variables are tested, with 1,024 methodological combinations for each. Compustat variables cover the years 1963 to 2018. Each regression includes year fixed effects and either industry fixed effects (two-digit SIC) or firm fixed effects, as noted. Standard errors are clustered at the firm level.

References

Adams, J., D. Hayunga, S. Mansi, D. Reeb, and V. Verardi. 2019. Identifying and treating outliers in finance. *Financial Management* 48:345–84.

Agrawal, A., and D. Matsa. 2013. Labor unemployment risk and corporate financing decisions. *Journal of Financial Economics* 108:449–70.

Almazan, A., A. De Motta, S. Titman, and V. Uysal. 2010. Financial structure, acquisition opportunities, and firm locations. *Journal of Finance* 65:529–63.

Andrews, I., and M. Kasy. 2019. Identification of and correction for publication bias. *American Economic Review* 109:2766–94.

Bae, K.-H., J.-K. Kang, and J. Wang. 2011. Employee treatment and firm leverage: A test of the stakeholder theory of capital structure. *Journal of Financial Economics* 100:130–53.

Baghai, R., H. Servaes, and A. Tamayo. 2014. Have rating agencies become more conservative? Implications for capital structure and debt pricing. *Journal of Finance* 69:1961–2005.

Baker, M., and J. Wurgler. 2002. Market timing and capital structure. *Journal of Finance* 57:1–32.

Banerjee, S., S. Dasgupta, and Y. Kim. 2008. Buyer-supplier relationships and the stakeholder theory of capital structure. *Journal of Finance* 63:2507–52.

Bebchuk, L., A. Cohen, and A. Ferrell. 2009. What matters in corporate governance? *Review of Financial Studies* 22:783–827.

Benmelech, E., and C. Frydman. 2015. Military CEOs. *Journal of Financial Economics* 117:43–59.

Berger, P., E. Ofek, and D. Yermack. 1997. Managerial entrenchment and capital structure decisions. *Journal of Finance* 52:1411–38.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119:249–75.

Bertrand, M., and A. Schoar. 2003. Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics* 118:1169–1208.

Bonica, A. 2016. Avenues of influence: On the political expenditures of corporations and their directors and executives. *Business and Politics* 18:367–94.

Brockman, P., X. Martin, and E. Unlu. 2010. Executive compensation and the maturity structure of corporate debt. *Journal of Finance* 65:1123–61.

Brodeur, A., N. Cook, and A. Heyes. 2020a. A proposed specification check for *p*-hacking. *AEA Papers and Proceedings* 110:66–69.

———. 2020b. Methods matter: *p*-hacking and publication bias in causal analysis in economics. *American Economic Review* 110:3634–60.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg. 2016. Star wars: The empirics strike back. *American Economic Journal: Applied Economics* 8:1–32.

Cain, M., and S. McKeon. 2016. CEO personal risk-taking and corporate policies. *Journal of Financial and Quantitative Analysis* 51:139–64.

Cain, M., S. McKeon, and S. Solomon. 2017. Do takeover laws matter? Evidence from five decades of hostile takeovers. *Journal of Financial Economics* 124:464–85.

Carlson, M., and A. Lazrak. 2010. Leverage choice and credit spreads when managers risk shift. *Journal of Finance* 65:2323–62.

Cassell, C., S. Huang, J. Sanchez, and M. Stuart. 2012. Seeking safety: The relation between CEO inside debt holdings and the riskiness of firm investment and financial policies. *Journal of Financial Economics* 103:588–610.

Chava, S., and A. Purnanandam. 2010. CEOs versus CFOs: Incentives and corporate policies. *Journal of Financial Economics* 97:263–78.

Chen, A. 2020. The limits of p-hacking: Some thought experiments. *Journal of Finance*, forthcoming.

Chen, A., and T. Zimmerman. 2020. Publication bias and the cross-section of stock returns. *Review of Asset Pricing Studies* 10:249–89.

Chordia, T., A. Goyal, and A. Saretto. 2020. Anomalies and false rejections. *Review of Financial Studies* 33:2134–79.

Christensen, G., and E. Miguel. 2018. Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature* 56:920–80.

Claessens, S., E. Feijen, and L. Laeven. 2008. Political connections and preferential access to finance. *Journal of Financial Economics* 88:554–80.

Coles, J., N. Daniel, and L. Naveen. 2014. Managerial incentives and risk-taking. *Journal of Financial Economics* 79:431–68.

DeAngelo, H., and R. Masulis. 1980. Optimal capital structure under corporate and personal taxation. *Journal of Financial Economics* 8:3–29.

De Long, B., and K. Lang. 1992. Are all economic hypotheses false? *Journal of Political Economy* 100:1257–72.

Denis, D., and J. Wang. 2014. Debt covenant renegotiations and creditor control rights. *Journal of Financial Economics* 113:348–67.

Denton, F. 1985. Data mining as an industry. *Review of Economics and Statistics* 67:124–27.

Derrien, F. and A. Kecskés. 2013. The real effects of financial shocks: Evidence from exogenous changes in analyst coverage. *Journal of Finance* 68:1407–40.

Dittmar, A., and A. Thakor. 2007. Why do firms issue equity? *Journal of Finance* 62:1–54.

Fang, V., T. Now, and S. Tice. 2009. Stock market liquidity and firm value. *Journal of Financial Economics* 94:150–69.

Faulkender, M., and M. Petersen. 2006. Does the source of capital affect capital structure? *Review of Financial Studies* 19:45–79.

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

Frésard, L. and P. Valta. 2016. How does corporate investment respond to increased entry threat? *Review of Corporate Finance Studies* 5:1–35.

Fukui, T., T. Mitton, and R. Schonlau. 2020. Determinants of capital structure: An expanded assessment. Unpublished manuscript.

Gao, H. 2010. Optimal compensation contracts when managers can hedge. *Journal of Financial Economics* 97:218–38.

Garvey, G., and G. Hanka. 1999. Capital structure and corporate control: The effect of antitakeover statutes on firm leverage. *Journal of Finance* 54:519–46.

Gaspar, J.-M., and M. Massa. 2007. Local ownership as private information: Evidence on the monitoring-liquidity trade-off. *Journal of Financial Economics* 83:751–92.

Geary, R., 1935. The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika* 27:310–32.

Gelman, A., and E. Loken. 2014. The statistical crisis in science. *American Scientist* 102:460–65.

Graham, J., M. Lang, and D. Shackelford. 2004. Employee stock options, corporate taxes, and debt policy. *Journal of Finance* 59:1585–1618.

Graham, J., and A. Tucker. 2006. Tax shelters and corporate debt policy. *Journal of Financial Economics* 81:563–94.

Harford, J., R. Schonlau, and J. Stanfield. 2019. Trade relationships, indirect economic links, and mergers. *Management Science* 65:3085–3110.

Harvey, C. 2017. Presidential address: The scientific outlook in financial economics. *Journal of Finance* 72:1399–1440.

Harvey, C. 2019. Replication in financial economics. *Critical Finance Review* 8:1–9.

Harvey, C., and Y. Liu. 2020. False (and missed) discoveries in financial economics. *Journal of Finance* 75:2503–53

Harvey, C., Y. Liu, and A. Saretto. 2020. An evaluation of alternative multiple testing methods for finance applications. *Review of Asset Pricing Studies* 10:199–248.

Harvey, C., Y. Liu, and H. Zhu. 2016. . . . and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Heider, F., and A. Ljungqvist. 2015. As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes. *Journal of Financial Economics* 118:684–712.

Hirsch, B., and D. Macpherson. 2003. Union membership and coverage database from the current population survey: Note. *Industrial & Labor Relations Review* 56:349–54.

Hoberg, G., and G. Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423–65.

Hoberg, G., G. Phillips, and N. Prabhala. 2014. Product market threats, payouts, and financial flexibility. *Journal of Finance* 69:293–324.

Hong, H., and M. Kacperczyk. 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93:15–36.

Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–2133.

Huang, J., and D. Kisgen. 2013. Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics* 108:822–39.

Ioannidis, J. 2008. Why most discovered true associations are inflated. *Epidemiology* 19:640–58.

Kale, J., and H. Shahrur. 2007. Corporate capital structure and the characteristics of suppliers and customers. *Journal of Financial Economics* 83:321–65.

Kim, J., and P. Ji. 2015. Significance testing in empirical finance: A critical review and assessment. *Journal of Empirical Finance* 34:1–14.

Kini, O., and R. Williams. 2012. Tournament incentives, firm risk, and corporate policies. *Journal of Financial Economics* 103:350–76.

Klasa, S., H. Ortiz-Molina, M. Serfling, and S. Srinivasan. 2018. Protection of trade secrets and capital structure decisions. *Journal of Financial Economics* 128:266–86.

Kostovetsky, L. 2015. Political capital and moral hazard. *Journal of Financial Economics* 116:144–59.

Kurshev, A., and I. Strebulaev. 2015. Firm size and capital structure. *Quarterly Journal of Finance* 5:1–46.

Larkin, Y. 2013. Brand perception, cash flow stability, and financial policy. *Journal of Financial Economics* 110:232–53.

Leamer, E. 1983. Let's take the con out of econometrics. *American Economic Review* 73:31–43.

Leamer, E., and H. Leonard. 1983. Reporting the fragility of regression estimates. *Review of Economics and Statistics* 65:306–17.

Lev, B., and S. Sunder. 1979. Methodological issues in the use of financial ratios. *Journal of Accounting and Economics* 1:187–210.

Li, J., and D. Tang. 2016. The leverage externalities of credit default swaps. *Journal of Financial Economics* 120:491–513.

Linnainmaa, J., and M. Roberts. 2018. The history of the cross-section of stock returns. *Review of Financial Studies* 31:2606–49.

Lovett, M., R. Peres, and R. Shachar. 2014. A data set of brands and their characteristics. *Marketing Science* 33:609–17.

Malmendier, U., G. Tate, and J. Yan. 2011. Overconfidence and early-life experiences: The effect of managerial traits on corporate financial policies. *Journal of Finance* 66:1687–1733.

Massa, M., A. Yasuda, and L. Zhang. 2013. Supply uncertainty of the bond investor base and the leverage of the firm. *Journal of Financial Economics* 110:185–214.

Matsa, D. 2010. Capital structure as a strategic variable: Evidence from collective bargaining. *Journal of Finance* 65:1197–1232.

McCloskey, D., and S. Ziliak. 1996. The standard error of regressions. *Journal of Economic Literature* 34:97–114.

McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–31.

Mitton, T. 2021. Economic significance in corporate finance. Unpublished manuscript.

Morey, M., and S. Yadav. 2018. Documentation of the file drawer problem in academic finance journals. *Journal of Investing* 27:143–47.

Mulherin, H., J. Netter, and A. Poulsen. 2018. Observations on research and publishing from nineteen years as editors of the *Journal of Corporate Finance*. *Journal of Corporate Finance* 49:120–24.

Rosenthal, R. 1979. The “file drawer problem” and tolerance for null results. *Psychological Bulletin* 86:638–41.

Saretto, A., and H. Tookes. 2013. Corporate leverage, debt maturity, and credit supply: The role of credit default swaps. *Review of Financial Studies* 26:1190–1247.

Sen, R., and R. Tumarkin. 2015. Stocking up: Executive optimism, option exercise, and share retention. *Journal of Financial Economics* 118:399–430.

Serfling, M. 2016. Firing costs and capital structure decisions. *Journal of Finance* 71:2239–86.

Shivdasani, A., and I. Stefanescu. 2010. How do pensions affect corporate capital structure decisions? *Review of Financial Studies* 23:1287–1323.

Shleifer, A., and R. Vishny. 1992. Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance* 47:1343–66.

Simmons, J., L. Nelson, and U. Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22:1359–66.

Simonsohn, U., J. Simmons, and L. Nelson. 2020. Specification curve analysis. *Nature Human Behaviour* 4:1208–14.

Smith, J. 2016. US political corruption and firm financial policies. *Journal of Financial Economics* 121:350–67.

Titman, S., and R. Wessels. 1988. The determinants of capital structure choice. *Journal of Finance* 43:1–19.

Volkova, E. 2018. Blockholders diversity: Effect of polyphony on the power of monitoring. Unpublished manuscript.

Xu, J. 2012. Profitability and capital structure: Evidence from import penetration. *Journal of Financial Economics* 106:427–46.

Yan, X., and L. Zheng. 2017. Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies* 30:1382–1423.

Young, C., and K. Holsteen. 2017. Model uncertainty and robustness: A computation framework for multimodel analysis. *Sociological Methods and Research* 46:3–40.