Identifying and Treating Outliers in Finance

John Adams, University of Texas Arlington
Darren Hayunga, University of Georgia
Sattar Mansi, Virginia Polytechnic Institute and University
David Reeb, National University of Singapore
Vincenzo Verardi, Université de Namur

February 2019

Abstract

Outliers represent a fundamental challenge in empirical finance research. We investigate whether the routine techniques used in finance research to identify and treat outliers are appropriate for the data structures we observe in practice. Specifically, we propose a multivariate identification strategy that can effectively detect outliers. We also introduce an estimator that minimizes the bias outliers caused in both cross-sectional and panel regressions and provide outlier mitigation guidance. Using replications of four recently published studies in premier finance journals, we show how adjusting for multivariate outliers can lead to significantly different results.

The authors would like to thank the editors Utpal Bhattacharya and Bing Han, an anonymous referee, Robert Andersen, Bo Becker, C.Y. Choi, Rachel Croson, Paul Goldsmith-Pinkham, Andrey Golubov, Jim Musumeci, David Rakowski, Jan Sokolowsky, Peter Westfall, Mahmut Yasar, Zeynep Senyuz, seminar participants at Virginia Tech and the University of Texas Arlington, and conference participants at the 2018 Financial Management Association Annual Conference, the 2018 Financial Management Association European Conference, and the 2017 World Finance Conference. Special thanks to Adam Harper, Vasanth Rajarajan, and Anurag Mehrotra for help with the data collection. Vincenzo Verardi gratefully acknowledges financial support from FNRS.

I. Introduction

Outliers represent a persistent concern in empirical finance research. We are all aware that outliers, or observations that deviate markedly from the data, potentially lead to biased coefficient estimates in least-square regressions (Edgeworth, 1887).[1] Researchers often seek to identify these potential outliers by examining descriptive statistics regarding the variables of interest (Dittmar and Duchin, 2016) effectively examining observations three standard deviations from the mean. After identifying these influential observations, the econometrician typically relies on mitigation techniques to remedy this outlier problem (Henry and Koski, 2017). A review of recent articles that identify outliers in prominent finance journals indicates that almost all studies rely on univariate identification.[2] Table I indicates that the vast majority of these studies winsorize the data or perform some sort of list-wise deletion. Yet, this identification and treatment of outliers implicitly relies on outliers arising in a univariate context.

We ask a fundamental, but simple question. Are the techniques we commonly use in finance to identify and treat outliers appropriate for the data structures we observe in practice? Many of the methods to identify and treat outliers, such as winsorizing, trimming, or dropping the affected observations, arose in a period with limited datasets and computer power. By necessity, these methods focused on identifying and treating outliers in a univariate setting, but studies in finance almost always require multivariate analysis. A simple example provides a useful illustration. Table II displays a small dataset, where descriptive statistics indicate that none of the observations contain univariate outliers. Yet, two of the observations include outliers in a

---

[1] While definitions vary, outliers describe observations that deviate so much from other observations as to arouse suspicions about the mechanism generating the data (Hawkins, 1980). We use the term bias to mean the difference between an estimator's expected value and the value of the parameter as determined by the bulk of the data.
[2] We examine articles in the *Journal of Finance*, the *Journal of Financial Economics*, *Review of Financial Studies*, and the *Journal of Financial and Quantitative Analysis*. We find that 66% use OLS as the primary statistical technique.

multivariate setting, dramatically influencing the coefficient estimates in an OLS regression framework.[3] Intuitively, if multivariate (i.e., regression) outliers arise in a non-random fashion, trimming and dropping potentially introduces sample selection problems and biased coefficient estimates (Heckman, 1979). Table II demonstrates that neither winsorizing nor trimming mitigates the influence of the multivariate outliers. Instead, these univariate outlier mitigation strategies actually exacerbate the multivariate outlier problem (our example is consistent with Bollinger and Chandra, 2005). The example provided in Table II clearly demonstrates that despite being the best linear unbiased estimator of the conditional expectation function from a purely statistical standpoint, naively using OLS can lead to incorrect economic inferences when there are multivariate outliers in the data.

Outliers arise in a variety of ways including data errors, variable construction, omitted variables, sampling errors, non-normality, or chance. Outliers can also be the most important data in a sample when they reflect some unusual fact that will lead to an improvement in economic theory or model specification (Zellner, 1981). Therefore, identifying multivariate outliers is a key step in evaluating their impact in empirical finance research. Traditional methods, such as studentized residuals or Cook's D, while simple to implement and easy to evaluate, suffer from a masking problem that occurs when specifying too few outliers in the test. For example, if we are testing for a single outlier when there are, in fact, two (or more) outliers, these additional outliers may influence the value of the test statistic enough so that no points are identified as outliers. Traditional methods also rely heavily on their assumptions of normality. In this paper, we propose

[3] To illustrate with a more real world example, consider a panel data of Minnesota employees containing information on natural hair color, height, weight, eye color, and ethnicity. In this sample, neither a 5'2" person nor an employee with blues eyes or an employee with blond hair would likely register as univariate outliers. Similarly, neither an observation regarding a Chinese male employee nor an employee weighing 235 pounds would appear as outliers. However, if all of these characteristics describe a single employee, then we might suspect this observation is an outlier.

an identification strategy using outlier robust estimation as in Rousseeuw and van Zomeren (1990). We find this method effectively identifies the outliers and tests for their influence. We further propose an outlier robust regression approach that minimizes the bias outliers caused in both cross-sectional and panel regressions. The objective is to provide an outlier robust estimator that is as efficient as OLS when the data is normally distributed, but is more efficient than OLS and unbiased when the data contains outliers.

Yet, there is a cost to our approach. In particular, it is computationally intensive. We use a combination of base robust estimators, specifically *M*-estimators, and *S*-estimators, which are termed *MM*-estimators as described in Yohai (1987). The *MM*-estimators combine a high breakdown point, or the largest percentage of outliers in a sample that the estimator can handle without producing arbitrary results, with relatively high efficiency when compared to OLS. To the best of our knowledge, there are no readily available procedures that compute *MM*-estimators or any other high breakdown point estimators with clustered standard errors. We rely on the theory of the generalized method of moments to calculate these clustered standard errors (Croux, Dhaene, and Hoorelbeke, 2008).

Researchers often need to control for group-specific fixed effects using binary variables. However, there are no outlier robust regression methods available that can account for a large number of fixed effects. We address this issue by presenting a new method that draws on insights from Aquaro and Čížek (2013). We conduct simulations to show this method effectively mitigates outlier bias in models that use fixed effects and/or clustered standard errors (see online appendix). Our analysis demonstrates that outliers in continuous variables can cause biases in the OLS coefficients of binary variables. We also find that outlier bias in binary variables increases when continuous independent variables are correlated, a common occurrence in finance samples. These findings lead to a concern that the results of studies that use binary variables suffer from outlier-induced biases. Our outlier robust estimators can be used as a diagnostic tool yielding different results than OLS when influential outliers are present.

Once identified, the origins of the outliers can be determined. The decision as to how, or even whether it is appropriate, to mitigate the influence of multivariate outliers depends upon their cause and economic theory. For example, in the case of data entry errors, the ideal mitigation strategy is dropping or correcting errant observations. Modern financial datasets are often large making it impractical or impossible to manually search for outliers at the observational level. Our approach effectively identifies the most influential outliers that change coefficient estimates. Since influential outliers often represent only a small fraction of the total observations, researchers can minimize manual examination and data cleaning costs using the proposed approach.

Of course, not all influential observations are the result of data errors. Some observations are informative in the sense they represent omitted variables or interesting phenomena not previously considered. For these informative and influential outliers, modeling enhancements (e.g., additional control variables) can be effective in mitigating the bias they cause in the estimated coefficients of the main variables of interest. Economic theory can help guide any further mitigation efforts. When the hypothesized relation between the dependent and independent variable is a general effect and not an outlier effect (i.e., driven by rarely occurring events or circumstances), outlier bias can be mitigated by either dropping the most extreme outliers or employing outlier robust regressions that place less weight on extreme observations than OLS does. For example, when theory suggests a general effect, influential outliers comprising a small fraction of the sample should not drive the empirical results.

However, these mitigation approaches are not appropriate when outliers, as tail risk events, are the most informative observations as removing the largest manifestation of an effect can make it appear insignificant when it is not. For example, when examining the impact of low probability economic disasters on equity premiums as in Reitz (1988), Barro (2006), and Welch (2016), naively dropping the most influential outliers would lead to incorrect inferences. [4] In these cases, outlier mitigation should be limited to removing or correcting data errors and improving model specifications to account for any omitted variables.

Our recommended approach to multivariate outliers is comprised of five steps: 1) test for the presence of multivariate outliers since they are suggestive of bad data (e.g., data entry errors, sampling errors, and omitted variables), 2) identify outliers robustly in a multivariate context, 3) carefully consider and examine the nature and origin of the outliers, 4) correct data and omitted variables errors, and 5) consider the nature of the research question and economic theory to determine whether to mitigate further by dropping the influential observations in the OLS regressions or by employing outlier robust estimators. While outliers potentially influence statistical and economic inferences, they may not systemically affect the results in finance research. We examine this possibility by replicating four recently published papers using our outlier robust estimator as a diagnostic tool. For tractability, we concentrate our analysis to two main areas in finance (i.e., corporate finance and investments) with relatively large datasets to minimize concerns about sample sizes. We identify two published articles in premier finance journals via a formal screening process that biases against finding outlier problems and where the

[4] In another example, Moeller, Schlingemann, and Stulz (2005) find large average losses in shareholder value in M&As. They report a small number of 87 announcements that resulted in a collective loss in acquiring-firm shareholder wealth of $397 billion, but for the overwhelming majority of 4,109 announcements, shareholders of the acquiring firms collectively gained $157 billion. Because the main research question in Moeller et al. (2005) is the average wealth effect of all M&A announcements on shareholder value, it would be incorrect to mitigate the influence of the 87 outliers.

authors made their datasets and code available. We find the estimated coefficients for the primary variables and covariates change in terms of statistical significance and economic importance after mitigating the influence of outliers. We also collect data on our own in order to evaluate two additional articles where the authors do not disclose this information, again finding evidence of multivariate outliers. We provide illustrations of robust identification and implementation across a variety of empirical settings demonstrating drastic changes in the magnitude and signs of the coefficient estimates of interest. In short, this initial analysis suggests that the common methods used to identify outliers in finance can lead to distinct distortions in empirical analyses. The results in the studies that we formally investigate, an admittedly small sample, appear to stem from multivariate outliers in the data.

Our paper makes two main contributions. First, this is the only research that proposes a comprehensive multivariate outlier identification strategy, and demonstrates how to effectively detect outliers, test for their influence, and mitigate (when appropriate) the bias they cause in both cross sectional and panel regressions. Our approach also provides outlier robust clustered standard errors and is capable of handling the large numbers of fixed effects common in finance models. We empirically test this approach using four replications from recently published papers in premier finance journals and show how adjusting for outliers can lead to different results. Current published work that incorporates replications highlights the importance of carefully examining and treating outliers. We argue that a comprehensive approach to addressing outliers improves internal and external validity thereby reducing the need for further investigations.[5]

[5] For example, in a study on board composition, Chhaochharia and Grinstein (2009) find a negative and significant relation between CEO pay and board independence enhancements. In a subsequent study, Guthrie, Sokolowsky, and Wan (2012) replicate Chhaochharia and Grinstein's (2009) work and show the relation is driven by two CEOs out of a sample of 865 firms. Thus, a comment by Chhaochharia and Grinstein (2009) and a rejoinder by Guthrie et al. (2012) attempt to confirm their reported results. Importantly, Chhaochharia and Grinstein (2009) and Guthrie et al. (2012)

In addition, we contribute to a growing literature on improving identification and estimation techniques in finance research. Angrist and Pischke (2010) argue that innovations in econometric identification techniques (e.g., unobserved heterogeneity, endogeneity, clustered standard errors in panel data, measurement errors, and natural experiments) represent a credibility revolution that enhances the ability of researchers to make accurate statements. Bowen, Frésard, and Taillard (2017) argue that researchers should quickly adopt useful econometric innovations, but various frictions hinder their diffusion. In the case of robust outlier technology, the frictions center on the lack of widely known and readily available methods to identify, test, and treat outliers. These frictions are coupled with the common belief that the standard outlier mitigation techniques, such winsorizing and trimming, provide protection against these extreme observations. In this article we attempt to alleviate these frictions.

II. Univariate vs. Multivariate Outliers

A. Limitations of Univariate Outlier Treatments

The most common outlier treatments in finance are winsorizing, trimming, and dropping (referred to hereafter as WTD). For example, of the 3,572 studies published in the top four finance journals from 2008-2017, only 999 (or 28%) mention outliers. Of the 717 studies that utilize OLS regression and mention outliers, the large majority use winsorizing (52%), trimming (16%), or dropping (17%). Researchers often argue these methods are effective or that outliers do not unduly influence their results. The following examples are excerpts from the top four finance journals over the study period. Authors note “to prevent outliers from influencing the analysis all variables

focus on outliers in a univariate context. Altogether, three publications and a tremendous amount of time was spent attempting to validate each author’s work. In our own replication of their work, we find neither Chhaochharia nor Grinstein (2009) nor Guthrie et al. (2012) reliably mitigate outliers in a multivariate context.

are winsorized" and "the dependent variable was winsorized to remove the effects of extreme outliers" and "to avoid potential problems with outliers all variables are winsorized." Similar language is also common when using trimming, dropping, or other techniques.

The main concern with WTD is that these are univariate attempts to correct a multivariate problem. In estimating a multivariate regression, observations that may not appear extreme in a single variable can exert outsized influence on the overall model. Accordingly, WTD can only be expected to reliably mitigate outlier-induced bias in univariate descriptive statistics. Since most empirical work in finance utilizes multivariate analysis, these outlier mitigation techniques are not consistently effective.

Univariate outlier treatments can also alter the data. For instance, across the distribution of a variable of interest, winsorizing requires identifying the h smallest and largest values and replacing them with the next smallest or largest values, where h is an integer or percentage. In a data panel with observations in rows and variables of interest in columns, the extreme value(s) within a column will be determined relative to the other column values and the researcher will alter both the highest and lowest amounts to what is deemed more reasonable. An issue with this procedure is that the setting of h is arbitrary. Also, replacing a column value for a particular observation with the value of another observation changes the information contained in that observation. Further, by identifying and targeting extreme values within a single column, trimming and dropping are similar to winsorizing, but instead of replacing the value of a column attribute, these techniques remove the entire row observation. And, while the observation is not maintained as a transformed record, trimming and dropping remove observations that may be well behaved in all other columns and can provide valuable insight into the data generating process. In fact, outliers as influential observations can be the most important data in a sample (Belsley, Kuh, and Welsh, 1980).

Another issue with WTD is that researchers continue to use OLS on the altered sample. Trimmed or winsorized least squares estimates can still be influenced (potentially greatly) by one remaining outlier. While employing a median rather than a mean estimator on the altered sample can provide more robustness, in simulations, we find that median regressions do not reliably mitigate the influence of multivariate outliers. The results of the simulations are provided by the authors upon request.

In addition, WTD can actually introduce new sample problems. Bollinger and Chandra (2005) and Verardi and Wagner (2011) find that trimming can lead to biased coefficient estimates even in samples without outliers. Moreover, arbitrarily winsorizing or removing observations with large values creates a sample selection problem (Heckman, 1979). The extreme values do not occur by accident. Rather, they arise from an underlying data generating process and removing them can introduce a new bias in parameter estimates.

B. Univariate Identification and Treatment of Multivariate Outliers: An Illustration

Panel A of Table II presents data from two small illustrative data sets. The first data set, labeled the "No Outlier Sample," features 20 simulated observations where the dependent variable, Y, equals $.5X1 + .5X2$ + a random error term. Independent variables $X1$ and $X2$ are randomly generated with values in the range of 1-20. The second dataset, labeled the "Multivariate Outlier Sample," is identical to the first except we replace the independent variables in Observation Number 18 and the dependent variable in Observation Number 19 with smaller values. These replacements represent multivariate outliers with dependent variable values that are larger

(Observation Number 18) and smaller (Observation Number 19), relative to their independent variable values, than the remaining observations.

Similar to Anscombe (1973), Table II demonstrates that descriptive statistics cannot reliably identify outliers.[6] The mean and median values of the two samples are essentially the same and the standard deviation values are identical. More importantly, the "No Outlier Sample" and the "Multivariate Outlier Sample" exhibit the same minimum and maximum values. Univariate identification essentially entails some selection of the largest or smallest values of each variable. However, in this example, the multivariate outliers do not have extreme dependent or independent variable values and, as such, are not identified as outliers.

Panel B of Table II reports the effects of univariate outlier mitigation and multivariate outliers on regression estimates. The estimated coefficients for the "No Outlier Sample" in Column 1 are reasonably close to their expected values and the adjusted R^2 is high. Winsorizing and trimming, univariate outlier mitigation techniques, have minimal impact on the regression estimates for the "No Outlier Sample" (Columns 2 and 3). Column 4 reports the results for the "Multivariate Outlier Sample." The regression coefficient estimates are far from their expected values and the adjusted R^2 is low. Columns 5 and 6 indicate that winsorizing and trimming do not effectively mitigate the influence of the multivariate outliers and actually appear to make the estimation worse. Column 7 illustrates the importance of identifying multivariate outliers. After removing the multivariate outliers, the regression estimates are similar to those obtained for the "No Outlier Sample" in Column 1.

C. Multivariate Outlier Identification

[6] Anscombe's (1973) Quartet illustrates the importance of visualizing the data prior to analysis and demonstrates how outliers can affect causality inferences.

Empirical findings and conclusions can vary based upon the type of outliers. Figure I illustrates the three multivariate outlier types: 1) vertical, 2) good leverage, and 3) bad leverage. A vertical outlier is an observation outlying in the dependent variable dimension, but not outlying in the independent variable space. A good leverage point is an extreme observation outlying in the independent variable space, but located near the regression line. When good leverage points are very close to the regression line, they marginally affect parameter estimation, but they can affect statistical inference by deflating standard error estimates. A bad leverage point is an observation that is outlying in the independent variable space and located far from the true regression line. Bad leverage points often significantly affect the estimation of both the intercept and the slope. Since the difference between good and bad leverage is a matter of degree, we focus on bad leverage points.[7]

The first step in assessing outliers is identification. Dehon, Gassner, and Verardi (2012) follow the logic of Hausman (1978) and develop a procedure that compares estimates from outlier robust estimators and OLS. Outlier robust estimation attempts to utilize all available data, but minimize the effect of extreme observations (see Appendix A for a primer on robust estimators). When the test fails to reject, the more efficient OLS is the best estimator.[8] After testing for the presence of influential observations, the researcher should next explicitly identify extreme values to check for correctness and gain further insight into the data generating process. Researchers have previously used leverage, studentized residuals, DFBETAs, DFFITS, Cook's D, and partial regression plots as a way to identify outliers. The main limitation with these tests is their attempt to produce normal looking residuals even when the data are not normal (Rousseeuw and van

[7] As an illustration, Observations 18 and 19 in Table II are examples of bad leverage points.
[8] When outlier mitigation is the objective and the test rejects OLS as unbiased, the next step is to perform a second Dehon et al. (2012) test that compares the robust *S*-estimator to a more efficient *MM*-estimator. The highest possible efficiency for the *MM*-estimator is determined via repeated application of the Dehon et al. (2012) test.

Zomeren, 1990). Also, the popular Cook's D often suffers from a masking effect that occurs when a group of extreme values mask the impact of each other.

Rousseeuw and van Zomeren (1990) argue that a better identification method is to plot robust standardized residuals against robust distances. The standardized residuals are from an outlier robust estimation procedure, such as S-estimation. On the other axis is a measure of multivariate outlyingness. Multivariate outlyingness is defined as the Mahalanobis distance where the multivariate location vector and covariance matrix are estimated robustly. Specific observations merit special attention if they exceed common y- and x-dimension boundaries. Standard practice is to use y limits of ±2.25 to represent the values from the standard normal that separate the 2.5% most remote region from the central mass. For the x dimension, robust distances have high leverage if their magnitude is greater than 0.975 of the chi-squared distribution with degrees of freedom equal to the number of parameters in the model ($\chi^2_{p,0.975}$). We demonstrate the usefulness of the Rousseeuw and van Zomeren (1990) detection method in several of our empirical tests (see Section IV and Figure II).

III. Outlier Mitigation

A. Outlier Robust Estimators

Regression analysis is the primary statistical method used in empirical finance research. When properly fitted, regression estimates provide a powerful summary of relations in the data. The goal of linear regression is to converge to the true values of the coefficients by minimizing a loss function on the residuals. When the assumptions of the classical linear regression model are violated, however, coefficients are estimated with error or bias, which can lead to spurious inferences. One of the main assumptions of linear regression is that estimation errors, or residuals, are distributed normally. In this case, OLS produces the best unbiased coefficient estimates with the smallest variance. Even when the residuals are not distributed normally, the OLS estimator is still the best linear unbiased estimator, a weaker condition indicating that among all linear unbiased estimators, OLS coefficient estimates have the smallest variance.

However, it is trivial to show that when a sample contains extreme observations, OLS becomes markedly inferior to outlier robust estimators.[9] The robustness of an estimator is the level of resistance to change that an estimator has to outliers (Andersen, 2008). When the residuals are not distributed normally, it is frequently possible to find robust estimators that are more efficient or produce estimates with smaller variances than OLS. Since the underlying error distribution is rarely known with certainty, robust estimation procedures that have attracted the greatest attention in the statistics literature are those that are concerned with finding estimators that are only slightly less efficient than OLS when the errors are normally distributed, but can be considerably less biased and more efficient in the presence of outliers in the data. The trade-off of efficiency and bias is critical to selecting the appropriate estimator.

Appendix A highlights the base parametric robust estimators: *L*-estimators, *R*-estimators, *M*-estimators, and *S*-estimators. Since each estimator has its trade-off, the most recent robust modeling uses combinations of the base robust estimators. The two primary estimators used are *M*-estimators and *S*-estimators, which are termed *MM*-estimators. To compute *MM*-estimators, Verardi and Croux (2009) program the algorithm of Salibian-Barrera and Yohai (2006) and use the iteratively reweighted OLS algorithm with the *S*-estimate as their initial value. The algorithm for computing the initial *S*-estimators begins by estimating regression parameters on randomly selected subsets. The intuition for multiple subsets is to obtain at least one subset without outliers

[9] In unreported simulations, we find the superiority of outlier robust regressions to OLS in samples with outliers. These results are available on request.

and the final *S*-scale estimate is from the subset with the smallest scale. Once the *S*-estimator is estimated, the *MM*-estimator is computed via the iteratively reweighted OLS algorithm. *MM*-estimators combine a high breakdown point (BDP) of 50% with relatively high efficiency, 95% relative to OLS under the Gauss-Markov assumptions.[10]

To the best of our knowledge, there are no readily available procedures that compute *MM*-estimators or any other high BDP estimators with clustered standard errors. Relying on the theory of the generalized method of moments (GMM), these standard errors can be calculated (Croux et al., 2008).[11] Specifically, a preliminary *S*-estimator is fit with a given loss function ρ_0. Parameters $\boldsymbol{\beta}$ and σ are estimated by $\hat{\boldsymbol{\beta}}_{S;\rho_0}$ and $\hat{\sigma}_{\rho_0}$. Then, an *M*-estimator with loss function ρ (with first derivative ψ) that allows a higher efficiency is estimated with the scale parameter fixed to $\hat{\sigma}_{\rho_0}$. The functions ρ_0 and ψ, which are chosen by the statistician, are non-constant, scalar-valued, and differentiable. Furthermore, ψ is odd, ρ_0 is even and non-decreasing on $[0,\infty[$with $\rho_0(0) = 0$, and b is a selected constant that is usually chosen to be $E_\Phi[\rho_0(u)]$ where Φ denotes the standard normal distribution. In the *MM*-estimation procedure, the estimations are such that:

[10] A breakdown point is the largest percentage of outliers in a sample that the estimator can handle without producing arbitrary results with relatively high efficiency when compared to OLS. For a more complete discussion of *MM*-estimators see Maronna, Martin, and Yohai (2006, p. 124).

[11] Employing Monte Carlo simulations to compute outlier robust clustered standard errors is not feasible given the computationally intensive nature of outlier robust regressions.

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{y_i - \mathbf{x}_{\boldsymbol{i}}^{'}\ \hat{\boldsymbol{\beta}}_{\boldsymbol{MM};\boldsymbol{\rho_0},\boldsymbol{\rho}}}{\hat{\sigma}_{\boldsymbol{\rho_0}}}\right)\mathbf{x}_i = 0 \\ \frac{1}{n}\sum_{i=1}^{n}\rho_0^{'}\left(\frac{y_i - \mathbf{x}_{\boldsymbol{i}}^{'}\ \hat{\boldsymbol{\beta}}_{\boldsymbol{S};\boldsymbol{\rho_0}}}{\hat{\sigma}_{\boldsymbol{\rho_0}}}\right)\mathbf{x}_i = 0 \\ \frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{y_i - \mathbf{x}_{\boldsymbol{i}}^{'}\ \hat{\boldsymbol{\beta}}_{\boldsymbol{S};\boldsymbol{\rho_0}}}{\hat{\sigma}_{\boldsymbol{\rho_0}}}\right) - b = 0. \end{cases} \tag{1}$$

The first line of Equation (1) is the first order condition of the final *M*-estimator, the second line is the first order condition for the preliminary *S*-estimator, and the last line corresponds to the sample equivalent of Equation (A-10) in Appendix A. The Tukey (1991) biweight function that has a smooth ψ is used for both the preliminary *S*-estimator, as well as the final *MM*-estimator.

$\left(\hat{\boldsymbol{\beta}}_{MM;\rho_0,\rho}^{'}, \hat{\boldsymbol{\beta}}_{S;\rho_0}^{'}, \hat{\sigma}_{\rho_0}\right)^{'}$ is the first order equivalent with the generalized method of moment estimator for $\left(\boldsymbol{\beta}^{'}, \boldsymbol{\beta}^{'}{}_0, \sigma_{\rho_0}\right)'$. Under usual GMM conditions:

$$\sqrt{n}\left(\begin{pmatrix}\hat{\boldsymbol{\beta}}_{MM;\rho_0,\rho}\\ \hat{\boldsymbol{\beta}}_{S;\rho_0}\\ \hat{\sigma}_{\rho_0}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\beta}_0\\ \sigma\end{pmatrix}\right)\xrightarrow{d}\mathcal{N}(0,\mathbf{V}_{MM}).$$

The asymptotic variance of the *MM*-estimator is available from the standard GMM estimator (top left block of V_{MM}). Thus, it is straightforward (Croux et al., 2008) to have a covariance matrix for the *S*- and the *MM*-estimator that is robust to heteroskedasticity, serial correlation, and/or clustered standard errors by relying on standard GMM theory.

The subsampling algorithm can result in collinear subsamples when there are multiple independent binary variables, which is a common occurrence in finance data that contain numerous binary variables to capture firm, year, and event effects. To remedy this issue, Maronna and Yohai (2000) introduce a *MS*-estimator that alternates an *S*-estimator and an *M*-estimator for continuous and binary variables, respectively, until convergence.

For the fixed effects panel data models in finance, Bramati and Croux (2007) recommend replacing the initial *S*-estimator with a *MS*-estimator in calculating *MM*-estimators. As stated by Aquaro and Čížek (2013), the problem with this method is that using the *MS*-estimator for panel data fixed effects estimations implies that the fixed effects must be explicitly estimated causing bias due to the nonlinearity of the procedure when the number of periods is fixed. Aquaro and Čížek (2013) argue the Bramati and Croux (2007) recommended method is consistent only if the number of time periods increases to infinity making them unsuitable for short panels. More practically, in replications and unreported simulations, we find that Bramati and Croux's (2007) method often does not converge when there are numerous fixed effects, an attribute of many finance models.

We address this issue by developing a procedure that computes *MM*-estimators with clustered standard errors and can handle a larger numbers of fixed effects relying on Aquaro and Čížek (2013). Consider a static linear fixed effect panel data model $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$, $i = 1,\ldots,n$ and $t = 1,\ldots,T_{max}$, where y_{it} is the dependent variable, $\mathbf{x}_{it}$ is the vector of the covariates, and $\boldsymbol{\beta}$ is the vector of parameters of interest. Indices *i* and *t* index individuals and time, respectively. T_{max} is the maximum value for the time index t. The unobservable term is a combination of the individual fixed effects α_i and the error term ε_{it}. Parameters $\boldsymbol{\beta}$ can easily be estimated if the individual fixed effects are removed from the model equation. A simple way of estimating the parameters of interest is to apply the well-known first-difference transformation $\Delta y_{it} = \Delta\mathbf{x}_{it}'\boldsymbol{\beta} +$

$\Delta\varepsilon_{it}$ and estimating the resulting model with a linear regression estimator. Subsequently, the researcher can use an *S*- and/or *MM*-estimator to mitigate outliers.

Alternatively, the researcher can obtain more accurate estimates by eliminating individual effects taking all pairwise differences within each individual (Aquaro and Čížek, 2013). This pairwise difference estimator transformation yields $\Delta^s y_{it} = \Delta^s \mathbf{x}'_{it}\boldsymbol{\beta} + \Delta^s \varepsilon_{it}$. This transformation removes the individual-specific effect, but generates a larger sample size of $nT(T-1)/2$ instead of $n(T-1)$ as differences for all $s = 1,\ldots,t-1$ are considered. To take into account the fact that individual specific observations are not independent, clustered standard errors have to be systematically considered at the individual level when running the linear regression estimator.

In practice, the idea is to plug $\Delta^s y_{it}$ into y_{it} and $\Delta^s \mathbf{x}'_{it}$ into $\mathbf{x}'_{it}$ in Equation (1) and estimate the parameters using GMM. Being a GMM estimator, the heteroskedasticity, clustering, and autocorrelation consistent estimator of the standard errors are readily available. The coefficients of interest will be $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{MM;\rho_0,\rho}$ and their estimated covariance will be the upper left square of matrix $\widehat{\mathbf{V}_{MM}}$.

Naturally, one might also want to cluster at an additional level than the individual one. As Thompson (2011) illustrates, computation of the two-way cluster-robust variance component estimation is straightforward. The two-way clustered variance can be calculated from $\mathbb{V}(\hat{\boldsymbol{\beta}}) = \mathbb{V}_1(\hat{\boldsymbol{\beta}}) + \mathbb{V}_2(\hat{\boldsymbol{\beta}}) - \mathbb{V}_{12}(\hat{\boldsymbol{\beta}})$, where the three variance estimates are derived from one-way clustering on the first dimension, the second dimension (the additional dimension we are interested in), and their intersection, respectively.

Both the first difference and the pairwise difference transformations are linear transformations of the data. Thus, robust linear *S*- or *MM*-estimation applied to such transformed data does not lose its equivariance properties. The first difference estimator has a BDP of (T-1)/4T for T>=3 where T is the number of periods. As such, first differencing approaches 25% BDP only for large T. In contrast, the BDP of the pairwise estimator is 25% for any T.

B. A Framework for Handling Outliers

Reproducibility is a fundamental requirement for empirical research. It is often impossible or difficult to replicate papers that do not carefully document how outliers are handled. Irreproducibility leads to questions about the validity of the research outcomes. For example, Adams, Hayunga, and Mansi (2018) find that outliers caused by data errors and comprising less than 2% of the original sample drive the Chen, Hong, Huang, and Kubik (2004) finding of mutual fund diseconomies of scale. Similarly, Adams, Hayunga, and Mansi (2019) revisit the nature of returns to scale following Pástor, Stambaugh, and Taylor (2015) and find that the documented negative relation between industry scale and return performance is an artifact of extreme observations that comprise less than 0.05% of the sample. Guthrie, Sokolowsky, and Wan (2012) report that two outliers out of a sample of 865 firms drive the Chhaochharia and Grinstein (2009) estimate on CEO pay decreases in noncompliant firms following the NYSE and NASDAQ revision of listing standards requiring majority independent boards.

We present a framework for handling outliers and improving empirical research replicability and robustness in Table III. The primary focus in this paper is the identification and mitigation of multivariate outliers. Table III includes commentary for each recommendation. We call attention to Items 1 and 2 that warn against winsorizing and univariate trimming, Items 7 and 11 for documenting data decisions, Item 8 for advising a formal outlier test, and the decision to mitigate beyond data errors in Items 12 and 13. We employ this framework in the replications that follow. However, in the interest of brevity and because univariate identification is commonly used, we largely restrict our discussion to the multivariate outliers (i.e., Items 8-13 in Table III).

IV. Replications

While multivariate outliers have the potential to significantly influence OLS estimated coefficients and standard errors, it does not necessarily follow they unduly affect statistical inference in practice. Outliers may not be a concern when they are not very influential or occur infrequently. If, however, outliers are common and influential enough to bias coefficient estimates, they merit extra consideration.

This section is divided into two subsections. First, we evaluate the incidence, characteristics, and influence of outliers by replicating a study from corporate finance (Peterson, 2009) and another from asset pricing (Wahal and Yavuz, 2013). In both replications, we collect the data and conduct the empirical analyses ourselves (i.e., no code or datasets were provided by the authors of the two studies). In addition, we replicate two published papers where the authors have provided their code and datasets using outlier robust regressions as a diagnostic tool.

Appendix B provides the variable definitions, data sources, and outlier methods for the four studies. The STATA code we use to generate the figures and tables in the replications is available at https://uta.box.com/v/FinancialManagement. The codes are fully documented to provide guidance for future researchers. We also provide the data we use for the corporate finance (Peterson, 2009) and another from the asset pricing (Wahal and Yavuz, 2013) applications. A STATA package including *MM*-estimators, outlier diagnostics, and the identification methods we use are available by typing, "net from http://homepages.ulb.ac.be/~vverardi/stata" from within STATA.

A. Incidence of Articles with Outlier Mention

Table I, Panel B provides the number and percentage of articles published annually from 1988-2017 in the top four finance journals mentioning outliers using OLS, as well as the combination of outliers and OLS. The data are collected from the EBSCO database for JF, RFS, and JFQA and the Science Direct database for JFE using keyword searches.[12] Three issues are noteworthy. First, as a percentage, over time, research papers increasingly mention outliers (7% in 1988 vs. 32% in 2017). In addition, the majority of recently published papers use OLS. Finally, most papers using OLS fail to mention potential outlier bias.

B. Outlier Influence in Corporate Finance and Asset Pricing

1. Peterson (2009) Study: Corporate Finance Application

Our first replication examines a typical specification found in the corporate finance literature. We follow Peterson (2009) and compute annual variables from the CRSP/Compustat merged database from 1973-2017. We begin by applying *S*-estimators to the following model separately for each of the years 1973-2017 to identify vertical outliers and bad leverage points.

$$\begin{aligned} \mathit{Debt\ Ratio}_{i,t} = {} & \beta_0 + \beta_1(\mathit{Ln}(\mathit{MV\ Assets}_{i,t})) + \beta_2(\mathit{Ln}(1+\mathit{Firm\ Age}_{i,t}) \\ & + \beta_3(\mathit{Ln}(1+\mathit{Firm\ Age}_{i,t}) + \beta_4(\mathit{Profits/Sales}_{i,t}) \\ & + \beta_5(\mathit{Tangible\ Assets}_{i,t}) + \beta_6(\mathit{Market\ to\ Book}_{i,t}) \\ & + \beta_7(\mathit{Advertising/Sales}_{i,t}) + \beta_8(\mathit{R\&D/Sales}_{i,t}) \\ & + \beta_9(\mathit{R\&D}>0\ (=1\ \mathit{if\ yes})_{i,t}) + \varepsilon_{i,t} \end{aligned} \qquad (2)$$

[12] We use keywords OLS and 'least squares' to capture the use of OLS and 'outlier(s), 'extreme value(s)' and 'extreme observation(s)' to identify papers mentioning outliers.

Panel A of Table IV reports the percentage of multivariate outliers (vertical and bad leverage points) occurring in each year. Overall, vertical outliers compromise about 2.6% of all observations. The data shows considerable variation in the annual incidence of outliers. Vertical outliers, for example, comprise only 0.6% of observations in 2003, but 8.3% in 1991. In terms of bad leverage points, the 44-year average annual incidence is 3.7%. Bad leverage point frequencies exhibit substantial variability, from a low of 2.4% in 1979, 2003, and 2005 to a high of 11.9% in 2016.

The results in Panel A provide context as to why winsorizing and trimming schemes are not consistently effective treatments for univariate outliers. First, winsorizing using the most common cutoff points is of limited value when outlier frequencies exceed the designated threshold. For example, 1% winsorizing is not likely to effectively mitigate outlier influence when outlier frequencies exceed 1%. Additionally, naïve winsorizing or trimming schemes that apply a uniform cutoff rule for all years run the risk of having too small a cutoff in some years and too large a cutoff in others. As a result, winsorizing and trimming could potentially exacerbate rather than mitigate outlier-induced biases.

Since multivariate outliers are observations whose dependent variable, in this case the debt ratio, is not consistent with the model's prediction, examining them carefully provides insights on model improvement. We begin by comparing outliers to typical non-outlier observations. Panel B of Table IV reports the mean (and median) values for non-outlier, vertical outlier, and bad leverage point observations, as well as mean (and median) difference testing results. The results report significant differences in non-outlier and outlier mean (median) values for many of the variables. Vertical outliers in Column 2 have a mean (median) debt ratio that is much larger than the mean (median) value for typical observations, a difference that is significant at the 1% level. Vertical outlier observations also have smaller market-to-book ratios and smaller research and development expenses. Bad leverage points in Column 3 also have higher market debt ratios, are younger, have more tangible assets, higher market-to-book ratios, and invest more in advertising and research and development.

Next, we visualize the outliers using the Rousseeuw and van Zomeren (1990) plots. Because the total dataset is large, Figure II plots observations occurring in 2017 only. Figure III illustrates the dataset containing numerous outliers and identifies four bad leverage point examples with internet service provider Windstream Holdings having the most extreme robust standardized residual (i.e., outlyingness in the dependent variable). The debt ratio for Windstream Holdings increased from around 40% in 2007 to over 80% in 2017. During this period, Windstream Holdings' stock price fell by more than 85% to less than $2.00. Windstream Holdings also reported a negative book value of equity in 2017. Figure II also shows specialty refiner Calumet Specialty Products as a bad leverage point. Further analysis reveals Calumet Specialty Products appears financially distressed with stock prices that fell about 75% and a book value of equity that fell over 80% from 2012-2017 period. Also during the same period, Calumet Specialty Products' debt ratio increased to over 65% from less than 30%. Ingles Markets is a supermarket chain operating in the southeastern U.S. Ingles Markets' debt ratio is about 50%, a value that is higher than the non-outlier sample average of less than 20%. Typical for the supermarket industry, profit margins are small (about 6% compared to over 20% for the sample) and tangible assets are high. Domino's Pizza is a high growth stock whose book and market values of equity were negative $2.7 billion and positive $8.1 billion, respectively, in 2017. Operating profits grew at an average annualized rate of about 15% and its stock price increased more than 200% from 2014-2017.

Of primary interest is whether multivariate outliers materially affect regression coefficients. We add year and firm fixed effects to the model presented in Equation 2. Panel C of Table IV reports OLS estimated coefficients with multiple winsorizing and trimming levels in Columns 1-5 and *MM*-robust regressions in Column 6 for the full sample period. Due to the large number of year and firm fixed effects, we use the panel data *MM*-robust regression extension developed in Section III. The results in Panel C report considerable variation in coefficient estimates and *t*-statistics across the six models. The largest differences are the loadings on Tangible Assets, Firm Age, and the Advertising/Sales Ratio. The estimated coefficients for Tangible Assets are significant in the five OLS models, but insignificant in the *MM*-robust regression suggesting the OLS results are driven by outliers. The Firm Age coefficients are positive and significant for two of the OLS models, positive but insignificant for three OLS models, but negative and significant for the *MM*-robust specification. Likewise, the Advertising/Sales Ratio OLS estimated coefficients are generally positive, but insignificant while the *MM*-robust coefficients are negative and statistically significant. The differences in the OLS and *MM*-robust results indicate the OLS results are driven by outliers.

Panel C also reports the Dehon et al. (2012) outlier test *p*-values for each OLS specification. The outlier tests reject the null hypothesis that OLS estimated coefficients are not materially influenced by outliers, which is consistent with Panel A of Table IV reporting that the incidence of outliers is large in certain years. Panel C also reports the maximum outlier robust efficiency for the *MM*-estimation is only about 31%, further evidence concerning the magnitude of the outlier bias problem for this model specification/sample combination. However, low efficiency in large data sets is less problematic than in small ones and this sample contains over 33,000 observations.

A few other determinants are statistically significant, but demonstrate altered economic inference in the robust model. The estimated coefficients on the market value of assets range from 200%-400% larger for the OLS specifications than the slope coefficient of 0.008 using the robust regression. The profit-to-sales ratio is insignificant using OLS in Model (1), but significantly negative for the remaining estimations. Again, the winsorizing and trimming estimates for the profit-to-sales ratio are considerably larger than the robust result. The market-to-book variable displays similar coefficients in the winsorized and trimmed results in Models 2 and 6, but these coefficients are different when compared to the untreated OLS results and the *MM*-robust models (by a factor of at least two). This finding highlights our concerns that univariate treatments, such as winsorizing and trimming, can exacerbate outlier influence.

Overall, there are considerable differences in the OLS and *MM*-robust regression results. Thus, we conclude influential outliers bias the reported OLS estimated coefficients. We identify several influential observations and find they are most often firms in financial distress and firms that increased leverage substantially in a short period. The most appropriate outlier mitigation strategy relies on the nature of the data, the hypothesized relation between debt ratios, and the independent variable of interest. For example, if our primary variable of interest is the Market-to-Book ratio, mitigation is not necessary so long as the hypothesized effect of size on leverage is a general one (i.e., as firms become more valued in the market relative to their book value they tend to become more levered) since outliers do not appear to be driving the untreated OLS estimated coefficients. In contrast, we see that outliers drive the significantly positive OLS estimated coefficients for Tangible Assets as the outlier robust estimated coefficient is insignificant. If theory predicts an outlier effect in that the average relation between tangible assets and leverage is determined by a handful of observations, mitigation should be limited to correcting influential outliers caused by bad data. Alternatively, if there is a hypothesized general effect, better inferences can be obtained by either dropping the influential outliers or performing outlier robust regressions.

2. Wahal and Yavuz's (2013) Study: Asset Pricing Application

Asset pricing studies commonly use Fama-MacBeth (1973) (FM) regressions in their analysis. However, Knez and Ready (1997) note that the OLS loss function employed in almost all applications of Fama-MacBeth (1973) is sensitive to both vertical outliers and bad leverage points. Knez and Ready (1997) develop robust Fama-MacBeth (1974) estimates by replacing the OLS loss function with least trimmed squares. They replicate Fama and French (1992, 1993) and find that the estimated relation between size and returns changes from negative when estimated using FM-OLS to positive when using outlier robust FM-LTS. Knez and Ready (1997) conclude the often-examined size effect in Banz (1981), where small firms outperform large firms, is driven by extreme observations accounting for as little as 1% of the data. Unfortunately, subsequent asset pricing studies appear to ignore robust estimation and instead winsorize or trim, often at the 1% level. One possible reason for the reluctance to adopt FM-LTS is that LTS estimates, while unbiased in that they fit most of the data, suffer from very low efficiency relative to OLS. Stromberg, Hössjer, and Hawkins (2000) find that LTS has a relative efficiency of only 7%.

We apply Knez and Ready's (1997) concept of robust Fama-MacBeth (1973) estimation, but instead use *MM* robust regressions that are more efficient than LTS. We replicate the models and sample generation process found in Wahal and Yavuz's (2013) examination of style investing's effect on return predictability. Specifically, we replicate Table 1 in the Wahal and Yavuz (2013) study, where the results from Fama-MacBeth (1973) regressions of monthly future stock returns on prior style and stock returns, book-to-market, and size are reported. In their discussion, Wahal and Yavuz (2013) focus on the prototypical six-month future return on six-month prior returns and, for brevity, we follow their example. The data cover 1973-2017. Stock return, book-to-market, and size (the market value of equity) data are from the merged CRSP/Compustat database. The model uses NYSE size breakpoints along with the full set of securities for book-to-market to obtain 5 x 5 size and book-to-market style portfolios. Wahal and Yavuz (2013) drop stocks with negative book-to-market values, while prior stock returns, book-to-market ratios, and size are winsorized at the 1% level in each month. We do not apply these univariate treatments to our primary sample.

Table V presents the results in two panels. Panel A provides the mean (and median) values for each variable segmented by observation type, as well as testing for the mean (and median) differences between the segmentations. Multivariate vertical outliers and bad leverage points are identified monthly via *S*-regression estimators. Panel B reports the coefficient estimates and Newey West *t*-statistics for the FM-OLS, FM-OLS with 1% winsorizing of all variables, FM-OLS with 1% trimming of all variables, and *MM*-robust Fama-MacBeth (FM-MM) specifications.

In terms of the dependent variable of future stock returns, Panel A reports the mean vertical and bad leverage values are about 10 and 25 times larger than the mean typical observation values. Mean size values are smaller for vertical outliers and leverage points as compared to the typical observation mean value. Another difference is that the vertical outliers have a mean prior stock return that is much smaller than the mean stock return for typical observations (-0.15% vs. 3.63%), while bad leverage points have a mean prior stock return that is a bit more than twice as large as the mean for typical observations. Also, prior six-month mean style returns are smaller for vertical outliers and larger for bad leverage points than for the typical observations.

Next, we investigate outlier frequency. Figures III and IV report the monthly percentage of vertical outliers and bad leverage points. Figure III illustrates vertical outliers occurring very frequently in some periods (e.g., mid 1970's) and not so much in others (1999-2000). The mean percentage of vertical outliers for all years is 1.8% with a minimum of 0.2% occurring in August 1999 and a maximum of 4.8% occurring in December 1974.

Similarly, Figure IV shows substantial monthly variation in bad leverage points. Bad leverage points occurred most frequently around 1996, 1999-2000, 2003-2004, and 2010. The high incidence of bad leverages points is cause for concern given that OLS and, by extension, FM-OLS, is especially sensitive to these types of outliers. The mean percentage of bad leverage points for all years is 6.4% with a minimum of 1.7% occurring in May 1981 and a maximum of 20.8% occurring in August 1999.

Panel B reports the estimated coefficients for four specifications, FM-OLS, FM-OLS with 1% winsorizing of all variables, FM-OLS with 1% trimming of all variables, and FM-*MM*-robust with 28.7% efficiency (*MM*-robust regression with 28.7% efficiency is equivalent to *S*-estimation). The estimated coefficients reported for the FM-OLS 1% winsorized specification in Column 2, with the exception of the Style return estimated coefficient, are not comparable to the six-month future return results reported in Panel A of Table 1 of the Wahal and Yavuz (2013) study. We suspect the discrepancies are due to their outlier mitigation approach (selective winsorizing and dropping) and differences in the sample beginning and end dates. Figures IV and V report the incidence of outliers varies considerably from month to month and year to year. Hence, different sample periods will contain different levels of outliers and FM-OLS estimates may be more biased for some sample periods and less biased for others. Also, we winsorize all of the variables and do not remove negative book-to-market ratio stocks (post winsorizing there are no negative book-to-market observations).

Panel B also reports the Dehon et al. (2012) outlier test *p*-value is 0.000, which indicates the FM-OLS monthly regression coefficients in Columns 1, 2, and 3 are biased and point to the necessity of robust regression, and the efficiency for the FM-*MM*-robust model is about 29%. The efficiency is a reflection of outliers in the data, but is still considerably higher than the 7% efficiency of the FM-LTS regressions in Knez and Ready (1997). The FM-MM Style return coefficient is about 15% larger than the FM-OLS estimate (10.32 vs. 8.96) and is significant at the 1% level in all specifications. However, because the estimated coefficients in the OLS and *MM*-robust regressions have the same sign and significance level, we conclude outliers are not driving the Prior Style Return empirical results. In contrast, the estimated coefficient for the prior stock return's estimated coefficient is insignificant for FM-OLS, marginally significant for FM-OLS with 1% winsorizing and FM-*MM*, and highly statistically significant when univariate outliers are trimmed at the 1% level (Column 3). Thus, we conclude the empirical support for stock level momentum at the six-month interval is fragile and influenced by outliers.

The estimated coefficients for size are negative and significant at the 1% level for the FM-OLS specifications and consistent with Asparouhova, Bessembinder, and Kalcheva (2013), Belo, Gala, and Li (2013), and Novy-Marx (2012). The FM-OLS results are also consistent with Fama and French (1992). In contrast, the FM robust estimated coefficient is positive and, while not quite statistically significant, is consistent with the Knez and Ready (1997) FM-LTS results. Thus, there is an outlier effect of size on future returns. For the overwhelming majority of stocks, there is a positive relation between size and future returns and not a negative relation as documented in most studies. Choosing the correct estimation method and mitigation strategy depends upon the hypothesized relation. If theory suggests an outlier effect (e.g., a few very large firms with low future returns or a few very small firms with high future returns), outlier mitigation beyond error correction is not desirable. Researchers will still want to know how many outliers are driving the result (e.g., 0.1%, 1%, 5%, or 10% of the sample). However, if theory predicts a general effect outlier mitigation, either dropping the most influential outliers or using outlier robust regressions will improve statistical inference. How outliers are treated in the empirical literature provides insight into the consensus as to whether the hypothesized relation is a general or an outlier effect. For example, in the stock momentum literature, it is common to winsorize. This suggests researchers in this area view momentum as a general effect and outlier mitigation beyond outlier correction will improve inference.

Overall, the capital structure and asset pricing replications demonstrate the practical importance of formal testing for multivariate outliers, identifying them, and potentially controlling for their influence. Table IV, using a capital structure setup, and Figures IV and V, focusing on asset pricing, indicate the incidence of outliers varies considerably from year to year and their occurrence is more frequent than implied by commonly used winsorizing and trimming levels. Figure III and Tables IV and V demonstrate that many of the vertical outliers and bad leverage points are extremely influential and point to the necessity of using outlier analysis for model improvement and potentially identifying omitted variables. Finally, the results suggest that multivariate outlier robust regressions often yield different coefficient estimates than OLS regardless as to winsorizing or trimming levels for both the capital structure and the asset pricing models.

C. Replications (Using Datasets/Code Provided by Authors)

Next, we replicate two recently published articles to further illustrate the practical importance of identifying and potentially mitigating outlier bias. These are Becker and Stromberg (2012) and Guthrie et al. (2012). We selected these papers for replication using a formal screening procedure. Of the studies in the top four journals in 2012 using OLS, there are 36 that use commercial databases available to us. We requested the data and code of a single regression model from the authors of these studies stating our interest in outlier investigation. Because these authors provided code and data, replications of these studies should bias against finding any outlier issues (i.e., authors concerned about outlier robustness likely declined to participate). These replications are not pre-selected to support a particular hypothesis, but instead we investigate all published articles where authors provide working code and data. Golubov, Petmezas, and Travlos (2012) and Panousi and Papanikolaou (2012) also provide their working code and datasets. Our outlier robust replications indicate significant changes in economic and statistical significance of the estimated coefficients in both studies. However, these changes are not large enough to unequivocally change the overall conclusions in the original papers. Due to space constraints, we do not report the results here.

1. Fixed Effects in Panel Data Models

In this section, we employ our newly developed panel data *MM* procedure to examine the prevalence of outlier bias by replicating Becker and Stromberg (2012) and Guthrie et al. (2012).

a. Becker and Stromberg (2012) Study

Becker and Stromberg (2012) examine the effect of managerial fiduciary duties on equity-debt conflicts using a 1991 legal ruling that changed corporate directors' fiduciary duties in Delaware firms. The 1991 ruling decrees that fiduciary duties are owed to all interested parties not only when a firm is insolvent (the pre-ruling standard), but also when it is in the "zone of insolvency." Consistent with their prediction that the ruling reduces affected equity holders' incentives to engage in risk shifting, Becker and Stromberg (2012) find a significant decrease in firm volatility following the ruling. Their unbalanced panel data set includes 2,145 observations for 745 Delaware firms and 653 non-Delaware incorporated firms.

Panel A of Table VI compares the Delaware (treatment group) and non-Delaware (control group) samples in terms of multivariate outlyingness for Becker and Stromberg's (2012) firm volatility results in Table 5 of their paper. Column 1 reports the mean values for Delaware firms, Column 2 presents the mean values for the non-Delaware firms, and Column 3 provides the mean differences. We also test whether the differences in the mean and median values between the two groups are significant. The objective is to assess the quality of the research design by determining whether there are significant differences in the incidence and magnitude of the multivariate outliers across the treatment and control groups. In terms of outlier classification types, the treatment and control groups have similar incidences of outliers. Likewise, there are no statistically significant differences in mean robust standardized residuals (outliers in the dependent variable space) across the Delaware and non-Delaware firms. Robust Mahalanobis distances are also similar across the two groups with none of the differences being statistically significant. However, we do note the difference in mean Mahalanobis distances for the Delaware and non-Delaware firms appears economically large.

Next, we examine the outlier detection plot in Figure V where the Delaware firms are represented with dots and the non-Delaware firms are represented with Xs. Overall and consistent with Table VI, Panel A, the treatment and control groups appear similar in terms of vertical, good, and bad leverage outliers. However, there appear to be differences in the extreme bad leverage outlier space (denoted by the letter "B"). For example, a control group firm has the largest Mahalanobis distance, while a treatment group firm has the largest robust standardized residual. More importantly, Figure VI demonstrates a large number of extreme bad leverage point outliers. Overall, Panel A and Figure V indicate the Delaware and non-Delaware firms are similar in their covariates and the quality of the research design is good.[13]

Panel B of Table VI reports the original OLS estimated coefficients from Becker and Stromberg's (2012) Table 5 in Column 1 and the outlier robust estimates in Column 2. The dependent variable in each model is the volatility of firm ROA. The estimated coefficient for the primary variable of interest, Delaware*Post-1991, is significant at the 5% level in the original OLS specification. In contrast, Model 2 reports the *MM*-estimated coefficient for Delaware*Post-1991 is economically and statistically insignificant. Because the OLS results differ from the outlier robust results, we conclude outliers drive the published relation of interest. We next examine what type of outliers, vertical or bad leverage point, are driving the OLS results. Column 3 reports the OLS estimated coefficient on Delaware*Post-1991 is significant at the 5% level after dropping the vertical outliers in Panel A and Figure V. In contrast, Column 4 indicates an insignificant Delaware*Post-1991 after dropping the bad leverage points.

[13] This approach can be extended to propensity score and other matching techniques to evaluate covariate balances across the treatment and control groups in a multivariate framework. The concern is that outliers may cause bias if they are more prevalent in either the treatment or the control group. How outliers affect design quality is an open question as some matching methods may drop outlier observations that do not have close peers.

We further confirm this in unreported analysis where we calculate percentiles of vertical and horizontal distances and rerun the original OLS regression. We find that less than 3% of the sample, 62 of 2,145 observations, are influential as they are responsible for the significant estimated coefficient for Delaware*Post-1991. Our identification of the influential outliers substantially reduces the number of observations subject to manual examination and the data cleaning costs from the full sample of 2,145 to 62. In the interest of brevity, we do not manually examine the influential outliers. However, we note these 62 influential outliers have much larger volatility of firm ROA (0.27 vs. 0.06), Tobin's Q (2.98 vs. 1.26), and two-year stock price change (0.32 vs. -0.02) than the non-influential observations. We also note that 38 of the influential outliers are Delaware firms and 24 are control (not Delaware) firms.

After correcting any data or omitted variable errors, any further mitigation depends upon the nature of the hypothesized relation. That is, does the 1991 legal ruling matter for all firms? If so, dropping the influential observations or using outlier robust regressions is appropriate. If not, and the ruling only matters for certain types of firms (as described above), further mitigation is not warranted since the influential outliers represent the manifestation of the 1991 legal ruling's effect (i.e., an outlier effect). In this case, identifying and describing the influential outliers provides new insights regarding the ruling's affects.

As for the control variables, the economic and statistical importance of several coefficient estimates varies considerably in the OLS and outlier robust estimations. For example, the OLS coefficient for Ln MV is economically and statistically significant, while the *MM*-estimate is insignificant. While maintaining statistical significance, the *MM*-estimate for Ln Assets is approximately one-sixth the size of the OLS estimate. In addition, the estimated OLS coefficient on Market Leverage is large and statistically significant, but the *MM*-estimated ROA coefficient is close to zero and insignificant.

b. Guthrie et al. (2012) Rebuttal to Chhaochharia and Grinstein (2009)

Bebchuk, Fried, and Walker (2002) argue that manager influence over boards of directors enables them to extract rents via compensation schemes that lower shareholder value (e.g., the managerial power hypothesis). If so, independent directors should be associated with better governance and lower CEO pay. Following the accounting scandals that led to the enactment of the Sarbanes-Oxley Act of 2002, the NYSE and NASDAQ began requiring majority independent director boards, as well as fully independent nominating and compensation committees. Chhaochharia and Grinstein (2009) examine the compliance status of firms prior to the NYSE and NASDAQ rule change and find that CEO pay decreases by about 17% more in non-compliant firms than in compliant firms. Chhaochharia and Grinstein's (2009) findings are consistent with the managerial power hypothesis in that non-independent directors appear to allow CEOs to extract rents in the form of higher pay.

In a subsequent study, using the data and methodology of Chhaochharia and Grinstein (2009, 2012), Guthrie et al. (2012) contend that the drop in CEO pay is primarily due to decreases for just two CEOs out of a sample of over 865 (12 firm-year observations from a 5,190 firm-year observation sample). After removing the two CEOs, the change in pay becomes economically and statistically insignificant. In a rejoinder, Chhaochharia and Grinstein (2012) extend the sample period and argue their results are robust to removing outliers, asymmetric winsorizing, and median regressions. However, our unreported simulations demonstrate that these attempts do not adequately address the multivariate outlier issue. In addition, Chhaochharia and Grinstein (2012) and Guthrie et al. (2012) focus on outliers in the dependent variable space (vertical outliers), but not outliers in the independent space (bad leverage points). This is an important consideration as multivariate bad leverage points can severely influence coefficient estimates. Moreover, in both cases, the authors neither conduct formal testing of the influence of outliers nor examine outlier robust regressions.[14]

Guthrie et al. (2012) use the actual sample created by Chhaochharia and Grinstein (2009) for their main findings. They also reconstruct the sample following Chhaochharia and Grinstein (2009). Guthrie et al. (2012) share with us their reconstructed sample, which we replicate in Table VII. The table provides the published results from Guthrie et al. (2012) and our outlier robust regressions. For the sake of brevity, we focus on the compensation committee models where Guthrie et al. (2012) reports statistically significant increases in CEO pay at noncompliant firms following the rule change. Panel A of Table VII reports the results of Model 5 in the Guthrie et al. (2012) study, Appendix A. Models 1-4 provide coefficients for four variations of Model 5. Models 1 and 2 include the two outliers that were excluded by Guthrie et al. (2012) and reports results using OLS and MM estimation. Models 3 and 4 repeat the specifications of Models 1 and 2, but exclude the two outliers. Panel B of Table VII repeats these variations for the Guthrie et al. (2012) Model 7 in Models 1-4. Guthrie et al.'s (2012) published Model 7 examines the effect of the rule change in firms with high and low concentrations of institutional ownership.[15] The dependent variable in all of the models is the natural log of CEO pay.

Figure V illustrates the outlier detection plot of the sample for Model 5 revealing several large outliers. Eight observations are particularly large vertical outliers (located below the lower

[14] In unreported simulations, we find that median regressions used by Chhaochharia and Grinstein (2009) and Guthrie et al. (2012) provide protection against vertical outliers, but are ineffective in mitigating bad leverage outlier influence.
[15] Guthrie et al. (2012) note the fragility of the results for institutional ownership in conjunction with compensation committee independence in their Footnote 21.

horizontal and to the left of the vertical boundaries). Of the eight, four are Kinder Morgan, two are Fossil, one is Gateway, and one is Apple. The large negative robust standardized residuals of these eight observations indicate CEO pay is much smaller than predicted by Model 5. Figure VI also identifies three bad leverage points with large negative robust standardized residuals (located below the lower horizontal and to the right of the vertical boundaries) and four bad leverage points with large positive robust standardized residuals. Figure V makes clear that the Apple and Fossil CEOs are not the only, or even the most extreme, outliers in the sample. This means that the Chhaochharia and Grinstein (2009) and Guthrie et al. (2012) results suffer from multivariate outlier induced bias. Next, we compare to the estimated coefficients from Guthrie et al.'s (2012) OLS regressions to outlier robust *MM*-estimate to determine the scope of the bias. The estimated coefficient on the main variable of interest, Noncompliant × After, changes in statistical significance and economic importance when the two outliers are removed for the OLS fixed effects regressions in Models 1 and 3. This change demonstrates the potential of a few outliers to affect inference in large datasets. In contrast, the MM estimated coefficients in Models 2 and 4 are very similar and statistically insignificant. More interestingly, the estimated coefficients on the main variable of interest in OLS Models 1 and 3, Noncompliant × High Inst Conc., are statistically significant at the 1% level, but are insignificant in MM Models 2 and 4 (Panel B).

In summary, the replication exercises demonstrate the importance of identifying and addressing multivariate outliers. We find evidence of outliers that deviate so much from other observations as to arouse suspicions about data quality, model specification, and the overall mechanisms generating the data. We also find small numbers of these outliers can drive published empirical results and commonly used mitigation techniques are not effective. The efforts of Chhaochharia and Grinstein (2009, 2012) and Guthrie et al. (2012) highlight that even when finance researchers attempt to control for outlier influence the remedies (i.e., identifying and treating outliers in a univariate rather than multivariate regression framework), they can be ineffective. In addition, the analysis of Becker and Stromberg (2012) indicates multivariate outliers also affect commonly used control variables including book-to-market ratios, market leverage, and other data from the Compustat and CRSP databases.

V. Conclusion

Upon examining research in premier finance journals, we find that the majority of studies employ OLS regression as the primary statistical inference technique. When the assumptions of the OLS regression are met, OLS estimates provide a precise summary of relations in the data. However, these assumptions are simplifications that do not necessarily reflect financial or economic reality. One assumption, in particular, that the observed data have a normal distribution is problematic in finance datasets in the presence of unusual (or extreme) observations. These outliers occurring far from the majority of the data can bias OLS coefficient estimates. To remedy this problem, most research efforts attempt to make the data appear normal by altering the characteristics of outliers using univariate identification (e.g., winsorizing, trimming, or removing them altogether using dropping or filtering before applying OLS regression). The problem with these alteration schemes is that they fundamentally change the data thereby introducing new inference limitations.

In this paper, we examine the limitations associated with the use of univariate identification in finance research to remedy the multivariate outlier problem. We propose an identification strategy for multivariate outliers and find this method effectively detects outliers. We then develop a robust regression method that minimizes the bias outliers caused in both cross-sectional and panel regressions. Specifically, we use a combination of base robust estimators (*MM*-estimators) as described in Yohai (1987). To the best of our knowledge, there are no readily available procedures that compute *MM*-estimators or any other high breakdown point estimators with clustered standard errors. We rely on the theory of the generalized method of moments to calculate these clustered standard errors. This method also provides improvements that address fixed effects in cross-sectional and panel regressions. Empirically, we employ this method as a diagnostic tool using replications of four recently published studies in the finance journals to demonstrate how adjusting for multivariate outliers can lead to significantly different results.

The infrequent use of methods in finance research to reliably identify multivariate outliers and, when appropriate, remedy the bias they cause suggests there are considerable impediments preventing their widespread use. One obstacle, in particular, is the lack of readily available methods for the types of models and data structures encountered in the finance field. There also appears to be a misplaced belief that common univariate outlier mitigation techniques provide protection against extreme multivariate observations. We help reduce these limitations by proposing a methodology to identify and treat multivariate outliers in the finance field.

Finally, we wish to be clear. We do not advocate simply removing outliers, but to find them and then decide whether to keep, correct, delete, or mitigate them is the most appropriate path.

Appendix A. Robust Estimators

A.1 Estimator Criteria

The ideal estimator efficiently provides precise (i.e., unbiased) coefficient estimates, but there is a trade-off between efficiency and precision. Consider sample $\mathbf{Z}^{(n)} = \{Z_1,\cdots,Z_n\}$, with n observations. Let $T(\mathbf{Z}^{(n)})$ represent an estimator for the parameter θ. Applying T to $\mathbf{Z}^{(n)}$ provides the estimate of the population parameter such that $T(\mathbf{Z}^{(n)}) = \hat{\theta}$. The estimator is unbiased if $\mathbb{E}\left[T(\mathbf{Z}^{(n)})\right] = \mathbb{E}(\hat{\theta}) = \theta$. It follows then that the bias of an estimator is given by:

$$\text{bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta) = \mathbb{E}(\hat{\theta}) - \theta \tag{A.1}$$

To provide unambiguous statistical inference, an estimator must converge to the population parameter θ and the variance approach zero as the sample size grows. A good method for accounting for bias and variance is to measure the mean squared error (MSE). The MSE of parameter $\hat{\theta}$ reduces to $\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] = \mathbb{V}(\hat{\theta}) + \left[\text{bias}(\hat{\theta})\right]^2$ and the desired estimator is one where:

$$\lim_{n\to\infty} \text{MSE}(\hat{\theta}) = 0. \tag{A.2}$$

Another way to account for bias and variance is efficiency. In the strictest case, an estimator's efficiency is the ratio of its minimum possible variance to the actual variance. More practically, an estimator is considered efficient if its sampling variance is relatively small. This leads to small standard errors. Since certain estimators are more efficient than others are, we use relative efficiency defined as:

$$\text{Efficiency}(T_1,T_2) = \mathbb{E}\left[(T_2-\theta)^2\right]/\mathbb{E}\left[(T_1-\theta)^2\right] \tag{A.3}$$

When considering outliers, Hampel (1974) introduces two criteria for evaluating an estimator's robustness to extreme observations. These are the influence function (IF) and the breakdown point (BDP). The IF is a measure of the dependence of the estimator on the value of a single sample observation y_1 on the theoretical distribution F. The IF for estimator T is:

$$IF(y_1,F,T) = \lim_{\lambda\to 0}\frac{T\left[(1-\lambda)F+\lambda\Delta_{y_1}\right]-T(F)}{\lambda} \tag{A.4}$$

where Δ_{y_1} is the cdf of the point mass distribution at y_1 and F is the cdf of the uncontaminated data generating process of Z_i for all i.[16] λ gives the proportion of contamination at y. The OLS estimator has an unbounded IF suggesting that the influence of a single outlier on the coefficient estimate grows steadily as that observation becomes more extreme. Practically speaking, the unbounded IF of OLS infers that just one outlier can severely bias the OLS slope coefficients, even in larger samples. A more robust estimator will have a bounded IF. As such, an outlier does not unduly influence its coefficient estimates.

While IF measures resistance to individual or local influential observations, the global measure of resistance is the BDP. BDP is the smallest percentage of outliers in a sample that the estimator can handle without producing arbitrary results. Following Andersen (2008), consider the replacement of m observations in the dataset with observations that do not fit the general

[16] For this definition, sample Z_1,... Z_n, is assumed to be independent and identically distributed.

trend in the data for all possible corrupted samples $\mathbf{Z}^{(n)*}$. The maximum effect from these substitutions is:

$$\text{effect}(m;T,\mathbf{Z}^{(n)}) = \sup_{\mathbf{Z}^{(n)*}} \left\| T(\mathbf{Z}^{(n)\star}) - T(\mathbf{Z}^{(n)}) \right\| \tag{A.5}$$

where the supremum is over all possible $\mathbf{Z}^{(n)\star}$. The estimator breaks down if the effect $(m;T,\mathbf{Z}^{(n)}$) is infinite and the m outliers have an arbitrarily large impact on T.

Fifty percent is the highest acceptable BDP indicating that the estimator withstands a contamination of up to half of the dataset. A BDP higher than 50% is nonsensical as it implies over half of the sample is not representative of the overall sample. Hampel, Ronchetti, Rousseeuw, and Stahel (1986) argue that the BDP should be at least 10%. OLS has *BDP* = 0.

Next, we provide a background of the robust estimators using the estimator metrics and Equations (1)-(5) above. We also discuss potentially ineffective methods that we find finance researchers consider useful for outlier mitigation (e.g., median regression). The parametric estimators are *L*-estimators, *R*-estimators, *M*-estimators, and *S*-estimators.[17] In most cases, the current preferred techniques use combinations of these base estimators.

A.2 *L*-estimators

Estimators that are linear combinations of the order statistics are *L*-estimators. A special class of *L*-estimators is the least power, L_p, estimators. These estimators result from minimizing the sum of the absolute values of the errors raised to the power of p, where p is usually between

[17] Non-parametric estimation, such as artificial neural networks or kernel estimation, can be found in Kennedy, 2001, pp. 302-303.

one and two.[18] The general form minimizes $\sum_i \left|y_i - \mathbf{x}_i' \boldsymbol{\beta}\right|^p$. Note that OLS is the case when p = 2. When p = 1, the estimator minimizes the sum of the absolute errors. This specification has many names and predates OLS by about 50 years. The most common designation for this specification is the least absolute values (LAV) estimator.[19] While the LAV regression is resistant to some specific types of outliers, it is not a good robust estimator as the *BDP* = 0. It is resistant to vertical outliers, but potentially breaks down in the case of presence of a single bad leverage point. LAV also exhibits low efficiency.

Another common type of *L*-estimator is the regression quantile. A regression quantile is an estimate of a coefficient that results from minimizing a weighted sum of the absolute values of the errors, with positive errors possibly weighted differently than negative errors. The objective function minimizes:

$$\sum_{i=1}^{n} \rho_{\alpha}(e_i),$$

where

$$\rho_{\alpha}(e_i) = \begin{cases} \alpha e_i & \text{if } e_i \geq 0 \\ (\alpha - 1)e_i & \text{if } e_i < 0 \end{cases}$$

and α is the order of the quantile to estimate. Thus, the αth regression quantile is the coefficient estimate that results from minimizing the weighted sum of the absolute values of the errors. For

[18] Harvey (1978) provides a simple proof of unbiasedness for $1 < \rho < \infty$.
[19] Other estimator names are least absolute deviations (LAD), least absolute residual (LAR), least absolute error (LAE), and the minimum absolute deviation (MAD).

instance, the 0.25 regression quantile places 0.25 weight on the positive errors and 0.75 on the negative errors. Edgeworth (1887) introduces the median regression estimator, L_1, where positive and negative errors are weighted equally ($\alpha = 0.5$).

Koenker and Bassett (1978) note that trimming or trimmed least squares (TLS) can be thought of as an extension of the quantile regression.[20] To obtain the TLS estimator, the econometrician first computes $\hat{\boldsymbol{\beta}}(\alpha)$ and $\hat{\boldsymbol{\beta}}(1-\alpha)$ where α is the desired trimming proportion $(0 < \alpha < 0.5)$. Then, observations where $y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}(\alpha) \leq 0$ or $y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}(1-\alpha) \geq 0$ are dropped and least squares is computed on the remaining observations.

The problem with all quantile estimators is that the *BDP* = 0. While quantile regressions mitigate bias from vertical outliers, they do not protect against bad leverage points. Quantile estimators also suffer from low efficiency relative to OLS when error terms are distributed normally (Huber, 1981).

Rousseeuw (1984) develops two other *L*-estimators: least median of squares (LMS) and least trimmed squares (LTS). LMS replaces the summing of the square errors in OLS with the median of squared residuals. The estimates equation is:

$$\min M\left(y_i - \mathbf{x}_i' \boldsymbol{\beta}\right)^2 = \min M\left(e_i^2\right), \tag{A.6}$$

Where *M* is the median. Since the base median location parameter is more resistant, the estimator and regression model are also resistant to influential observations. While the estimator has a BDP = 0.5, LMS does not have a well-defined influence function (Rousseeuw and Croux, 1993) and has a slow convergence rate.

[20] Chan and Lakonishok (1992) introduce the trimmed regression quantile estimators.

The LTS estimator, which should not be confused with TLS, minimizes the sum of the squared residuals such that $\min \sum_{i=1}^{q} e_{(i)}^2$, where $q = [n(1-\alpha) + 1]$ is the number of subsample observations used to calculate the estimator, and α is the proportion of trimming. Algorithms are employed to find the subsample that yields the minimum sum of the squared residuals. Using $q = n/2 + 1$ yields a *BDP* = 0.5. However, Stromberg et al. (2000) find that LTS has a relative efficiency of only 7%. This excludes its use in most instances, but the LTS estimator can be useful as a first stage in a multi-step robust regression technique.

A.3 *R*-estimators

Jaeckel (1972) proposes a set of estimators that use dispersion measures based on the linear combinations of ordered residuals. The *R*-estimators are scale equivariant, which is advantageous over the *M*-estimators discussed next. However, the *BDP* = 0 and there are issues with the intercept and determining a score function necessary for use. Accordingly, we do not recommend *R*-estimators and refer the interested reader to Davis and McKean (1993) and Huber (2004).

A.4 *M*-estimators

Huber (1964) generalized the median regression to a wider class of *M*-estimators by considering functions other than the absolute value of the residuals. The maximum-likelihood type or *M*-estimator minimizes the sum of a less rapidly increasing loss function of the errors such that:

$$\sum_{=1}^{n} \rho(y_i - \mathbf{x}_i') = \min \sum_{i=1}^{n} \rho(e_i)\,. \qquad \text{(A.7)}$$

The concept is to use weights that do not continue to grow in magnitude as the absolute value of the error term grows. Equation (A-7) is not scale equivariant, so the errors must be standardized by a robust estimate of their scale $\hat{\sigma}_e$, which is also estimated on the data as:

$$\min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\sigma_e}\right). \qquad \text{(A.8)}$$

M-estimation requires an iterative procedure. Iteratively reweighted least squares, also known as Iterative weighted least squares, is the standard process to find *M*-estimates. The steps are: 1) fit an OLS model to the data to obtain the initial $\hat{\boldsymbol{\beta}}$, 2) use the errors to calculate the initial estimates for the weights, 3) the analyst then choses a weight function and applies it to the initial OLS residuals to create preliminary weights, and (4) the analyst uses weighted least squares (WLS) to minimize $\sum w_i e_i^2$ and obtain updated $\hat{\boldsymbol{\beta}}$. This is the standard solution in matrix form of $\hat{\boldsymbol{\beta}}_{\mathbf{1}} = (\mathbf{X}' \ \mathbf{W}\mathbf{X})^{-1}\mathbf{X}' \ \mathbf{W}\mathbf{y}$, where *W* is a (*N* x *N*) diagonal matrix of individual weights. The process continues by using residuals from the WLS model to calculate new weights to use in a new iteration of the WLS, and this is repeated until the $\hat{\boldsymbol{\beta}}'s$ converge. In practice, the finance researcher need not manually code this process as typical software, such as SAS and Stata, include the routines.

This *M*-estimator is efficient and an improvement with respect to outliers, but it is not reliably robust to bad leverage points.[21] This is because the iteratively reweighted OLS algorithm is only guaranteed to converge to the global minimum for monotone *M*-estimators and monotone *M*-estimators are not robust to bad leverage points.[22] This shortcoming is not limited to this particular method of *M*-estimation. In general, all methods of computing *M*-estimators suffer from the global minimum problem or an inability to identify all leverage points when outliers are clustered (Rousseeuw and van Zomeren 1990). Because of this issue ,*M*-estimators are usually combined with other robust estimators in a multi-step process.

A.5 *S*-Estimators

Rousseeuw and Yohai (1984) develop *S*-estimators that seek to minimize a measure of residual dispersion that is less sensitive to outliers than variance. The *S*-estimators are the solution with the smallest possible dispersion of the residuals:

$$\min \hat{\sigma}^S\left[e_1(\hat{\boldsymbol{\beta}}),\cdots,e_n(\hat{\boldsymbol{\beta}})\right]. \tag{A.9}$$

Note that OLS is a special less robust case of *S*-estimators. OLS minimizes the variance of the residuals. The problem can be seen as looking for the smallest σ that satisfies the equality $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{e_i}{\sigma_e}\right)^2 = 1$, which is the definition of the variance. *S*-estimation replaces the square in the variance with another loss function, ρ_0, that awards less importance to large residuals. Specifically, *S*-estimation minimizes a robust *M*-estimate of the residual scale:

---

[21] Theoretically there are some robust *M*-estimators that estimate σ and β simultaneously but these cannot be fit using the IRWLS algorithm.
[22] M-estimators are monotone if *p* is convex over the entire domain.

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{e_i}{\hat{\sigma}^S}\right) = b\,, \tag{A.10}$$

where b is a constant defined as $b = \mathbb{E}_{\Phi}[\rho_0(e)]$ and Φ is the standard normal distribution. The value of β that minimizes $\hat{\sigma}^S$ is the *S*-estimator. While *S*-estimators address the low breakdown point with *BDP* = 0.5, it comes at a cost of low efficiency compared to OLS that Croux, Rousseeuw, and Hössjer (1994) indicate is approximately 30%. Consequently, the benefits of *S*-estimators are commonly combined with the efficiency characteristic of *M*-estimation to compute *MM*-estimators.

A.6 Choosing an Estimator

The ideal estimator is both efficient and robust to outliers. There is a tradeoff as highly efficient estimators are often not robust to outliers and robust estimators tend to be less efficient when there are no outliers. The issue with low efficiency is that inference is problematic if the errors are normally (or nearly so) distributed. As such, selecting the best estimator requires an examination of the sample. OLS is highly efficient, but has a *BDP* = 0 and outliers can have unbounded influence (i.e., parameter estimate bias increases with outlier size). LMS and LTS have a high BDP and can identify outliers, but are not a practical estimation choice since standard errors are not available and the results are not stable in larger samples. S-estimation is robust to outliers with a *BDP* = 0.5 and is more efficient than LMS, but still much less efficient than OLS. *M*-estimates have poor resistance to outliers and are less efficient than OLS.

In contrast, the *MM*-estimator has high outlier resistance of up to *BDP* = 0.5 and a bounded influence with respect to outliers.[23] In addition, *MM*-estimation can be nearly as efficient as OLS. This suggests that *MM*-estimators can provide coefficient estimates with less bias than OLS when datasets contain outliers and coefficient estimates that are similar to those provided by OLS in datasets without outliers. Dehon et al. (2012) follow the logic of Hausman (1978) to develop a testing procedure that compares estimates from outlier robust estimators and OLS. When the test fails to reject OLS, there are no significant outliers and the more efficient OLS is the best estimator. In cases where the test rejects OLS, the next step is to perform a second test that compares the robust, but less efficient *S*-estimator to a more efficient *MM*-estimator. The highest possible efficiency for the *MM*-estimator is determined via repeated testing. Thus, OLS, *S*-, or *MM*-regressions can be the most appropriate estimator depending upon the severity of the outlier problem.

Alternatively, researchers can use *S*- and *MM*-estimations to identify outliers and then drop some or all prior to implementing OLS. This approach has the advantage of using the more efficient OLS, but information from the dropped observations is lost. This approach typically yields similar results to the *MM*-estimation. At the very least, we recommend researchers compare estimated coefficients from OLS to coefficients from *S*- or *MM*-estimators to ease concerns that outliers are biasing estimates. STATA packages to implement the robust methods used in this paper's replications are available by typing, 'net from http://homepages.ulb.ac.be/~vverardi/ stata' from within STATA.

[23] There are multiple loss functions available to compute *MM*-estimators.

Appendix B. Variable Definitions

This table provides variable definitions, data sources, and outlier mitigation efforts for the six samples used in the outlier analysis. The data definitions and outlier mitigation details are obtained from the replicated articles.

Variable	Definition (Data Source)	Winsorize (W), Trim (T), or Drop (D) Levels in OLS Regressions
Capital Structure Application (Table IV)		
Debt Ratio	Book value of debt divided by the sum of the book value of assets minus the book value of equity plus the market value of equity (Compustat).	None
Ln (MV Assets)	Log of the sum of the book value of assets and the market value of equity minus the book value of equity (Compustat).	None
Ln (1+ Firm Age)	Log of one + Firm Age where Firm Age is difference years between the current year and the first trading date (Compustat).	None
Profit/Sales	Ratio of operating income before depreciation to sales (Compustat).	W & T @ 0%,1%,2.5%
Tangible Assets	Ratio of property, plant, and equipment to the book value of total assets (Compustat).	W & T @ 0%,1%,2.5%
Market to Book	Ratio of the market value of assets to the book value of total assets (Compustat).	W & T @ 0%,1%,2.5%
Advertising/Sales	Ratio of advertising expenses to sales (Compustat).	W & T @ 0%,1%,2.5%
R&D/Sales	Ratio of research and development expenses to sales (Compustat).	W & T @ 0%,1%,2.5%
R&D Dummy	Dummy variable that is equal to one when the research and development expense is positive and zero otherwise (Compustat).	None
Asset Pricing Application (Table V)		
Future Stock Return	Six month geometric stock return beginning in month *t*+1 (CRSP).	None
Prior Style Return	Six month geometric value-weighted return on a style portfolio constructed using the intersection of NYSE, Amex, and NASDAQ size and book-to-market quintiles for months *t*-5 through *t*=0 (CRSP).	None
Prior Stock Return	Six month geometric stock return for months *t*-5 through *t*=0 (CRSP).	W @ 1%
Log Size	Log of the market value of equity (Compustat).	W @ 1%
Log BM	Log of the ratio of the book value of equity to the market value of equity (Compustat).	Negative book value of equity stocks dropped and then W @ 1%
Becker and Stromberg (2012)- Equity-debtholder conflicts (Table VI)		
ROA Volatility	Standard deviation of the previous eight quarterly changes in the return on assets. (Compustat via authors).	T if outside [0,1]
Delaware*Post-1991	Dummy variable: one if the firm is incorporated in Delaware and the observation year is post 1991 (Compustat via authors).	None
Post 1991	Dummy variable: one for years after 1991 and zero otherwise (Compustat via authors).	None
Return on Assets	EBITDA divided by total assets (Compustat via authors).	T if outside [-.5,5]
Return on Sales	EBITDA divided by sales (Compustat via authors).	T if outside [-1,1]
Ln Assets	Log of the book value of assets (Compustat via authors).	None
Ln Sales	Log of sales (Compustat via authors).	None
Ln MV	Log of the market value of equity (Compustat via authors).	None
Depreciation/Assets	Depreciation divided by the book value of assets (Compustat via authors).	T if outside [0,.3]
Book Leverage	Assets minus common equity (book value) and minus tax liabilities divided by assets (Compustat via authors).	T if outside [0,1]

Market Leverage	Assets minus common equity (book value) and minus tax liabilities divided by assets minus common equity (book value) and minus tax liabilities plus the market value of equity (Compustat via authors).	T if outside [0,1]
Q	Assets minus common equity (book value) plus the market value of equity minus tax liabilities divided by assets minus 0.1 times the common equity (book value) and plus 0.1 times the market value of equity (limits q to a maximum value of 10) (Compustat via authors).	None
Two-Year Stock Price Change	Two-year log change in stock price (CRSP via authors).	
Guthrie, Sokolowsky, and Wan (2012)- CEO Compensation (Table VII)		
CEO pay	Log of CEO pay (Execucomp via authors).	None
Noncompliant	Binary variable that takes a value of one if the firm did not have a majority of independent directors on the board in 2002 and zero otherwise (IRRC via authors).	None
Before/After	Period indicators taking a value of one if the observation is in the premandate (before) period (2000–2002) or postmandate (after) period (2003 to 2005) and zero otherwise (via authors).	None
High/Low Inst Conc	Binary variable that takes a value of one if a firm's institutional ownership concentration falls into the top quartile (high) or bottom quartile (low) (Thomson Financials 13F database via authors).	None
Sales	Log of sales (Compustat via authors).	None
ROA	Log of one plus net income before extraordinary items scaled by the book value of assets (Compustat via authors).	None
RET	Log of one plus the annual stock return (dividends reinvested) in the prior year (CRSP via authors).	None
Tenure	Log of one plus the number of years the CEO served in the firm (Execucomp via authors).	None

References

Adams, J., D. Hayunga, and S. Mansi, 2018, "Diseconomies of Scale in the Actively-Managed Mutual Fund Industry: What Do the Outliers in the Data Tell Us?" *Critical Finance Review* 7(2), 1-48.

Andersen, R., 2008, *Modern Methods for Robust Regression*, No. 152, Los Angeles, CA, Sage.

Angrist, J. and J. Pischke, 2010, "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con Out of Econometrics," *Journal of Economic Perspectives* 24, 330.

Anscombe, F.J., 1973, "Graphs in Statistical Analysis," *American Statistician* 27(1), 17-21.

Aquaro, M. and P. Čížek, 2013, "One-step Robust Estimation of Fixed-effects Panel Data Models," *Computational Statistics and Data Analysis* 57, 536-548.

Asparouhova, E., H. Bessembinder, and I. Kalcheva, 2013, "Noisy Prices and Inference Regarding Returns," *Journal of Finance* 68, 665-714.

Banz, R., 1981, "The Relationship Between Return and Market Value of Common Stock," *Journal of Financial Economics* 9, 3-18.

Barro, R.J., 2006, "Rare Disasters and Asset Markets in the Twentieth Century," *The Quarterly Journal of Economics* 121(3), 823-866.

Bebchuk, L., J. Fried, and D. Walker, 2002, "Managerial Power and Rent Extraction in the Design of Executive Compensation," *University of Chicago Law Review* 69, 751-846.

Becker, B. and P. Stromberg, 2012, "Fiduciary Duties and Equity-debtholder Conflicts," *Review of Financial Studies* 25, 1931-1969.

Belo, F., V. Gala, and J. Li, 2013, "Government Spending, Political Cycles, and the Cross Section of Stock Returns," *Journal of Financial Economics* 107, 305-324.

Belsley, D., E. Kuh, and R. Welsch, 1980, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York, NY: John Wiley.

Bollinger, C.R. and A. Chandra, 2005, "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data," *Journal of Labor Economics* 23(2), 235-257.

Bowen, D., L. Frésard, and J. Taillard, 2017, "What's Your Identification Strategy? Innovation in Corporate Finance Research," *Management Science* 63(8), August 2529-2548.

Bramati, M. and C. Croux, 2007, "Robust Estimators for the Fixed Effects Panel Data Model," *Econometrics Journal* 10, 521-540.

Chan, L., Lakonishok, J., 1992. Robust measurement of beta risk, Journal of Financial and Quantitative Analysis 27, 265–282.

Chhaochharia, V. and Y. Grinstein, 2009, "CEO Compensation and Board Structure," *Journal of Finance* 64, 231-261.

Chhaochharia, V. and Y. Grinstein, 2012, "CEO Compensation and Board Structure – There is an Effect After All," *Journal of Finance*, Comments and Rejoinders.

Chen, J., H. Hong, M. Huang, and J.D. Kubik, 2004, "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization," *American Economic Review* 94(5), 1276-1302.

Croux C., G. Dhaene, and D. Hoorelbeke, 2008, "Robust Standard Errors for Robust Estimators," Catholic University of Leuven Manuscript.

Croux, C., P. Rousseeuw, and O. Hössjer, 1994, "Generalized S-estimators," *Journal of American Statistical Association* 89, 1271-1281.

Davis, J. and J. McKean, 1993, "Rank-based Methods for Multivariate Linear Models," *Journal of the American Statistical Association* 88, 245-251.

Dehon, C., M. Gassner, and V. Verardi, 2012, "Extending the Hausman Test to Check for the Presence of Outliers," *Advances in Econometrics* 29, 435-453.

Dittmar, A. and R. Duchin, 2016, "Looking in the Review Mirror: The Effect of Managers' Professional Experience on Corporate Financial Policy," *Review of Financial Studies* 29, 565-602.

Edgeworth, F., 1887, "On Observations Relating to Several Quantities," *Hermathena* 6, 279-285.

Fama, E. and K. French, 1992, "The Cross-section of Expected Stock Returns," *Journal of Finance* 47, 427-465.

Fama, E. and K. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.

Fama, E. and J. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy* 81, 607-636.

Golubov, A., D. Petmezas, and N. Travlos, 2012, "When it Pays to Pay Your Investment Banker: New Evidence on the Role of Financial Advisors in M&As," *Journal of Finance* 67, 271-311.

Guthrie, K., J. Sokolowsky, and K. Wan, 2012, "CEO Compensation and Board Structure Revisited," *Journal of Finance* 67, 1149-1168.

Hampel, F., 1974, "The Influence Curve and Its Role in Robust Estimation," *Journal of American Statistical Association* 69, 383-393.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel, 1986, *Robust Statistics. The Approach Based on Influence*, New York, NY, Wiley.

Harvey, A., 1978, "On the Unbiasedness of Robust Regression Estimators," *Communications in Statistics* 7, 779-783.

Hausman, J., 1978, "Specification Tests in Econometrics," *Econometrica* 46, 1251-1271.

Hawkins, D., 1980, *Identification of Outliers*, London, UK, Chapman and Hall.

Heckman, J., 1979, "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153-161.

Henry, T. and J. Koski, 2017, "Ex-Dividend Profitability and Institutional Trading Skill," *Journal of Finance* 72, 461-494.

Huber, P., 1964, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics* 35, 73-101.

Huber, P., 1981, *Robust Statistics*, New York, NY, Wiley.

Huber, P., 2004, *Robust Statistics*, New York, NY, Wiley.

Jaeckel, L., 1972, "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals," *Annals of Mathematical Statistics* 43, 1449-1458.

Kennedy, P., 2001, *A Guide to Econometrics*, Cambridge, MA, MIT Press.

Knez, P. and M. Ready, 1997, "On the Robustness of Size and Book-to-market in Cross-sectional Regressions," *Journal of Finance* 52, 1355-1382.

Koenker, R. and G. Bassett, Jr., 1978, "Regression Quantiles," *Econometrica* 46, 33-50.

Maronna, R., D. Martin, and V. Yohai, 2006, *Robust Statistics, Theory and Methods*, New York, NY, John Wiley and Sons.

Maronna, R. and V. Yohai, 2000, "Robust Regression with Both Continuous and Categorical Predictors," *Journal of Statistical Planning and Inference* 89, 197-214.

Moeller, S.B., F. Schlingemann, and R. Stulz, 2005, "Wealth Destruction on a Massive Scale? A Study of Acquiring-Firm Returns in the Recent Merger Wave," *Journal of Finance* 60(2), 757-782.

Novy-Marx, R., 2012, "Is Momentum Really Momentum," *Journal of Financial Economics* 103 429-453.

Panousi, V. and D. Papanikolaou, 2012, "Investment, Idiosyncratic Risk, and Ownership," *Journal of Finance* 67, 1113-1148.

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

Rietz, T.A., 1988, "The Equity Risk Premium a Solution," *Journal of Monetary Economics* 22(1), 117-131.

Roberts, M. and T. Whited, 2013, "Endogeneity in Empirical Corporate Finance," in G. Constantinides, R. Stulz, and M. Harris, *Handbook of the Economics of Finance*, Vol 2, Part A. Amsterdam, Netherlands, Elsevier, 493-572.

Rousseeuw, P., 1984, "Least Median of Squares Regression," *Journal of the American Statistical Association* 79, 871-880.

Rousseeuw, P. and C. Croux, 1993, "Alternative to the Median Absolute Deviation," *Journal of American Statistical Association* 88, 1273-1283.

Rousseeuw, P. and B. van Zomeren, 1990, "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association* 85, 633-639.

Rousseeuw, P. and V. Yohai, 1984, "Robust Regression by Means of S-estimators," *Nonlinear Time Series Analysis: Lecture Notes in Statistics* 26, 256-272.

Salibian-Barrera, M. and V. Yohai, 2006, "A Fast Algorithm for S-regression Estimates," *Journal of Computational and Graphical Statistics* 15, 414-427.

Stromberg, A., O. Hössjer, and D. Hawkins, 2000, "The Least Trimmed Difference Regression Estimator and Alternatives," *Journal of the American Statistical Association* 95, 853-864.

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

Tukey J., 1991, "Graphic Displays for Alternative Regression Fits," in W. Stahel and S. Weisberg, Ed., *Direction in Robust Statistics and Diagnostics*, Part 2, New York, NY, Springer-Verlag.

Verardi, V. and C. Croux, 2009, "Robust Regression in Stata," *The Stata Journal* 9, 439-453.

Verardi, V. and J. Wagner, 2011, "Robust Estimation of Linear Fixed Effects Panel Data Models with an Application to the Exporter Productivity Premium," *Jahrbucher f. Nationalokonomie u. Statistik*, 231(4), 546-557.

Wahal, S. and M. Yavuz, 2013, "Style Investing, Comovement and Return Predictability," *Journal of Financial Economics* 107, 136-154.

Welch, I., 2016. "The (Time-Varying) Importance of Disaster Risk," *Financial Analysts Journal* 72(5), 14-30.

Yohai, V., 1987, "High Breakdown-point and High Efficiency Robust Estimates for Regression," *The Annals of Statistics* 15, 642-656.

Zellner, A., 1981, "Philosophy and Objectives of Econometrics," in D.A. Currie, A.R. Nobay, and D. Peel, Eds., *Macroeconomic Analysis: Essays in Macroeconomics and Econometrics*, London, UK, Croom Helm.

Figure I. Outlier Types

The figure illustrates the three types of outliers. Vertical outliers are extreme observations in the dependent variable. Good leverage points are extreme in both the dependent and the independent variables space, but fall on or near the regression line. Bad leverage points are outlying in the independent variables.

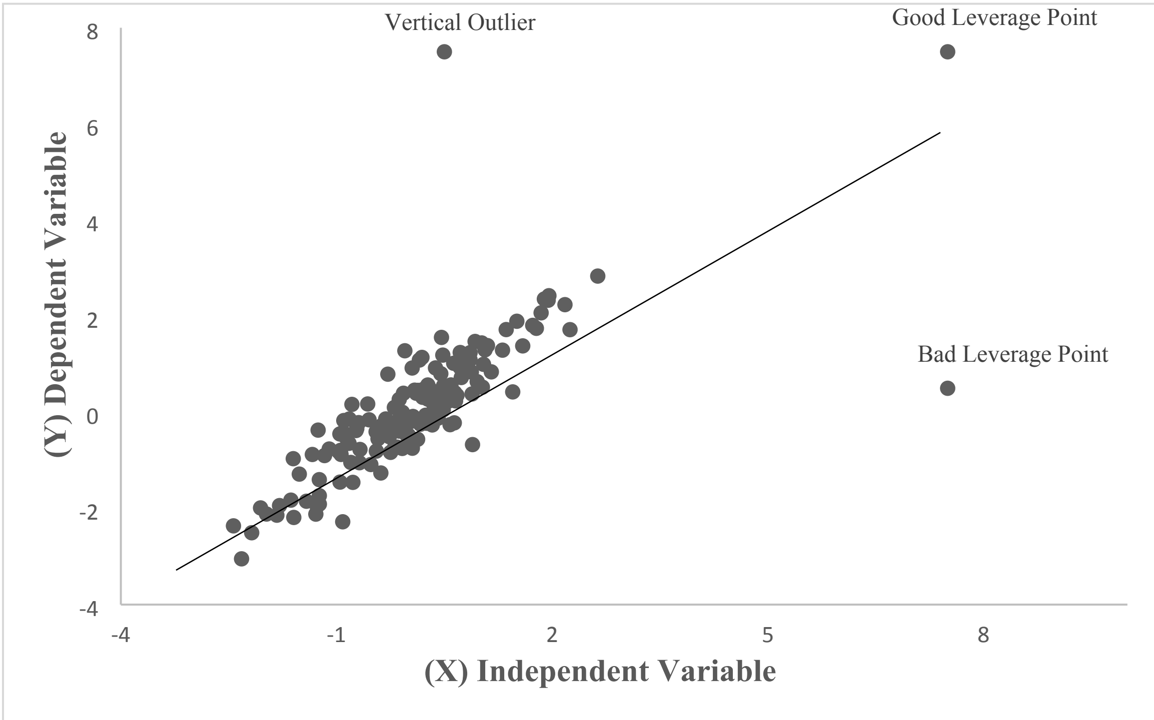

Figure II. Outlier Detection Plot for 2017 Capital Structure Data

Outlier detection plot for the year 2017 capital structure data. Robust standardized residuals measure the outlying of each observation in the (Y) dependent variable. Observations with robust standardized residuals outside the region identified by the two horizontal boundaries located at +/- 2.25 (values from the standard normal distribution that separate the +/- 1.25% most remote regions from the central mass of observations) are classified as extreme. Robust Mahalanobis distance measures the multivariate outlying of the (X) independent variable space. Observations with robust distances to the right of the vertical boundary located at ($\chi^2_{p,0.975}$), where p is the number of parameters in the model, are high leverage points. Vertical outliers are located outside of the horizontal and to the left of the vertical boundaries (regions labeled "V"). Good leverage points are inside the horizontal and to the right of the vertical boundaries (region labeled "G"). Bad leverage points are outside the horizontal and to the right of the vertical boundaries (regions labeled "B").

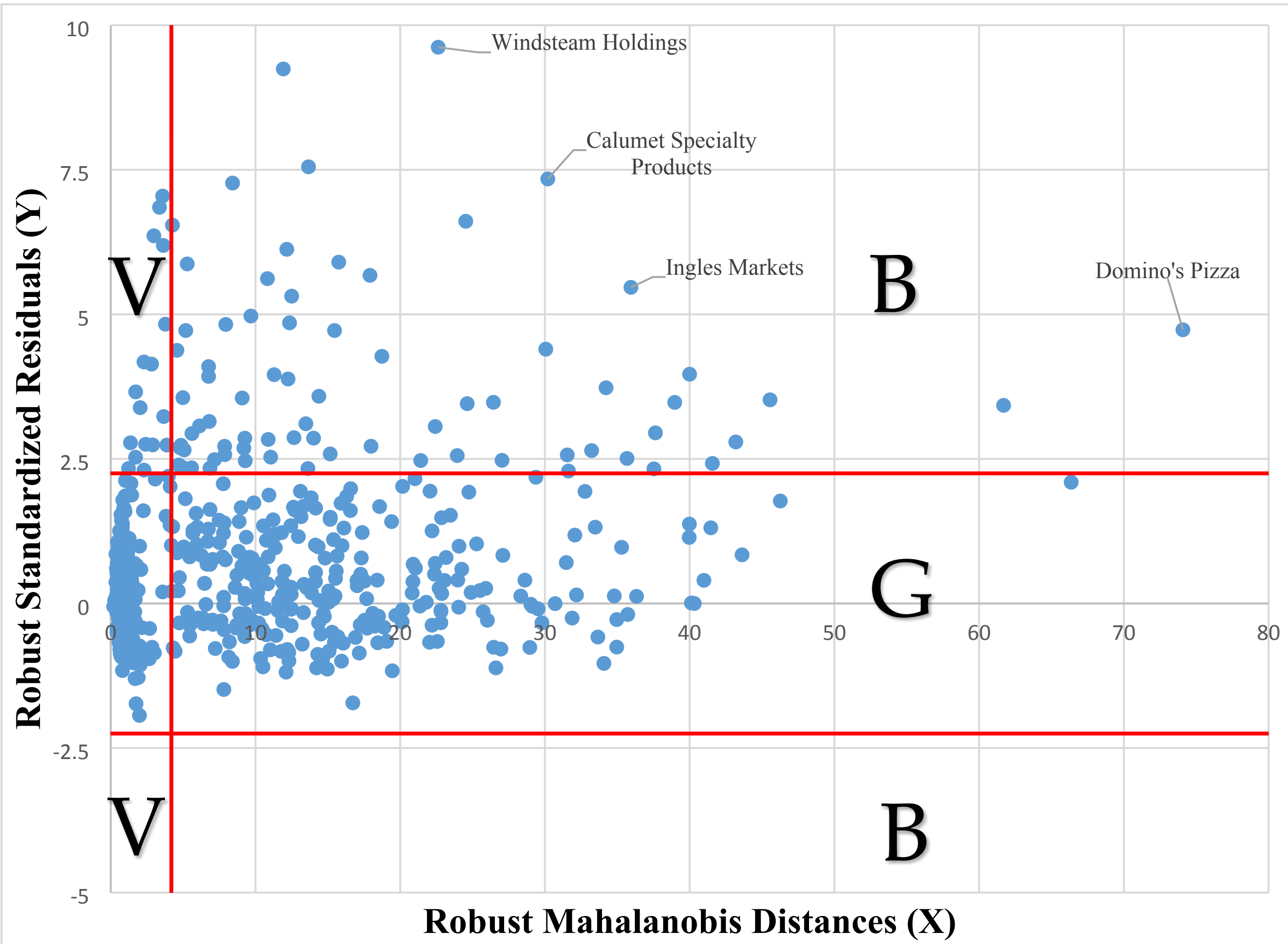

Figure III. Asset Pricing Vertical Outliers

This figure reports the monthly average incidence of vertical (independent variable) outliers in the asset pricing replication model/data for the years 1973-2017.

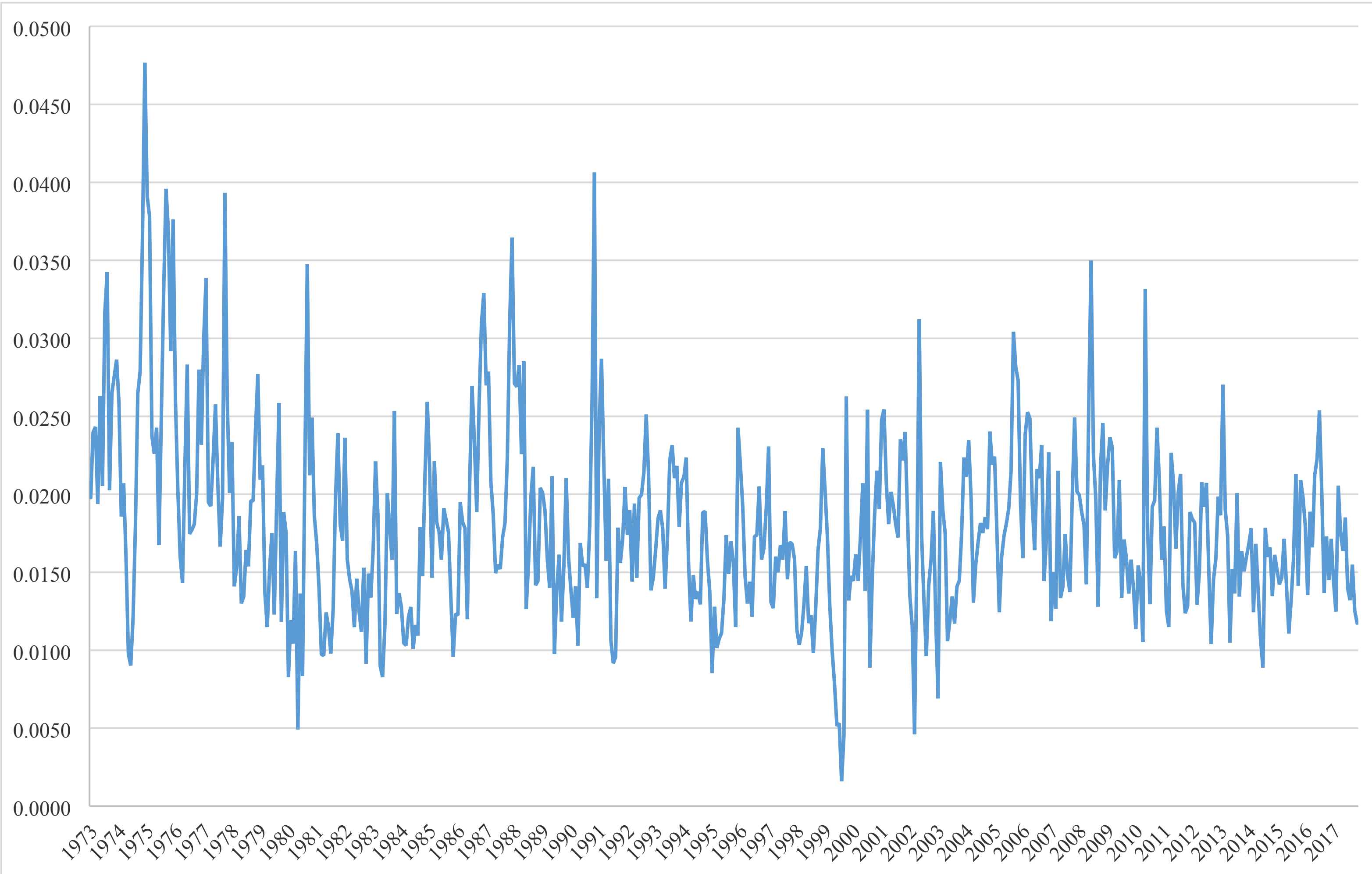

Figure IV. Asset Pricing Bad Leverage Points

This figure presents the monthly average incidence of bad leverage points (outliers in multivariate (X) independent variable space, but not in the (Y) dependent variable) in the asset pricing replication model/data for the years 1973-2017.

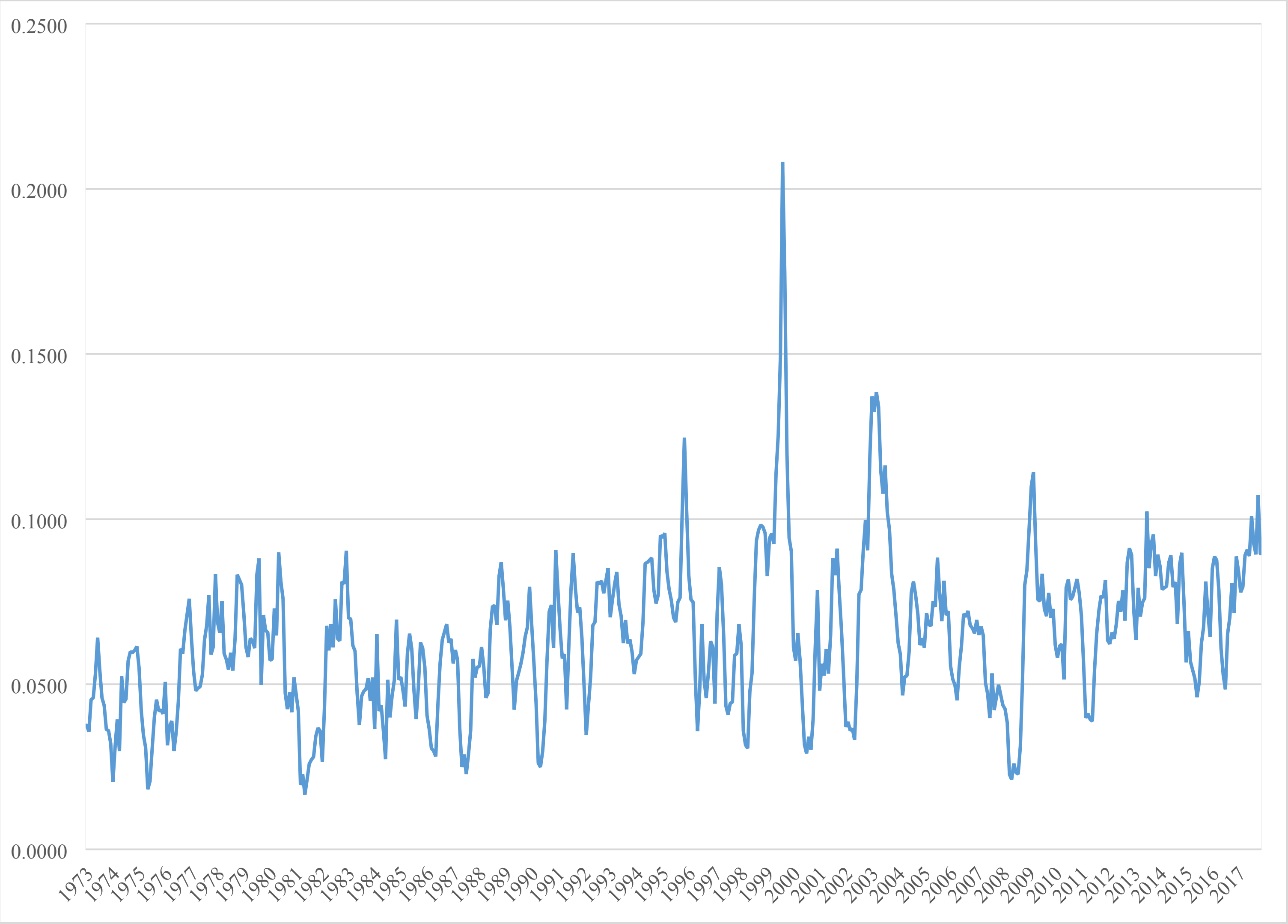

Figure V. Outlier Detection Plot for Becker and Stromberg's (2012) Fiduciary Duties (Table V) Data

This figure displays the outlier detection plot for Becker and Stromberg's (2012) Fiduciary Duties (Table V) data. Observations with robust standardized residuals outside the region identified by the two horizontal boundaries located at +/- 2.25 (values from the standard normal distribution that separate the +/- 1.25% most remote regions from the central mass of observations) are classified as extreme. Robust Mahalanobis distance measures multivariate outlying of the (X) independent variable space. Observations with robust distances to the right of the vertical boundary located at ($\chi^2_{p,0.975}$), where p is the number of parameters in the model, are high leverage points. Vertical outliers are located outside of the horizontal and to the left of the vertical boundaries (regions labeled "V"). Good leverage points are inside the horizontal and to the right of the vertical boundaries (region labeled "G"). Bad leverage points are outside the horizontal and to the right of the vertical boundaries (regions labeled "B").

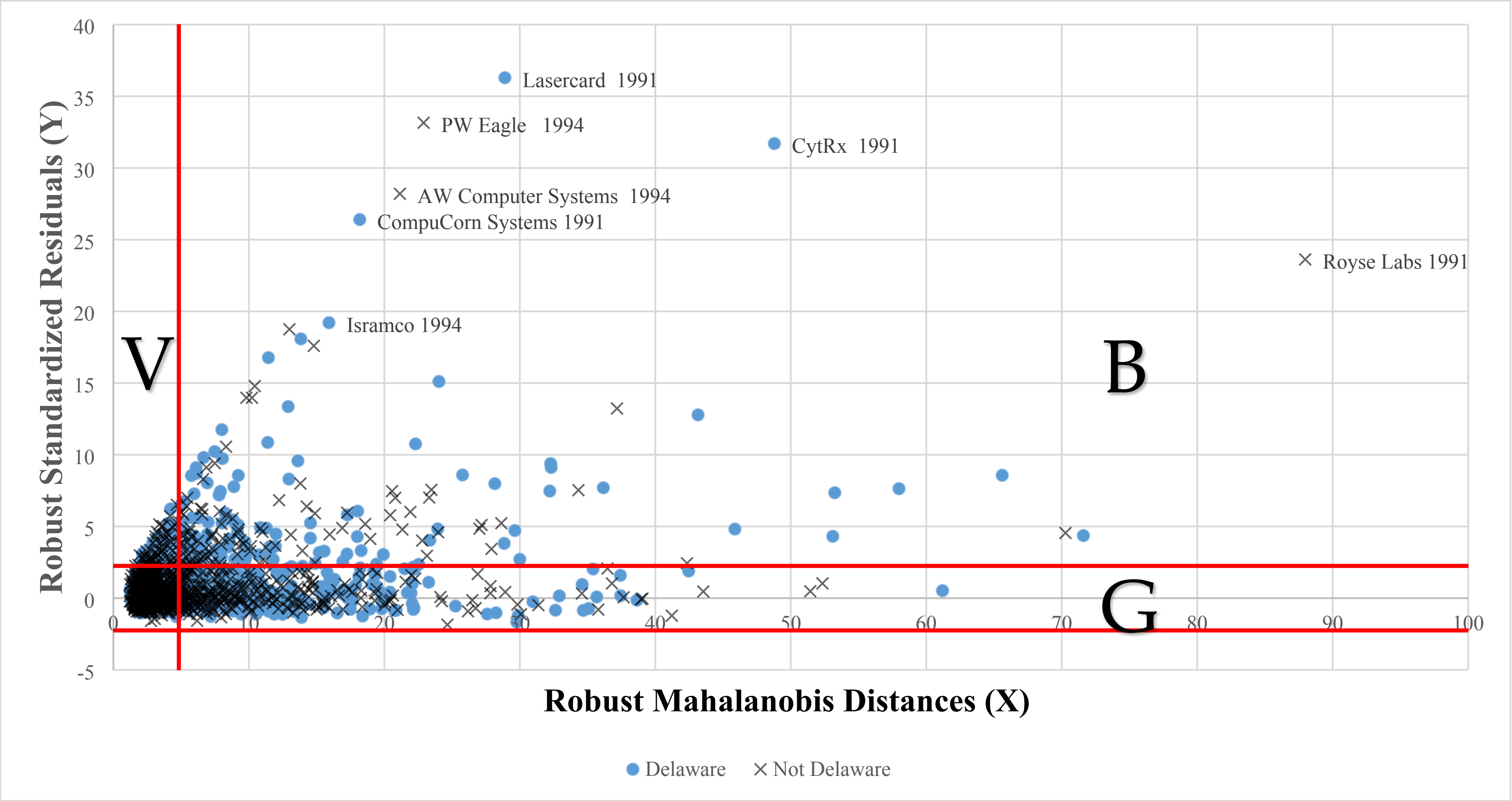

Figure VI. Outlier Detection Plot for Guthrie, Sokolowsky, and Wan's (2012) CEO Compensation (Model 5) Data

This figure displays the outlier detection plot for Guthrie et al.'s (2012) CEO compensation (Model 5) data. Observations with robust standardized residuals outside the region identified by the two horizontal boundaries located at +/- 2.25 (values from the standard normal distribution that separate the +/- 1.25% most remote regions from the central mass of observations) are classified as extreme. Robust Mahalanobis distance measures multivariate outlying of the (X) independent variable space. Observations with robust distances to the right of the vertical boundary located at ($\chi^2_{p,0.975}$), where p is the number of parameters in the model, are high leverage points. Vertical outliers are located outside of the horizontal and to the left of the vertical boundaries (regions labeled "V"). Good leverage points are inside the horizontal and to the right of the vertical boundaries (region labeled "G"). Bad leverage points are outside the horizontal and to the right of the vertical boundaries (regions labeled "B").

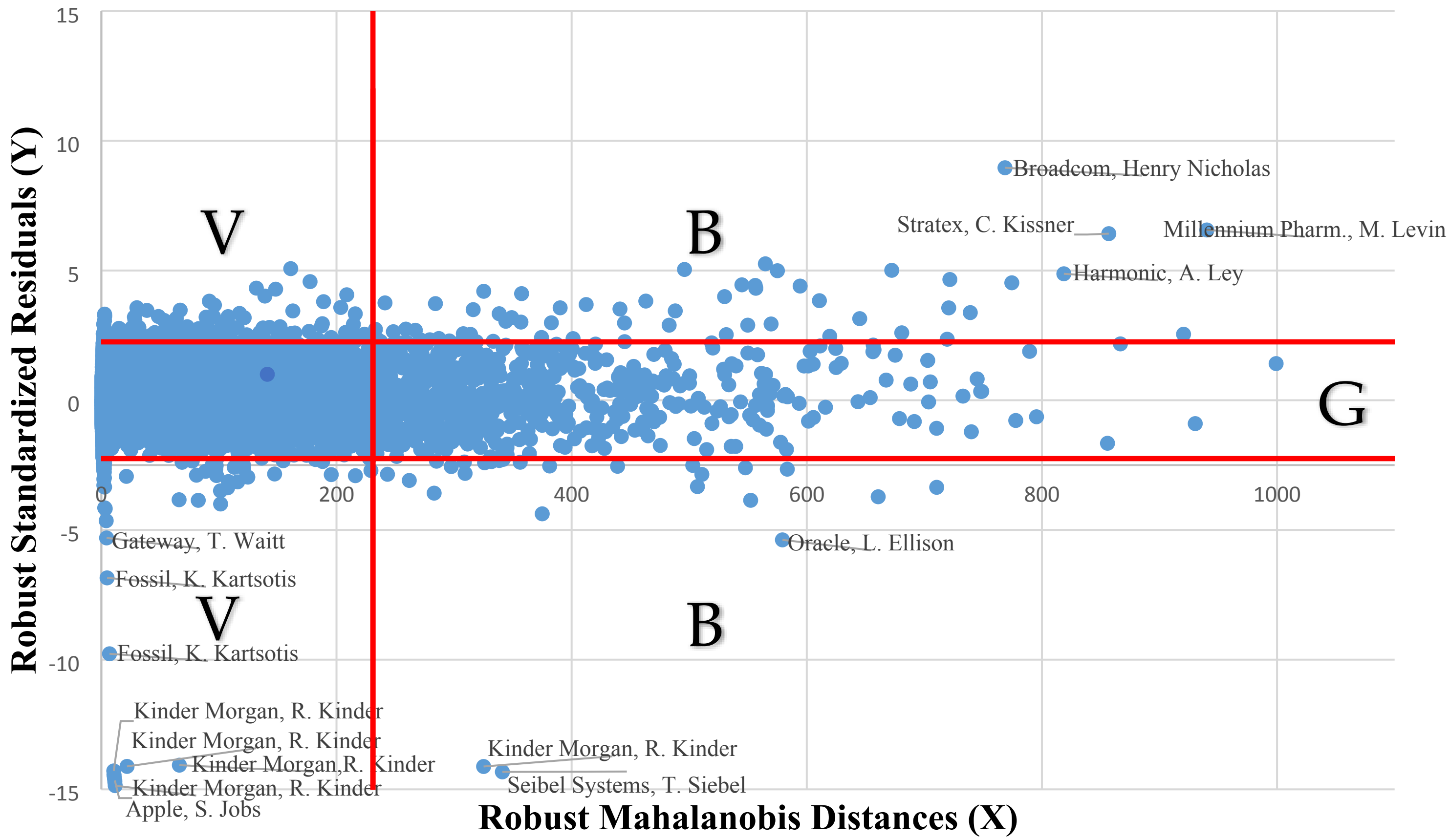

Table I. Incidence of Articles in Historical Finance Journals with Outlier Mention and Treatments

This table provides the number and percentage of articles published each year in the historical finance journals [*Journal of Finance* (JF), *Journal of Financial Economics* (JFE), *Review of Financial Studies* (RFS), and the *Journal of Financial and Quantitative Analysis* (JFQA)]. Panel A reports the outlier mitigation methods used in the historical finance journal articles from 2008-2017 using hand collection. Percentages total more than 100% due to multiple treatments in some papers. Panel B presents the incidences of articles with outlier mention, with OLS mention, and with OLS and outlier mentions from 1988-2017 using keyword searches. The data is from the EBSCO database for JF, RFS, and JFQA and the Science Direct database for JFE.

Panel A. Outlier Treatments in Articles Using OLS

Year	% Winsorize	% Trim	% Drop	% Winsorize, Trim, and/or Drop	% All Other Treatments
2008	35	11	31	75	38
2009	46	14	21	80	25
2010	41	24	10	75	29
2011	53	12	12	78	29
2012	64	20	6	91	21
2013	54	15	39	109	11
2014	35	13	30	78	35
2015	69	7	17	93	14
2016	56	26	21	103	12
2017	72	10	8	90	5
Average	52	16	17	85	24

Panel B. Incidence of Articles using OLS and Mentioning Outliers

Year	All Papers in JF, JFE, RFS, JFQA	% All Papers Mentioning Outliers	All Papers Utilizing OLS	% All Papers Utilizing OLS	% OLS Papers Mentioning Outliers
1988	194	7%	64	33%	17%
1989	209	5%	71	34%	13%
1990	228	7%	89	39%	15%
1991	195	5%	58	30%	10%
1992	201	6%	66	33%	15%
1993	207	7%	77	37%	9%
1994	182	13%	68	37%	21%
1995	201	12%	58	29%	28%
1996	199	14%	86	43%	26%
1997	223	11%	106	48%	19%
1998	201	14%	85	42%	28%
1999	208	8%	96	46%	23%
2000	216	9%	90	42%	23%
2001	220	16%	102	46%	26%
2002	243	13%	110	45%	25%
2003	241	18%	118	49%	34%
2004	250	15%	105	42%	28%
2005	248	23%	137	55%	35%
2006	259	24%	163	63%	32%
2007	288	23%	185	64%	30%
2008	298	27%	200	67%	37%
2009	379	25%	252	66%	35%
2010	381	24%	234	61%	38%
2011	400	26%	279	70%	34%
2012	365	27%	231	63%	27%
2013	364	30%	178	49%	44%
2014	316	28%	152	48%	33%
2015	328	30%	137	42%	36%
2016	355	31%	163	46%	36%
2017	386	32%	195	51%	37%

Table II. An Illustration of the Multivariate Outlier Issue

The table provides simulated data and regressions to illustrate the multivariate outlier problem. Panel A presents the data where the dependent variable, *Y*, equals *.5X1* + *.5X2* + the random error term. *X1* and *X2* are randomly generated with values in the range of 1-20. Columns (2)-(4) report the no outlier sample observations. Columns (5)-(7) report the multivariate outlier sample. The observations in the multivariate outlier sample are the same as the no outlier sample with the exceptions of Observations 18 and 19 where the dependent variable is much larger and smaller, respectively, relative to the independent variables as indicated by the other observations and the data generating process. Values are rounded to the nearest whole number to ease exposition. Panel B reports the regression estimates for the no outlier [Columns (1)-(3)] and the multivariate outlier samples [Columns (4)-(7)].

Panel A. Samples

	No Outlier Sample			Multivariate Outlier Sample		
Observation	**Y-Value**	**X1**	**X2**	**Y-Value**	**X1**	**X2**
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
1	4	10	2	4	10	2
2	5	7	3	5	7	3
3	6	4	7	6	4	7
4	7	15	1	7	15	1
5	7	3	12	7	3	12
6	7	3	10	7	3	10
7	11	10	9	11	10	9
8	11	7	12	11	7	12
9	11	20	3	11	20	3
10	11	16	4	11	16	4
11	11	12	8	11	12	8
12	12	3	12	12	3	12
13	12	19	2	12	19	2
14	12	19	10	12	19	10
15	13	20	5	13	20	5
16	15	19	9	15	19	9
17	16	15	10	16	15	10
18	17	16	17	***17***	***5***	***2***
19	18	19	19	***5***	***19***	***19***
20	21	17	20	21	17	20
Mean	11	13	9	11	12	8
Median	11	15	9	11	14	9
Minimum	4	3	1	4	3	1
Maximum	21	20	20	21	20	20
Std. Dev.	4	6	5	4	6	5

Panel B. Regressions

	No Outlier Sample			Multivariate Outlier Sample			
	All Observations	**Winsorized**	**Trimmed**	**All Observations**	**Winsorized**	**Trimmed**	**Multivariate Outliers Removed**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Intercept	0.554	0.893	1.511	6.348	7.414	10.557	-0.331
	(0.550)	(0.935)	(1.203)	(2.558)	(3.208)	(3.564)	(0.297)
X1	0.443	0.451	0.448	0.219	0.196	0.154	0.478
	(7.723)	(8.062)	(6.018)	(1.476)	(1.403)	(0.816)	(7.978)
X2	0.574	0.516	0.463	0.193	0.075	-0.234	0.653
	(8.887)	(8.401)	(5.875)	(1.091)	(0.453)	(1.006)	(8.261)
Observations	20	20	13	20	20	13	18
Adjusted R^2	87%	89%	89%	7%	<1%	<1%	86%

Table III. A Framework for Handling Outliers

Item	Comment
1) Do not winsorize	Winsorizing does not mitigate data errors and the choice of treatment level is arbitrary. It alters the data so that inferences are not generalizable.
2) Do not trim based on univariate statistics	Similar to winsorizing, this ad hoc univariate approach does not reliably mitigate multivariate outliers and may introduce new statistical biases.
3) Decompose constructed variables	Constructed variables using two or more variables can potentially misrepresent economic reality or mask data errors. Use both constructed variables and their underlying variables to identify outliers. For example, consider return on equity (ROE) defined as the ratio of net income (NI) to book equity (BE). A researcher may assume that a firm with a ROE of 20% is performing better than a firm with 10% ROE. However, if the 20% ROE firm has a negative 20 NI and negative 100 BE, it is clearly doing worse than a 10% ROE firm with positive 10 NI and positive 100 BE. Using both the constructed variables and their underlying variables helps to identify outliers that can lead to incorrect inferences.
4) Drop observations with missing values for the variables used in regressions	Observations with missing information for any of the dependent or independent observations are dropped in OLS regression computations. Prior to running the regression, dropping observations with missing data for key variables helps ensure that the univariate summary statistics are meaningful representations of the sample used in the regression analysis. In addition, removing observations that will not be included in the subsequent regression analysis reduces manual checking costs.
5) Identify extreme values of all dependent and independent variables	Not all univariate outliers are data errors. Likewise, not all univariate outliers are multivariate (regression) outliers (though many are). Identify the minimum and maximum value observations and those in the 1^{st}, 5^{th}, 50^{th}, 95^{th}, and 99^{th} percentiles. Include the underlying variables used in the construction of other variables (Item 3).
6) Correct or remove impossible and implausible observations	Extreme values that are impossible or implausible are generally data errors. Examine observations with extreme values to determine whether they are impossible or highly improbable. The preferred course of action is to correct data errors thereby avoiding potential sample selection bias. However, when the number of potential errors is large and manual investigation is prohibitively expensive, the next best alternative is to remove unlikely observations. The exception to the removing alternative is when there is structure in the measurement error.
7) Report and document removals and sample extremes	Reproducibility is a fundamental research requirement. It is difficult to replicate papers that do not explain how outliers are treated. Carefully document and justify data decisions and their effects in the data section. Report detailed summary statistics including minimum, the 1^{st}, 5^{th}, 50^{th}, 95^{th}, and 99^{th} percentile and maximum values.
8) Test for multivariate (regression) outliers	Conduct a formal test to determine whether multivariate outliers significantly influence OLS regression coefficient estimates. We recommend the DGV outlier test. If the test fails to reject the null of no outlier bias, report the OLS results. If the test rejects the null, continue to Item 9. The DGV outlier text is an option for the ROBREG command and is available by typing, "net from http://homepages.ulb.ac.be/~vverardi/stata" from within STATA.
9) Identify multivariate outliers	Identify outliers robustly in a multivariate context. We recommend using *S*-estimation to compute robust standardized residuals and Mahalanobis distances. Provide outlier detection plots to help identify and label particularly large multivariate outliers. *S*-estimation and outlier plots are options for the SREGRESS command that is available by typing, "net from http://homepages.ulb.ac.be/~vverardi/stata" from within STATA.
10) Evaluate the multivariate outliers	Carefully consider and examine the nature and origin of the outliers identified in Item 9 to identify potential data entry, sampling, omitted variables problems, and other errors. Manual examination costs can often be reduced by focusing on vertical and bad leverage point outliers with the largest robust standardized residuals and robust Mahalanobis distances. That is, by examining the most influential observations, those that heavily influence coefficient estimates, it is possible to reduce manual evaluation costs. Specifically, compute percentiles of the robust standardized residuals and Mahalanobis distances and re-run the OLS regressions excluding the extremes to identify influential observations. For example, if a previously significant OLS regression coefficient

	becomes insignificant when observations with the largest 1% robust standardized residuals and the largest 1% Mahalanobis distances are dropped, those observations in the top 1% are influential.
11) Correct the data and document	Correct data entry, sampling, omitted variables, and other errors. Remedy data entry errors by replacing erroneous entries with correct values. If this is not feasible, explain why and remove the suspected data entry errors. Remove sampling errors and add omitted variables to the regression model. Carefully document data decisions to preserve reproducibility and validity. Report the coefficient estimates.
12) Decide if further mitigation is prudent and how treat outliers	Identify, examine, discuss, and document the remaining influential outliers. Consider the nature of the research question and economic theory. If the research question involves tail-risk events or phenomenon, further mitigation can lead to incorrect inferences. However, even for tail-risk research questions, we recommend to identify, report, and examine influential observations (see Items 9, 10, and 11). Reporting influential observations improves inference when the influential observations are consistent with the tail-risk events or phenomenon of the research question. Reporting influential observations also removes the incentives for unscrupulous researchers to falsely argue research questions are about tail-risk phenomena and not general effects. For general effect research questions, which we contend represent the majority of finance research, mitigate multivariate outliers by dropping extreme influential observations and repeating the OLS regressions. That is, when theory suggests a general effect, influential outliers comprising a small fraction of the sample should not drive the empirical results. We generally recommend a maximum 10% cutoff rule for dropping influential observations. For example, if influential observations comprise less than 1% of the sample, dropping them improves inferences, but if they account for 30% of the sample, dropping can lead to incorrect inferences. Justify the cutoff rule in the context of the research question. Alternatively, for general effect research questions, employing outlier robust estimators, by design, yields general effect coefficient estimates. The robust estimators we employ in this paper can be found at http://homepages.ulb.ac.be/~vverardi/stata.
13) Recommendations for outlier robust estimators	We recommend using *MM* robust regression or *S*-estimation to mitigate outlier influence as they provide a good balance of robustness and efficiency. *MM* robust estimators can have greater efficiency than *S*-estimations when outlier bias is less severe. Perform repeated DGV outlier bias tests to determine the highest efficiency possible with outer robust *MM*-estimators. Median regressions and other quantile estimators have a BDP = 0 (sensitive to even a single outlier) and while they mitigate bias from vertical outliers, they do not protect against bad leverage points. Appendix A provides a summary of the costs and benefits of several outlier robust estimators. Finally, *MM* robust regressions and *S*-estimation regressions require patience. These outlier robust estimators are computationally intensive, particularly in asset pricing applications where robust Fama-Macbeth (1973) regressions may take several hours to process on a reasonably powerful desktop computer.

Table IV. Capital Structure Application Based on Peterson (2009) Study

This table provide percentage of outliers in the capital structure data identified from robust regressions of the market debt ratio on the log of market value of assets, log of 1 plus firm age, the profit to sales ratio, tangible assets ratio, market to book ratio, advertising to sales ratio, R&D to sales ratio, and a dummy to capture R&D spending. This table reports the annual percentage of outliers in the capital structure model using a sample of domestic firms with data in Compustat from 1973 to 2017.

Panel A. Incidence of Outliers (Vertical and Bad Leverage Points)

Years	Vertical Outliers	Bad Leverage Points	Years (cont.)	Vertical Outliers	Bad Leverage Points
1973	0.015	0.038	1996	0.062	0.060
1974	0.017	0.026	1997	0.060	0.075
1975	0.0026	0.033	1998	0.050	0.079
1976	0.035	0.036	1999	0.024	0.044
1977	0.039	0.027	2000	0.037	0.059
1978	0.024	0.029	2001	0.020	0.052
1979	0.028	0.031	2002	0.027	0.064
1980	0.028	0.023	2003	0.027	0.040
1981	0.021	0.022	2004	0.035	0.042
1982	0.066	0.044	2005	0.036	0.053
1983	0.081	0.045	2006	0.014	0.032
1984	0.049	0.057	2007	0.026	0.069
1985	0.049	0.050	2008	0.012	0.102
1986	0.036	0.063	2009	0.029	0.097
1987	0.033	0.062	2010	0.038	0.070
1988	0.062	0.073	2011	0.042	0.071
1989	0.064	0.069	2012	0.042	0.100
1990	0.072	0.056	2013	0.040	0.097
1991	0.097	0.067	2014	0.037	0.089
1992	0.097	0.054	2015	0.034	0.108
1993	0.115	0.058	2016	0.031	0.119
1994	0.056	0.105	2017	0.033	0.109
1995	0.065	0.074	All	0.041	0.062

The table reports mean (median) values for the capital structure sample that contains 37,430 firm-year observations from 1973-2017. The results provide segmentations for each observation type, as well as mean (median) differences in the non-outlier and outlier values. The notations [a], [b], and [c] denote statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Table I.

Panel B. Descriptive Statistics

	Observation Type Mean (Median) Values			**Mean (Median) Differences**	
	Non-Outliers (1)	**Vertical Outlier (2)**	**Bad Leverage (3)**	**Vertical - Non Outliers (2) – (1)**	**Bad - Non Outliers (3) – (1)**
Debt Ratio	0.174	0.448	0.453	0.274[a]	0.279[a]
	(0.152)	(0.437)	(0.447)	(0.285)[a]	(0.295)[a]
Ln(MV Assets)	6.280	6.229	6.839	-0.051	0.559[b]
	(6.259)	(6.102)	(6.771)	(0.157)	(0.512)
Ln(1+Firm age)	2.245	2.234	2.225	-0.011	-0.020
	(2.398)	(2.303)	(2.398)	(-0.095)	(-0.000)
Profits/Sales	0.215	0.175	0.229	-0.039[a]	0.014[a]
	(0.184)	(0.140)	(0.177)	(-0.044)[a]	(0.007)
Tangible Assets	0.182	0.174	0.315	0.008	0.132[a]
	(0.080)	(0.070)	(0.238)	(0.012)[a]	(0.158)[a]
Market-to-Book	1.101	1.033	1.738	-0.068[a]	0.637[a]
	(1.036)	(1.009)	(1.159)	(-0.027)[a]	(0.123)[a]
Advertising/Sales	0.016	0.016	0.029	0.001[c]	0.013a
	(0.013)	(0.012)	(0.015)	(0.001)	(0.002)[a]
R&D/Sales	0.003	0.001	0.009	-0.002[a]	0.006[a]
	(0.000)	(0.000)	(0.000)	(0.000)[a]	(0.000)[a]
R&D Dummy	0.195	0.117	0.273	-0.078[a]	0.078[a]
	(0.000)	(0.000)	(0.000)	(0.000)[a]	(0.000)[a]

This table reports the regression coefficients of the market value of debt to assets ratio on firm measures for the complete sample of domestic firms with data in Compustat. The sample contains annual observations on 4,919 firms from 1973-2017. Outlier test *p*-values are computed as in Dehon et al. (2012). All models include firm and year fixed effects. Firm level clustered robust *t*-statistics are in parentheses. The notations [a], [b], and [c] denote significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Table I.

Panel C. Regressions

	OLS No Winsorizing	OLS 1% Winsorizing	OLS 1% Trim	OLS 2.5% Winsorizing	OLS 2.5% Trim	MM Robust Regression
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	0.138[a]	0.145[a]	0.142[a]	0.156[a]	0.145[a]	-0.004[a]
	(7.448)	(8.314)	(8.284)	(9.350)	(7.914)	(4.625)
Ln(MV Assets)	0.017[a]	0.027[a]	0.030[a]	0.027[a]	0.033[a]	0.008[a]
	(3.226)	(7.538)	(8.248)	(7.813)	(8.148)	(4.207)
Ln(1+Firm Age)	0.009[a]	0.004	0.004	0.002	0.005[c]	-0.005[b]
	(2.744)	(1.316)	(1.361)	(0.841)	(1.749)	(2.044)
Profits/Sales	-0.012	-0.099[a]	-0.129[a]	-0.091[a]	-0.127[a]	-0.063[b]
	(0.757)	(6.020)	(7.146)	(5.271)	(6.463)	(5.460)
Tangible Assets	0.054[c]	0.052[b]	0.054[b]	0.044[b]	0.051[c]	0.007
	(1.94)	(2.020)	(1.989)	(1.723)	(1.829)	(0.318)
Market-to-Book	-0.021[a]	-0.043[a]	-0.059[a]	-0.053[a]	-0.062[a]	-0.020[a]
	(2.617)	(18.363)	(19.017)	(20.683)	(19.434)	(12.357)
Advertising/Sales	-0.087	0.028	0.026	0.040	0.118	-0.198[b]
	(0.770)	(0.291)	(0.228)	(0.371)	(0.889)	(2.664)
R&D/Sales	-0.024	-0.126	-0.159	-0.151	-0.279	-0.155
	(0.610)	(1.079)	(1.066)	(0.937)	(1.196)	(4.842)
R&D > 0 (=1 if yes)	0.007	0.006	0.008	0.008	0.010	0.010
	(0.978)	(0.908)	(1.094)	(1.107)	(1.395)	(1.211)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Outlier Test p-value	0.000	0.000	0.000	0.000	0.000	
MM Efficiency						30.7%
Adjusted R^2	0.748	0.756	0.739	0.756	0.738	0.033
Observations	37,430	37,430	33,543	37,430	28,384	37,430

Table V. Asset Pricing Application Based on Wahal and Yavuz (2013) Study

This table reports the results from the univariate and cross-sectional analysis of style and stock level momentum anomalies. Panel A provides mean (median) values for non-outlier observations, vertical outliers, and bad leverage points, as well as mean (median) differences. Panel B presents the monthly average coefficient estimates obtained from regressing future stock returns on prior style returns, stock returns, log size, and log BM. Newey West *t*-statistics with nine-month lags are reported in brackets. Future Stock Return is computed for the six-month period beginning the following month, Prior Style Return is the prior six-month value-weighted return on a style portfolio constructed using the intersection of the size and book-to-market quintiles, Prior Stock Return is each stock's prior six-month return, Log size is the natural logarithm of each stock's market value, and Log BM is the log of the book-to-market ratio. All of the variables are winsorized (trimmed) at the 1% level in Column 2 (3). The sample consists of all NYSE, Amex, and NASDAQ stocks from 1973-2017. Outlier test *p*-values are computed as in Dehon et al. (2012). The notations [a], [b], and [c] indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Table I.

Panel A. Descriptive Statistics

	Observation Type Mean (Median) Values			Mean (Median) Differences	
	Non-Outlier (1)	**Vertical (2)**	**Bad Leverage (3)**	**Vertical – Non-Outlier (2) – (1)**	**Bad – Non-Outlier (3) – (1)**
Future Stock Return	2.959	29.427	73.119	26.467[a]	70.159[a]
	(3.122)	(52.102)	(75.000)	(48.980)[a]	(71.878)[a]
Prior Style Return	2.363	1.997	5.153	-0.336[a]	2.791[a]
	(2.834)	(2.277)	(3.365)	(-0.557)[a]	(0.531)[a]
Prior Stock Return	3.625	-0.153	14.238	-3.778[a]	10.613[a]
	(3.321)	(-0.634)	(-0.606)	(-3.955)[a]	(-3.923)[a]
Log Size	12.760	12.112	11.274	-0.647[a]	-1.485[a]
	(12.692)	(12.086)	(11.154)	(-0.607)[a]	(-1.539)[a]
Log BM	-0.475	-0.460	-0.887	0.016[a]	-0.411[a]
	(-0.466)	(-0.475)	(-0.883)	(0.008)[b]	(-0.417)[a]

Panel B. Regressions

	FM-OLS	**FM-OLS 1% Winsorized**	**FM-OLS 1% Trimmed**	**FM-MM Robust**
	(1)	**(2)**	**(3)**	**(4)**
Prior Style Return	8.960[a]	9.525[a]	10.891[a]	10.330[a]
	(3.966)	(4.797)	(5.440)	(5.296)
Prior Stock Return	1.047	2.134[c]	3.311[a]	2.200[c]
	(0.745)	(1.847)	(3.165)	(1.685)
Log Size	-2.041[a]	-1.520[a]	-0.977[a]	0.215
	(7.000)	(7.005)	(5.351)	(1.566)
Log BM	-7.673[a]	-6.577[a]	-5.237[a]	-3.661[a]
	(9.250)	(10.998)	(10.383)	(7.401)
Outlier Test p-value	0.000	0.000	0.000	
MM Efficiency				28.7%
Adjusted R^2	8.02	8.50	7.47	3.32
Number of Observations	1,745,845	1,745,845	1,663,268	1,745,845

Table VI. Becker and Stromberg (2012) Study - Equity-Debtholder Conflicts (Replication using BS Data and OLS Code)

This table replicates the fixed effects panel OLS regressions of Table V, Model 1 in Becker and Stromberg (2012). The dependent variable is the volatility of ROA. Panel A reports the incidence of outliers by classification type and *S*-estimates of the robust standardized residuals and Mahalanobis distances. Panel B provides the regression results. Standard errors are clustered, where clusters are defined by the interaction of year and the firm's state of incorporation. Outlier test *p*-values are computed as in Dehon et al. (2012). The notation [a], [b], and [c] denote statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Table I.

Panel A. Comparing Treatment and Control Samples

	Observation Type Mean Values		
	Delaware (Treatment) (1)	**Not Delaware (Control) (2)**	**Treatment-Control (3)**
Incidence by Treatment Type			
Overall Sample	0.524	0.476	0.058
Non-Outliers	0.558	0.583	-0.025
Vertical Outliers	0.062	0.053	0.009
Good Leverage Points	0.277	0.263	0.014
Bad Leverage Points	0.102	0.100	0.002
Robust Standardized Residuals			
Overall Sample	1.087	1.008	0.079
Non-Outliers	0.318	0.248	0.070
Vertical Outliers	3.198	3.459	-0.261
Good Leverage Points	0.240	0.315	-0.075
Bad Leverage Points	6.287	5.954	0.333
Robust Mahalanobis Distances			
Overall Sample	6.369	6.148	0.221
Non-Outliers	2.696	2.676	0.020
Vertical Outliers	3.301	3.368	-0.067
Good Leverage Points	11.219	11.694	-0.475
Bad Leverage Points	15.145	13.309	1.836

Panel B. Regressions				
	Published Results OLS	**MM Robust Regression**	**OLS w/o Vertical Outliers**	**OLS w/o Bad Leverage Points**
	(1)	**(2)**	**(3)**	**(4)**
Intercept	0.1182[a]	-0.0019[a]	0.1023[a]	0.0585[a]
	(6.709)	(4.6702)	(6.5723)	(6.7103)
Delaware*Post-1991	-0.0076[b]	-0.0015	-0.0061[c]	-0.0004
	(2.0898)	(1.2174)	(1.8780)	(0.3424)
Controls				
Post 1991	0.0003	-0.0016	-0.0010	-0.0039[a]
	(0.0953)	(1.1748)	(0.4092)	(3.2219)
Return on Assets	0.0001	-0.0129	-0.0032	-0.0341[b]
	(0.0020)	(1.6326)	(0.0949)	(2.2451)
Return on Sales	-0.0020	0.0078	0.0065	-0.0069
	(0.0706)	(1.3474)	(0.1888)	(0.7308)
Ln Assets	-0.0296[a]	-0.0054[c]	-0.0270[a]	-0.0122[c]
	(3.9021)	(1.6944)	(3.1548)	(1.9279)
Ln Sales	-0.0101	0.0012	-0.0091	0.0010
	(1.5968)	(0.4359)	(1.3602)	(0.1828)
Ln MV	0.0191[a]	-0.0006	-0.0189[a]	0.0070[a]
	(3.1806)	(0.3126)	(3.0831)	(3.1192)
Depreciation/Assets	-0.0804	0.0047	-0.0811	0.0071
	(1.2762)	(0.2084)	(1.2921)	(0.2486)
Book Leverage	-0.0244	0.0105[c]	-0.0143	0.0210
	(0.7507)	(1.7469)	(0.4042)	(1.2035)
Market Leverage	0.1086[b]	0.0010	0.1028[c]	0.0197
	(2.0677)	(0.1364)	(1.8792)	(1.2150)
Q	0.0082	-0.0003	0.0088	0.0020
	(0.4929)	(0.1687)	(0.5063)	(0.6895)
Two-Year Stock Price Change	0.0006	0.0004	0.0007	-0.0015[c]
	(0.2347)	(0.5228)	(0.2664)	(1.7813)
Firm/Year Fixed Effects	Yes	Yes	Yes	Yes
Outlier Test p-value	0.000		0.000	0.000
MM Efficiency		50.1%		
Adjusted R-Squared	0.710	0.250	0.736	0.679
Observations	2,145	2,145	2,074	1,874

Table VII. Guthrie, Sokolowsky, and Wan (2012) Study - CEO Compensation (Revisited)
(Replication Using GSW Data and OLS Code)

This table replicates select fixed effects regressions of Appendix A in Guthrie et al.'s (2012). The dependent variable is the natural log of CEO pay. Standard errors are clustered at the firm-period level. Outlier test *p*-values are computed as in Dehon et al. (2012). The notations [a] and [c] denote statistical significance at the 1% and 10% levels, respectively. Variable definitions are provided in Table I.

Panel A. GSW (2012) – Replications of Model 5

	GSW Model 5 Including Apple & Fossil *OLS Replication*	Model 5 Including Apple & Fossil *MM Replication*	GSW Model 5 Excluding Apple & Fossil *Published Results*	Model 5 Excluding Apple & Fossil *MM Replication*
	(1)	(2)	(3)	(4)
Noncompliant x After	-0.014	0.041	0.060[c]	0.041
	(0.209)	(1.126)	(1.781)	(1.126)
Sales x Before	0.313[a]	0.328[a]	0.354[a]	0.329[a]
	(5.249)	(6.545)	(6.985)	(6.561)
Sales x After	0.290[a]	0.330[a]	0.342[a]	0.331[a]
	(4.578)	(6.653)	(6.775)	(6.668)
ROA x Before	0.480	0.291	0.378	0.281
	(1.428)	(1.134)	(1.131)	(1.095)
ROA x After	0.191	0.142[a]	0.120	0.142[a]
	(1.445)	(2.430)	(1.144)	(2.429)
RET x Before	0.083[a]	0.122[a]	0.087[a]	0.123[a]
	(2.689)	(4.564)	(2.829)	(4.576)
RET x After	0.285[a]	0.306[a]	0.319[a]	0.306[a]
	(5.754)	(10.58)	(7.454)	(10.599)
Tenure	-0.015	0.039[a]	-0.007	0.039[a]
	(0.831)	(2.568)	(0.419)	(2.565)
Outlier test p-value	0.000		0.002	
Max MM Efficiency		59.7%		68.5%
Adjusted R^2	0.096	0.045	0.121	0.045
Observations	5,318	5,318	5,306	5,306

Panel B. GSW (2012) – Replications of Model 7

	GSW Model 7 Including Apple & Fossil *OLS Replication*	Model 7 Including Apple & Fossil *MM Replication*	GSW Model 7 Excluding Apple & Fossil *Published Results*	Model 7 Excluding Apple & Fossil *MM Replication*
	(1)	(2)	(3)	(4)
Noncompliant	0.156[a]	0.089	0.167[a]	0.088
x high inst conc	(2.684)	(1.306)	(2.956)	(1.298)
	-0.068	0.031	0.026	0.031
x low inst conc	(0.863)	(0.742)	(0.669)	(0.737)
Sales x Before	0.317[a]	0.314[a]	0.356[a]	0.315[a]
	(5.334)	(6.338)	(7.023)	(6.353)
Sales x After	0.296[a]	0.316[a]	0.346[a]	0.317[a]
	(4.740)	(6.444)	(6.849)	(6.458)
ROA x Before	0.462	0.386	0.367	0.375
	(1.387)	(1.387)	(1.103)	(1.354)
ROA x After	0.192	0.176[a]	0.121	0.176[a]
	(1.452)	(3.016)	(1.151)	(3.014)
RET x Before	0.082[a]	0.127[a]	0.087[a]	0.128[a]
	(2.655)	(4.837)	(2.807)	(4.850)
RET x After	0.284[a]	0.306[a]	0.318[a]	0.306[a]
	(5.735)	(10.543)	(7.444)	(10.557)
Tenure	-0.015	0.042[a]	-0.007	0.042[a]
	(0.815)	(2.770)	(0.410)	(2.767)
Outlier test p-value	0.001		0.002	
Max MM Efficiency		68.9%		69.5%
Adjusted R^2	0.096	0.038	0.122	0.038
Observations	5,318	5,318	5,306	5,306