Online Appendix Table of Contents

List of Figures

List of Tables

Online Appendix

I. Additional Insights on the Case Selection Process by Plaintiffs' Lawyers

In Section III.E in the manuscript, we examine plaintiff-lawyer daily views of SEC filings around key litigation dates (Figure 1) for the full sample of cases constructed in Panel A of Table A.2. To further validate our data, we exploit our ability to distinguish between views from plaintiffs' lawyers participating in the initial filing and those that do not, and we examine *when* the views occur. One would expect participating plaintiffs' lawyers to be relatively more active using EDGAR in the days before the filing, while non-participating lawyers would be more active in the days afterward as they react to the revelation of the case and consider suing peer firms (see Gande and Lewis (2009)). However, as no news is available before the bad news announcement, we expect all plaintiffs' lawyers to exhibit a similar pattern in views in the days before and after the class period end. Figure A.1 presents this within-views analysis, based on the daily percentage of total views since our focus is on when the activity occurs. Results are consistent with our expectations.

Because we find some univariate evidence that plaintiff-lawyer views of SEC filings are associated with case merits in Table 1 in the manuscript, we next use plaintiff-lawyer daily views to provide further insight. First, we separately examine daily views around key litigation dates using our partitions for case merits. Results are shown in Figures A.2 and A.3.

Despite no significant difference in the number of plaintiff-lawyer views for settled versus dismissed cases over the 20 days preceding the filing date in Panel B of Table 1, it appears these insignificant differences are largely driven by the fact that views are relatively similar across these different cases until around three days before the filing date (Graph A of Figure A.2). However, starting two days before the filing date, plaintiff-lawyer views are significantly higher through the filing date for cases that settle ($p < 0.1$, untabulated). Moreover, we continue to see higher daily views for cases that settle over the following 10 days. Similarly, while the differences are generally insignificantly different, we observe higher plaintiff-lawyer views for cases that eventually settle after the class period end in Graph B, consistent with scrutinizing the event that triggered the litigation through the SEC filings to build a stronger case. Figure A.3 provides generally similar inferences, although the differences are less pronounced between cases alleging accounting fraud and those that do not. Collectively, these figures are consistent with our takeaway from Table 1 that plaintiffs' lawyers appear to view more SEC filings for more meritorious cases.

In Figure A.4, we capitalize on our ability to observe the different filings viewed by plaintiffs' lawyers to provide insight into *which* SEC filings plaintiffs' lawyers are viewing. Similar to Figure 1, we plot the number of daily EDGAR views in the 20 days before and after the filing date (Graph A) and class period end (Graph B) for the full sample of cases, but we separate views into different groups based on the nature of the SEC filing. We group the SEC filings into the following categories: 1) 10-Ks, 20-F, or 10-Qs (i.e., annual and quarterly filings); 2) 8-Ks or 6-Ks (i.e., material information disclosures by domestic filers and required disclosures by foreign filers); 3) Forms 3–5 (i.e., insiders' holdings, purchases, and sales of company securities); 4) comment letter-related filings; and 5) other filings. Some filings contain multiple exhibits (e.g., 8-Ks), so to hold that constant across filing types, we only keep one instance of IP-CIK-Date-Accession Number (i.e., the overall filing) for this and subsequent analyses that examine views of individual SEC forms.

Graph A of Figure A.4 shows that the most accessed filing before the filing date is 8-Ks/6-Ks, consistent with our observation and that of Rogers, Van Buskirk, and Zechman (2011) that earnings announcements are the most commonly cited type of firm filings in the complaints. The next most viewed categories are annual and quarterly filings and "other" filings. Views related to insider trading and comment letters do not appear to attract any attention before the filing date, suggesting that insider trading may be less important in building a case than is often alleged. Graph B presents the same analysis using the class period end, and we observe no meaningful views before the class period end, other than a spike in EDGAR views of insider trading filings on one day, which we find is driven by a single observation (untabulated).

In the prior analysis, 8-K/6-K filings are the most viewed types of SEC filings by plaintiffs' lawyers around the filing date in Figure A.4, but there are myriad types of such filings, some of which are relatively more accounting or non-accounting based. Accordingly, to provide more insight into the types of information viewed, we plot the daily plaintiff-lawyer views for the five most common types of 8-K filings in Figure A.5. We do not examine 6-Ks in this analysis because there are fewer 6-Ks in our sample in total than each of these types of 8-K filings. Additionally, 6-K filings lack item codes, impeding classification.

The most viewed 8-K filings are those related to financial statements and exhibits, including pro forma financial information (Item 9.01), which is viewed significantly more than any other type on multiple days after the litigation filing ($p < 0.05$). The next most frequently viewed type of 8-K filing is those that contain information on the firm's financial condition (e.g., earnings releases filed according to Item 2.02). Thus, while largely non-accounting-related 8-Ks are also viewed around this window (e.g., other events disclosed under Item 8.01), we find that 8-Ks viewed around the filing date are largely accounting-based.

To provide insight into whether the types of SEC filings viewed by plaintiffs' lawyers differ based on case merits, we perform a similar analysis as in Figure A.4, but we separate views based on whether cases are settled (Figure A.6) or whether they allege accounting fraud (Figure A.7). To maximize readability, given the combination of five groups of SEC filings that we use to classify views and the partition of our sample into two types of cases based on merits, we decompose our analyses into five separate sub-figures to correspond to each group of SEC filings and split the views based on our partitions for case merits in each. We use the same scale on the Y-axis for each of these sub-figures to facilitate comparison. Consistent with greater use of SEC filings by plaintiffs' lawyers being associated with case merits, in Figure A.6, we observe generally higher plaintiff-lawyer views for each type of

SEC filing after the filing date for settled cases, although these differences are generally insignificant. Once again, we obtain generally similar inferences for cases alleging accounting fraud versus those that do not, as shown in Figure A.7. Consistent with the nature of the allegations in the case, however, we observe higher views of annual or quarterly filings by plaintiffs' lawyers for cases alleging accounting fraud (Graph A of Figure A.7), relative to cases that settle (Graph A of Figure A.6) before the filing date.

Finally, we switch to the full-sample from Table A.2 Panel B and examine daily EDGAR views around major firm events: major restatement announcements[1] (Graph A), large negative stock market crashes (Graph B), earnings announcements (Graph C), and disclosures of ICWs (Graph D). This analysis provides insight into how frequent and infrequent firm news events affect plaintiff-lawyer scrutiny. We examine the 20 days before and after the events similar to Figure 1. Results show plaintiffs' lawyers scrutinize these events as expected, which is apparent from the spike in views on the event date. Plaintiffs' lawyers appear to most (least) scrutinize major restatement announcements (earnings announcements).

However, our biggest takeaway from this analysis is that *none* of these events result in substantial plaintiff-lawyer views. For example, even major restatements result in less than 1.5 average views on the announcement date. These limited views around bad news events are consistent with background discussions with a senior plaintiffs' lawyer, who noted that his firm did not heavily scrutinize restatements and other bad news events at companies unless they were already monitoring the firm and there were other significant bad news facts in the announcement. Consistent with many restatements not creating substantive litigation risk, over 90% of total restatements and over 55% of restatements classified as "fraudulent" by Audit Analytics do not result in subsequent securities class actions or SEC accounting and auditing enforcement releases (AAERs) that allege

[1]We focus on major ("Big R") restatements, which require an 8-K filing under Item 4.02 (Tan and Young (2015)), as many restatements do not cause significant market reactions (Hennes, Leone, and Miller (2008)).

accounting fraud and ultimately settle. Similarly, most stock price crashes do not result in litigation (Donelson, Kartapanis, et al. (2021)). Many of these news events are also relatively common, which further explains why they do not result in substantial scrutiny by plaintiffs' lawyers. For example, our sample has an average of more than 18,000 earnings announcements per year (untabulated), the vast majority of which are highly unlikely to contain negative news. As such, unless otherwise warranted, plaintiffs' lawyers are unlikely to examine every single one of those announcements. Further, we have an average of about 5,800 stock price crashes per year, many of them due to smaller firms that plaintiffs' lawyers would have weaker incentives to target because the litigation is less likely to be profitable (untabulated). These analyses suggest that bad news events may play a smaller role in plaintiff-lawyer scrutiny and triggering litigation than commonly believed.

II. Using a Longer Sample Period

As discussed in Section III.A in the manuscript, we begin our sample period in 2012 because this is the first year with sufficient data on plaintiff-lawyer investigations. These investigations imply that firms have high litigation risk and are frequently sued (see Section IV.A.2), so we believe this sample period results in better-specified models and increased ability to detect whether we can better predict litigation risk and market outcomes using our novel measure. Nonetheless, plaintiff-lawyer views are available in earlier periods, so we want to ensure the main predictive analyses reported in Tables 6 and 8 in the manuscript are not period-specific and continue to hold using an expanded sample period.

As shown in Graph A of Figure A.9, the number of plaintiffs' law firms for which we can identify IP addresses is relatively constant from 2010 through our main sample period but drops significantly in earlier years, largely due to challenges identifying IP addresses in earlier years as discussed in Section V.B. However, the percentage of securities class actions for which we can identify the IP address for at least one of the involved law firms remains relatively high from 2001 onward because the dropped law firms are disproportionately smaller and involved in relatively few cases (see Graph B). Thus, the main limitation for extending our tests to earlier years is not due to an inability to identify plaintiff lawyer IP addresses, but rather due to limited observable EDGAR views in earlier years. Specifically, as noted by Ryan (2017), there are significant gaps in the EDGAR search data before March 2003 and from September 2005 to May 2006 due to lost or corrupted log files. Consistent with this, the first year that we observe significant plaintiff-lawyer views is 2007 (see Graph C). Accordingly, because we use lagged views in our tests, we use 2008 as the first fiscal year in this expanded sample period.

Results using the sample period from 2008–2016 are tabulated in Tables A.4 and A.5. We continue to find plaintiff-lawyer views improve the precision and sensitivity of predictions of realized litigation risk, relative to the model from Kim and Skinner (2012), and predict future abnormal returns and return volatility in this expanded sample. Thus, plaintiff-lawyer views continue to have significant predictive ability in earlier years.

III. Do ***Lagged*** Plaintiff-Lawyer Views Proxy for Ex Ante Litigation Risk?

As discussed in Section IV.A.3 of the manuscript, the underlying premise of our use of lagged plaintiff-lawyer views to predict future litigation filings is that they proxy for ex ante litigation risk. If, for example, plaintiff-lawyer views are only due to litigation triggering events, such as shown in Figure 1 in the manuscript and Figure A.8, or existing litigation, they would have limited usefulness to future researchers. Rather, it is important to focus on inherent firm characteristics, rather than solely bad news events, when proxying for ex ante litigation risk. Accordingly, our primary tests use plaintiff-lawyer scrutiny in year $t{-}1$ to examine realized litigation year t.

If plaintiff-lawyer views proxy for the observable and unobservable factors that make firms good litigation targets (e.g., "bad" firms and/or firms with deep pockets or insurance policies), we would expect plaintiff-lawyer scrutiny to be relatively persistent. Conversely, if adverse events, such as litigation filings and restatement announcements, are the primary driver of plaintiff-lawyer scrutiny, the level of plaintiff-lawyer scrutiny should vary significantly over time. This is because neither restatements announcements nor litigation filings are persistent as shown in Table A.6. Specifically, only 0.10% of firm-years in our sample are associated with litigation filings in both the prior and current years. Similarly, only 0.13% of firm-years have major restatements in both the prior and current year.

To examine the persistence of plaintiff-lawyer views, due to their non-binary nature, we concentrate on firm-years in which there is at least one plaintiff-lawyer view during the current year and then independently sort current and prior year's total plaintiff-lawyer views into terciles (low, medium, and high scrutiny). We split firms yearly based on the 33rd and 67th percentile of total plaintiff-lawyer views. Given that views are discrete in nature and many firms can have a similar number of views, the number of observations per group slightly differs. We then plot a 3x3 matrix listing all possible combinations (e.g., low scrutiny last year and low scrutiny this year). If relatively stable firm characteristics drive plaintiff-lawyer scrutiny, one would expect most of the observations to lie on the diagonal (e.g., low scrutiny in both years). In particular, we would expect a large group of firms to persistently reside in the low scrutiny tercile, given that most firms face no substantive litigation risk (Nelson and Pritchard (2016)), and a large group of firms to persistently remain in the top scrutiny tercile, due to the importance of relatively persistent factors, such as firm size and industry in ex ante litigation risk (Kim and Skinner (2012), Brochet and Srinivasan (2014)). On the other hand, if adverse events mostly drive scrutiny, the level of scrutiny from plaintiffs' lawyers should vary significantly over time, so we should observe that firms frequently switch between scrutiny terciles from the prior year to the current year.

Table A.7 presents results. We find plaintiff-lawyer scrutiny is relatively persistent (i.e., mostly driven by firm characteristics). For example, about 50% (52%) of the firms classified as low (high) scrutiny firms during the prior year are also low (high) scrutiny firms during the current year. In contrast, only 21% of the observations classified as low scrutinized firms last year are in the highly scrutinized group this year. For comparison purposes, only 4% (8%) of firms that are sued (announce a major restatement) in year t

also faced a litigation (announced a major restatement) during t–1. That said, the fact that some firms move from the low scrutiny to high scrutiny tercile in subsequent years or vice versa, potentially due to changes in the litigation environment or company management, emphasizes it is important to have time-varying measures of firm litigation risk, such as our own.

We believe that the predictive ability of plaintiff-lawyer views in Table 6 in the manuscript is due to plaintiff-lawyer views proxing for these relatively persistent firm characteristics that make firms more likely to face future realized litigation risk. Consistent with this, the results in Table 6 Panel C in the manuscript, which uses t–2 values for all the independent variables to predict realized litigation in year t, are particularly compelling. Such twice-lagged variables long precede any bad news events around the class period end and litigation filing. In fact, plaintiff-lawyer views in year t–2 outperform the Kim and Skinner (2012) variables from year t–1 (untabulated). For example, we obtain in-sample (out-of-sample) precision and sensitivity improvements of 10%–15% (17%–22%) when examining $\text{SUED\_INV}_{i,t}$ in Column 5 of Table 6 Panel C as compared to Column 4 of Table 6 Panel A. This remarkable predictive ability of plaintiff-lawyer views even when twice-lagged is inconsistent with any explanation other than lagged plaintiff-lawyer views being a high-quality proxy for the firm characteristics that create litigation risk and may motivate firms to take actions accordingly, long before any public bad news events that may incite litigation.

The analyses in Table 6 indicate that the improved predictive ability of plaintiff-lawyer views is not due to a mechanical relation with *future* litigation (i.e., plaintiff-lawyers building the case in year t–1 that is ultimately filed in year t). Next, we perform five additional primary analyses to ensure that our improved specification is not due to plaintiff-lawyer views associated with *contemporaneous* litigation (i.e., views and litigation filings taking place in year t–1) and other bad news events. While we perform these analyses to be thorough, we note that such an alternative explanation is unlikely given the lack of persistence of realized litigation and bad news events as shown in Table A.6. For example, because litigation is not persistent, if plaintiff-lawyer views are primarily due to past or existing litigation, they should be *negatively* associated with future realized litigation, which is opposite to our findings in Table 6. Additionally, the predictive ability of plaintiff-lawyer views in year t–2 indicates that results cannot be due to bad news revealed in year t–1 that drives both views in year t–1 and litigation filings in year t. That said, it is also worth noting that the concern of confounding events in year t–1 is not unique to plaintiff-lawyer views as a proxy for ex ante litigation risk. Bad news events in year t–1 are also likely to affect many of the Kim and Skinner (2012) variables (e.g., stock returns and volatility).

First, we include an additional indicator variable in the models set to one for firms that are sued in year t–1. If plaintiff-lawyer views in year t–1 are proxying for litigation in year t–1, the inclusion of this predictor in the Kim and Skinner (2012) model should negate the superior predictive ability of our model. As shown in Table A.8, we continue to predict future realized litigation by an even larger margin relative to Kim and Skinner (2012). Notably, the coefficient on $\text{SUED}_{i,t-1}$ is negative and significant across most models, further demonstrating the lack of persistence of realized litigation. Second, we exclude firm-years that had a lawsuit in both years t–1 and t. We note that such an analysis induces look-ahead bias, but finding similar results to our main analyses ensures our results are not driven by these few firm-years. Results are shown in Table A.9. We continue to find similar inferences and note the sample size only decreases by 0.10%, due to the rarity of serial litigation. Third, we exclude all firm-years that face litigation in year t–1. This approach allows us to avoid look-ahead bias but results in larger (but still modest) sample attrition of 2.6%, relative to the prior analysis. Results are presented in Table A.10. Contradicting the notion that plaintiff-lawyer views proxy for prior year litigation, this specification results in our most powerful predictive model for $\text{SUED}_{i,t}$. The intuition behind this improvement is simple: by excluding those cases, we eliminate all views resulting from prior litigation, and as such, plaintiff-lawyer views are almost entirely due to ongoing monitoring.

Fourth, we exclude firm-years that in year t–1 include *any* major bad news event we can identify. Specifically, we exclude firm-years with litigation filings; earnings warnings; or announcements of major restatements, ICWs, CEO or CFO turnover, non-timely filings, or auditor changes in year t–1. While no vector of bad news events is exhaustive and this design creates significant sample attrition, it is a powerful test to show our results are not due to the correlation of plaintiff-lawyer views with other bad news events that instead predict future realized litigation risk. Results are presented in Table A.11. We continue to find plaintiff-lawyer views are positive and significantly associated with future realized litigation risk ($p < 0.1$), and models with views have significantly improved predictive ability.[2] Fifth, as an alternative to the sample attrition in the prior test, we attempt to strip out the portion of plaintiff-lawyer views due to these bad news events. We orthogonalize views to bad news events by individually regressing top and remaining plaintiff-lawyer views on individual indicator variables set to one for firm-years with litigation filings; earnings warnings; announcements of major restatements, ICWs, CEO turnover, CFO turnover, non-timely filings, and auditor changes; and a count of days in the firm-year with at least –10% market-adjusted returns. We then repeat the analysis in Table 6 Panel A using residuals from these regressions. We find similar results as shown in Table A.12. Further showing these bad news events have minimal effect on annual plaintiff-lawyer views, the correlation between these residuals and raw logged plaintiff-lawyer views exceeds 0.9 for both types of law firms (untabulated). The results are unambiguous: the predictive ability of plaintiff-lawyer views is not due to a relation between plaintiff-lawyer views and contemporaneous litigation or other bad news events or a mechanical relation with future litigation.

The prior findings and our assertion that lagged plaintiff-lawyer views proxy for ex ante litigation risk may appear in contradiction to the fact that bad news events do

[2] We also exclude firm-years with any days of at least –10% market-adjusted returns, which reduces the sample by 40% due to the frequency of extreme negative returns, and find similar results (untabulated).

increase plaintiff-lawyer views as shown in Figure A.8, Table 9, and Appendix C. However, as discussed with Figure A.8, we note that the economic magnitude of these bad news events on plaintiff-lawyer views is relatively small. Additionally, as discussed above, lagged plaintiff-lawyer views predate the class period end for virtually all cases in our sample and thus predate any litigation-associated bad news events.

However, just as our prior analyses demonstrate that the predictive ability of actual plaintiff-lawyer views is not due to these bad news events, we also test whether the predictive ability of *predicted* plaintiff-lawyer views is primarily due to inherent and relatively persistent firm characteristics, rather than non-persistent bad news events. We first re-estimate the determinants model in Appendix C but exclude the bad news events in year t. Specifically, we exclude variables in the following categories: *Accounting Events* (i.e., auditor changes, major restatements, non-timely SEC filings, and ICW announcements), *Personnel Events* (i.e., CEO and CFO turnover), and *Disclosure* (i.e., voluntary 8-K filings and earnings warnings). If the predictive ability of plaintiff-lawyer views is due to its positive association with these events, their exclusion should negate the ability for predicted plaintiff-lawyer views to predict future realized litigation risk.

Results of this simplified determinants model are shown in Table A.13. We continue to find similar inferences on the remaining variables in the determinants model. We then calculate predicted plaintiff-lawyer views using these relatively persistent firm characteristics and benchmark these predicted views against the measure of predicted litigation risk from Kim and Skinner (2012). As shown in Table A.14, we continue to significantly outperform their measure. For example, when tested after our sample period, we improve in-sample (out-of-sample) model precision and sensitivity by around 25%–30% (20%–30%).

Collectively, while bad news events often will result in a modest increase in plaintiff-lawyer views, these results demonstrate that it is not these views that explain the ability of plaintiff-lawyer views to predict future litigation. In contrast, these analyses further validate that plaintiff-lawyer views appear to be relatively persistent measures of firm quality that are strong predictors of future realized litigation risk.

IV. Alternative Model Specifications

Virtually all research designs include at least some potentially arbitrary design choices. To ensure that our results are not due to these choices, we discuss a variety of alternative specifications in our analyses to demonstrate the robustness of our findings.

The combination of a high frequency of zero plaintiff-lawyer views and a low frequency of extremely high views make it difficult to interpret the economic effect of views with logged values. To ensure results are not due to their functional form, we replicate our case-level and firm-year level predictive analyses using the inverse tangent (i.e., arctan) as an alternative non-linear model (Freeman and Tse (1992)). Relative to a logarithmic transformation, arctan: 1) has no issues with cases where the value is zero and, as such, requires no linear transformation before the arctan transformation and 2) results in smaller increases when views are small (e.g., increases from one to two views) and less incremental effect from extremely large views. We also use the inverse hyperbolic sine, which has similar properties to arctan for values near zero and is similar to log transformations with larger values (Johnson (1949), Burbridge, Maggee, and Robb (1988)). We find similar results as shown in Tables A.15–A.18.

We separate views from the top and remaining plaintiffs' lawyers in most of our tests due to their potentially different associations with litigation risk. However, to ensure this choice does not drive our results, we replicate the analysis in Table 6 Panel A in the manuscript using combined plaintiff-lawyer views. We find similar results as shown in Table A.19.

We use multiple industry indicator variables following Brochet and Srinivasan (2014) in our model of litigation risk in Table 6 due to substantial differences in the litigation rates before our sample in high litigation industries, as shown in Figure A.10. However, we retain the single high-risk indicator variable in the model from Kim and

Skinner (2012) to faithfully represent their model and the way that it has been used by extensive subsequent research.

We conduct two alternative specifications to ensure the differences in the litigation indicator variables across our models are not driving our results. First, we find similar inferences when using the single high-risk industry indicator from Kim and Skinner (2012) in our model as shown in Table A.20. Second, we adapt their model to include the multiple industry indicator variables from Brochet and Srinivasan (2014). Results are presented in Table A.21. Two takeaways are apparent from these tests. First, models using plaintiff-lawyer views are superior to those from Kim and Skinner (2012), regardless of the granularity of industry indicator variables. Second, both the model from Kim and Skinner (2012) and models including plaintiff-lawyer views improve when using the multiple industry indicator variables. Thus, it seems clear that the use of a single high-risk indicator variable to proxy for litigation risk is no longer appropriate due to changes in the litigation environment over time.

We primarily demonstrate the superiority of plaintiff-lawyer views as a proxy for litigation risk by comparing our model to that of Kim and Skinner (2012) because it is the most used and best-regarded model in prior research. It is in no way our intention to criticize this or any other model from prior research, but we need some baseline by which to examine whether the use of plaintiff-lawyer views provides an important contribution to the literature. However, we ensure that our improved predictability is not only due to the comparison used. Some papers in finance continue to only use industry indicator variable(s) to proxy for litigation risk (e.g., Brockman, Khurana, and Martin (2008), Jiang, Petroni, and Wang (2010), Callen and Fang (2015), Bird, Karolyi, and Ruchti (2019), and Hutton, Shu, and Zheng (2022)). We compare our model to these alternative approaches in Table A.22. We note our model has improvements of over 100% in both precision and sensitivity relative to models using only a single industry indicator variable and size when predicting our broader measure of realized litigation risk. We also find improved predictive ability relative to the lagged model in Brochet and Srinivasan (2014) (Table A.23).

Finally, Huang, Hui, and Li (2019) create a federal circuit-level measure of judge ideology to proxy for litigation risk (LIBERAL_COURT$_{i,t-1}$) based on research that liberal judges are more likely to favor investor plaintiffs (Fedderke and Ventoruzzo (2016)). While their measure is positively associated with litigation filings, which we confirm in our sample, by construction, it assumes that litigation is 1) homogeneous across firms in a given circuit and 2) does not greatly vary across time. This limited time-series variation is because circuit ideology can change only if a judge retires *and* the president that appoints the judge's successor is from a different party. In contrast, while admittedly less exogenous to the firm, plaintiff-lawyer views have significant firm-specific and time-series variation. Consistent with these differences, the univariate correlation between views (TOP_PLF_LN_VIEWS$_{i,t-1}$ and REM_PLF_LN_VIEWS$_{i,t-1}$) and LIBERAL_COURT$_{i,t-1}$ is relatively low (6%) ($p < 0.01$, untabulated).

We compare our model of litigation risk to LIBERAL_COURT$_{i,t-1}$ in Table A.24. We note that this table exhibits some sample attrition relative to the prior models from when firms lack data for headquarter locations and because LIBERAL_COURT$_{i,t-1}$ requires firms to be headquartered in the United States. We estimate equation (3) in the manuscript after replacing TOP_PLF_LN_VIEWS$_{i,t-1}$ and REM_PLF_LN_VIEWS$_{i,t-1}$ with LIBERAL_COURT$_{i,t-1}$. The precision and sensitivity of our model (Column 4), relative to theirs (Column 1), are over 50% (90%) higher in Panel A (B). Thus, while LIBERAL_COURT$_{i,t-1}$ is a statistically significant determinant of securities litigation, the enhanced firm-specific variation in our measure makes it a significantly better general measure of litigation risk.

V. Using AUC in Imbalanced Data Sets

While the AUC remains the main model performance metric in the finance and accounting literature, as discussed in Section IV.A.3, it is problematic for use in imbalanced data sets with binary dependent variables like ours. Recent advances in the machine learning and medical literature have identified serious limitations with the use of AUC in the analyses of rare events (e.g., cancer identification, gene detection, network intrusions, and disaster management) (see Swamidass et al. (2010), Saito and Rehmsmeier (2015), Brabec and Machlica (2018), and Brabec, Komárek, et al. (2020)). Namely, the AUC measures the ability of a model to distinguish between classes *over the entire range* of cutoffs used to classify observations into the different classes. In our setting, if each firm facing realized litigation risk is compared to each firm that does not face realized litigation risk, the AUC measures the proportion of comparisons that the firm facing realized litigation risk has a higher estimated probability from the model (see Hosmer and Lemeshow (2000)).

However, realized litigation risk is rare, and few firms face any substantive risk of litigation (Kim and Skinner (2012), Nelson and Pritchard (2016)). Accordingly, the relevant region to examine when evaluating whether a model correctly classifies firms on litigation risk is the region that identifies firms with the *highest* predicted probabilities (i.e., the region of interest). As summarized by Swamidass et al. (2010, p. 1348), "in many fundamental problems ranging from information retrieval to drug discovery, only the very top of the ranked list of predictions is of any interest and ROCs and AUCs are not very useful."

To illustrate this issue in our setting, in Figure A.11, we graph the ROC curve for models 1 (i.e., the full model from Kim and Skinner (2012)) and 2 (i.e., our simplified model with plaintiff-lawyer views) from Table 6 in the manuscript. As shown in Table 6, the in-sample AUC for model 1 is 0.674 versus 0.635 for model 2, so researchers only examining the AUC would conclude model 1 has improved predictive ability. However, as shown in Figure A.11, the two models perform almost identically in the high specificity cut points, with model 2 slightly outperforming (see bottom left; also known as the early retrieval area), which is the relevant region for researchers focused on minimizing litigation false positives (i.e., classifying as firms likely to be sued only those with the highest predicted probabilities). It is not until around 0.20 on the x-axis, which indicates the false positive rate equals 20%, that model 1 demonstrates improved predictive ability, resulting in its higher reported AUC. In other words, the model from Kim and Skinner (2012) only has an advantage, relative to our model, if researchers have wrongly classified more than 3,000 firm-years as $\text{SUED}_{i,t}$ observations, in addition to the cases correctly classified as $\text{SUED}_{i,t}$. However, it would not be reasonable to classify over 20% of the nonsued observations as sued firms because we can observe a true sued rate of 3% in our sample. Thus the ability of model 1 to outperform model 2 after this cut point, which is the reason model 1 has a higher AUC, is irrelevant for researchers attempting to identify the few firms that credibly may expect to be sued in a given year.

Inferences from using the AUC are even more misleading when using our combined measure of realized litigation risk ($\text{SUED\_INV}_{i,t}$) as the dependent variable in models 4 and 5 in Table 6, as shown in Figure A.12. Here, it is clear model 5 has improved ability to identify the few firms that face ex ante litigation risk (i.e., when the false positive rate is small), whereas model 4 only has improved predictive ability once the false positive rate is unacceptably high, with a false positive rate of over 35%. Given that the AUC is calculated over the entire distribution, it simply cannot meaningfully compare two models with such different distributional properties. As Saito and Rehmsmeier (2015, p. 12) note, "AUC (ROC) can be inaccurate for fair comparisons when two ROC curves are crossing each other."

For these reasons, as well as because AUC does not allow researchers to focus relatively more on false positives, which are relatively more costly in our setting, we focus our analysis on evaluating the models' precision and sensitivity using a classification rate that matches one's priors (i.e., we expect 3% of our firms to be sued, similar to Bao et al. (2020)). We believe that these performance metrics are the most appropriate statistical methodology, given our research question of identifying the relatively rare, high litigation risk firms. We note that the false negative rate can be calculated as 1 – sensitivity, so it is also informative of false negatives. Additionally, for completeness, we report specificity, but we do not expect this metric to be diagnostic in our setting, given our focus on examining the accuracy of the models' classifications of firms in the top 3% of predicted probabilities. Rather, we expect all of the models we examine to have very high accuracy at identifying true negatives using this threshold, given the rare incidence of litigation (i.e., most firms face no substantive litigation risk, and we classify most of them as facing no litigation risk using this threshold). Consistent with our expectations, *all* of the models that we examine have specificity of over 97% for largely mechanical reasons. Relative to the model from Kim and Skinner (2012), our simplified model has the same or larger specificity, but the increase is not statistically meaningful, which is why we do not discuss this metric in detail.

Another common metric used to evaluate models in machine learning (e.g., Saito and Rehmsmeier (2015), Brabec, Komárek, et al. (2020)) is the F1 score, calculated as $2 \times (\frac{precision \times recall}{precision + recall})$ (where recall refers to sensitivity). The intuition for this score is that a model can achieve high precision by simply predicting very few observations as high litigation firms. For example, if a model predicts only one observation as a high litigation firm and is correct, the precision will be 100%, although clearly sensitivity will greatly suffer. The F1 score thus forces a model to balance between the two. In our case, however, this is not an issue, as we predefine the cutoffs (3%), constraining all models to have the same number of predictions. Note, however, that readers can calculate this score using statistics already included in our models if desired. As our model improves on both precision *and* sensitivity, relative to that of Kim and Skinner (2012), we similarly achieve higher F1 scores (untabulated).

VI. Model Performance in Samples with Fewer Data Restrictions

Our argument for using plaintiff-lawyer views is not only that it is a better measure of ex ante litigation risk but also that it allows researchers to examine more firms. For example, roughly 20% of the number of stocks traded on major exchanges are traded OTC, but these firms are generally excluded from prior research due to lack of CRSP data coverage (Ang, Shtauber, and Tetlock (2013)). While these firms face lower litigation rates than those traded on major exchanges due to their generally smaller size, which makes litigation less profitable (Field, Lowry, and Shu (2005)), they are still sued. Further, litigation risk can have an even larger effect on their actions than for other firms (Donelson and Yust (2014)). Relative to exchange-traded firms, OTC firms have fewer investor protections, higher crash risk, greater insider trading prior to bad news, more opaque information environments, and less scrutiny from market participants, such as institutional investors, equity analysts, and short sellers (Gosnell, Keown, and Pinkerton (1992), Karmel (2001), Ang et al. (2013), Eraker and Ready (2015), and Brüggemann et al. (2018)). Given these higher agency costs and fewer alternative governance mechanisms, litigation risk plays a relatively more important governance mechanism role for OTC firms (Donelson and Yust (2014)). Thus, for example, OTC firms may respond more strongly to changes in litigation risk, such as the decreased risk from the 1999 Silicon Graphics case in the Ninth Circuit (Houston et al. (2019), Arena, Wang, and Yang (2021)). While researchers could estimate litigation risk for these firms using our measure, they could not do so following Kim and Skinner (2012) due to data requirements.

To validate that our measure retains strong predictive ability in larger samples, we first remove the requirement of having non-missing data for accruals. We require accruals in the analysis in Table 6 to maintain a constant sample in the manuscript where possible across our tables because accruals data is required in Table 9. As shown in Table A.25, we find similar results after removing this requirement.

Next, for the largest possible sample, we also remove the requirement for variables in the Kim and Skinner (2012) model. Validating that OTC firms face, albeit lower, litigation risk, average values of $\text{SUED}_{i,t}$ ($\text{SUED\_INV}_{i,t}$) for firms without CRSP coverage in the current or prior fiscal year in this sample is 0.6% (0.9%) versus 2.9% (6.2%) for other firms (untabulated). Results are shown in Table A.26. Demonstrating the effect of sample attrition from a lack of data, the sample size increases by over 80%, relative to

Table 6. Comparing performance metrics across different samples is problematic (Hosmer and Lemeshow (2000)), but it is worth noting that, while model precision decreases relative to Table 6, sensitivity increases by a larger percentage. As such, plaintiff-lawyer views appear to measure ex ante litigation risk in large samples with minimal data requirements.

VII. Determinants of Plaintiff-Lawyer Views versus Realized Litigation Risk

We examine the determinants of plaintiff-lawyer views in Table 9 in the manuscript. To further highlight the importance of this analysis, as opposed to simply examining the determinants of litigation itself, we examine the determinants of realized litigation risk. We also compare the ability of this expanded model to predict realized litigation risk relative to plaintiff-lawyer views, similar to the Table 6 analysis in the manuscript.

We first examine the association between the expansive vector of lagged firm characteristics and bad news events from Table 9 and future litigation filings (Column 1) or the combined measure of future litigation filings or plaintiff-lawyer investigations (Column 3) in Table A.27. We benchmark this model against our model with lagged plaintiff-lawyer views. This analysis yields two primary takeaways. First, we note that variables one would expect to be associated with future realized litigation risk (e.g., $\text{MAJOR\_RESTATE\_ANNCT}_{i,t}$, $\text{CEO\_TURNOVER}_{i,t}$, $\text{CFO\_TURNOVER}_{i,t}$) are insignificantly associated with future realized litigation. Accordingly, if plaintiff-lawyer views are mostly driven by these events, then the use of lagged views should also be insignificantly related to future realized litigation. We believe that this further demonstrates the enhanced construct validity of plaintiff-lawyer views as a proxy for litigation risk. Notably, these tests also demonstrate the enhanced construct validity of our broader measure of realized litigation risk ($\text{SUED\_INV}_{i,t}$). This is because the precision of these determinants to predict realized litigation risk increases by about 100% moving from Column 1 to Column 3. Second, the model with plaintiff-lawyer views significantly outperforms the extensive model of potential litigation determinants. For example, we find in-sample (out-of-sample) precision and sensitivity that is over 25% (70%) higher in

Column 2 than Column 1 when predicting future litigation filings.

Second, we perform the same analysis with non-lagged values of the variables from Table 9 and contemporaneous plaintiff-lawyer views. We expect significant increases in model performance in these tests due to capturing plaintiff-lawyer views associated with ongoing litigation and bad news events caused by the litigation itself (e.g., CEO or CFO turnover). Results are shown in Table A.28. Consistent with contemporaneous views capturing litigation-in-process, we observe staggering precision and sensitivity, both in-sample and out-of-sample, of around 50% when predicting litigation filings using contemporaneous plaintiff-lawyer views. This precision and sensitivity are nearly five times what we observe when we examine lagged plaintiff-lawyer views. These significantly higher values further illustrate that our tests using lagged plaintiff-lawyer scrutiny are capturing something fundamentally different: scrutiny of a firm *before* litigation is being considered. Notably, we continue to obtain in-sample and out-of-sample improvements in precision and sensitivity in this analysis using contemporaneous variables of over 170% using our model (Columns 2 and 4), relative to the extensive vector of litigation determinants (Columns 1 and 3). Thus, it seems clear plaintiff-lawyer views uniquely summarize the factors that create future realized litigation risk in a way that existing bad news events and firm characteristics from prior research cannot.

To further illustrate the advantage of examining the determinants of plaintiff-lawyer views versus just examining the determinants of realized litigation, which is what virtually all prior research is forced to do, we examine the differential ability of predicted plaintiff-lawyer views versus predicted litigation probability to predict future realized litigation risk. If examining the determinants of plaintiff-lawyer views provides no incremental value for future researchers, predicted plaintiff-lawyer views from Appendix C should not better predict future realized litigation risk than measures of predicted litigation probability. We measure predicted litigation probability similar to our measure of predicted plaintiff-lawyer views in Table 10 by using the coefficients estimated from

Column 1 of Table A.28 in sample periods outside the period in which they were estimated.[3] To ensure results are not due to differences in the estimation techniques—we estimate predicted plaintiff-lawyer views using OLS and predicted litigation probability using logistic regression—we also estimate litigation probability using these variables in OLS (untabulated).

Results are presented in Table A.29. We find that predicted plaintiff-lawyer views outperform measures of litigation probability. In Panel A, which examines a period before our main sample period, we can better predict future litigation filings using predicted plaintiff-lawyer views than either measure of predicted litigation probability. We find similar inferences examining a period after our main sample period in Panel B.[4] As a result, knowing the determinants of plaintiff-lawyer views can better predict litigation than knowing the determinants of litigation itself. Thus, this analysis further demonstrates that the use of lagged predicted plaintiff-lawyer views can also significantly decrease measurement error in measures of ex ante litigation risk.

VIII. Avoiding Litigation under High Scrutiny

While litigation itself is not the focus of our paper, the ability to observe firms that are scrutinized but not sued by plaintiffs' lawyers provides insights into the plaintiff-lawyer case selection process. Thus, we examine determinants of litigation *conditional on* a firm facing high scrutiny by plaintiffs' lawyers to identify potential tiebreak factors the lawyers may use when deciding which firms to sue (see Donelson, Hopkins, and Yust (2018)). We define highly scrutinized firms as those in the top 25% or 50% of top or remaining plaintiff-lawyer views in a given year.[5] We include several potential categories of variables that may affect case selection by plaintiffs' lawyers based on prior research and estimate

[3] We find similar inferences if we estimate litigation probability using Column 1 of Table A.27 (untabulated).

[4] We find similar inferences if we immediately evaluate the precision and sensitivity of $\text{PRED\_LITIG\_PROB-DETERM\_MODEL}_{Logit\ i,t-1}$ and $\text{PRED\_LITIG\_PROB-DETERM\_MODEL}_{OLS\ i,t-1}$ versus $\text{PRED\_LN\_VIEWS}_{i,t-1}$ by examining the top 3% of predicted values (untabulated).

[5] We calculate the percentiles after excluding observations with no views. We do not separately examine scrutinized firms for top versus remaining plaintiffs' lawyers as we lack power due to the small sample sizes.

the following logistic regression:

$$\begin{aligned}(1)\qquad \text{SUED}_{i,t} = \beta_0 &+ \beta_{1-2}\text{DEFENSE RESOURCES} + \beta_{3-4}\text{LITIGATION PROCEEDS} \\ &+ \beta_{5-7}\text{RISK ENVIRONMENT} + \beta_{8-9}\text{VOLUNTARY DISCLOSURE} \\ &+ \beta_{10}\text{INSIDER TRADING} + \beta_{11-13}\text{ACCOUNTING EVENTS} \\ &+ \beta_{14-19}\text{CONTROLS} + \text{YEAR FE} + \epsilon\end{aligned}$$

where *Defense Resources* is a vector of characteristics that may increase the ability and desire of a firm to contest litigation and is comprised of indicator variables set to one if the firm has a relationship with a top defendants' law firm ($\text{TOP\_SECURITIES\_LAW\_FIRM}_{i,t}$) and a top paid general counsel ($\text{HIGH\_COMP\_GC}_{i,t}$) (Bozanic, Choudhary, and Merkley (2019), Holzman, Marshall, and Schmidt (2020)). These variables may deter litigation, as they make it more costly and less likely to succeed. *Litigation Proceeds* is a vector of variables that may affect litigation recoveries. Higher stock returns ($\text{CAR}_{i,t}$) indicate less bad news and potential damages, decreasing litigation incidence (Kim and Skinner (2012)). In contrast, New York incorporated firms ($\text{NY\_INC}_{i,t}$) must disclose directors' and officers' (D&O) insurance premiums, which is alleged to be a potential tiebreak factor that increases litigation incidence (Donelson, Hopkins, et al. (2018)).

Risk Environment is a vector of characteristics related to the firms' risk environment ($\text{HIGH\_COMP\_CRO}_{i,t}$, $\text{RISK\_MNGT\_COMMITTEE}_{i,t}$, and $\text{SUED}_{i,t-1}$). The first two variables ($\text{HIGH\_COMP\_CRO}_{i,t}$ and $\text{RISK\_MNGT\_COMMITTEE}_{i,t}$) are indicator variables set to one if the firm has a top-paid chief risk officer (CRO) and risk management committee, respectively, and may indicate a commitment to risk management (see Ellul and Yerramilli (2013)). Risk management is believed to reduce litigation risk by D&O insurers (Baker and Griffith (2010)) and may make plaintiffs' lawyers believe litigation is more likely to be dismissed. However, these indicators of risk management also may indicate the firm is riskier and thus more likely to have behaved in ways that appear fraudulent (Pagach and Warr (2011)), making them positively associated with litigation. Relatedly, it is unclear what the relation will be when the firm has been sued in the prior year ($\text{SUED}_{i,t-1}$). Plaintiffs' lawyers may view the litigation as a sign that the firm is high risk and worth suing, making it positively associated with litigation. On the other hand, plaintiffs' lawyers may believe that D&O insurance coverage is already depleted, resulting in lower potential settlements (Donelson, Hopkins, and Yust (2015)). $\text{SUED}_{i,t-1}$ also controls for the fact that a firm may be misclassified as a plaintiff-lawyer high-scrutiny firm due to EDGAR views related to the prior year's litigation.

Voluntary Disclosure is a vector of voluntary disclosure ($\text{VOLUNTARY\_8-Ks}_{i,t}$ and $\text{EARN\_WARN\_ANNCT}_{i,t}$) that have been argued to reduce litigation incidence and settlement amounts (Skinner (1997), Field et al. (2005), and Donelson, McInnis, et al. (2012)).[6] Thus, these variables may reduce the threat of litigation when plaintiffs' lawyers scrutinize firms. However, these disclosures may also highlight bad news that plaintiffs' lawyers could target, making firms more likely to be sued. *Insider Trading* is comprised of Forms 3–5 filings ($\text{FORMS\_3-5}_{i,t}$), which have been used as evidence managers have intentionally misled shareholders and are associated with case merits (Johnson, Nelson, and Pritchard (2007)). Thus, similar to Rogers et al. (2011) and Billings and Cedergren (2015), who find that insider trading increases litigation risk, we expect it has a positive association with litigation. *Accounting Events* is comprised of $\text{MAJOR\_RESTATE\_ANNCT}_{i,t}$, $\text{ICW\_ANNCT}_{i,t}$, and $\text{NON-TIMELY\_FILING\_ANNCT}_{i,t}$. These negative events should increase the likelihood of a filing and may be deemed hard evidence that is particularly useful when building a case (Johnson et al. (2007)). We exclude auditor and executive turnover announcements, as they usually occur after a litigation filing. Finally, we include *Controls*, a vector of firm characteristics that may be correlated with the other variables ($\text{LN\_MVE}_{i,t}$ and $\text{FPS}_{i,t}$), and year fixed effects. Variables are defined in detail in Table A.41.

Columns 1–2 (3–4) of Table A.30 present the results of a logistic regression that

[6] We use $\text{VOLUNTARY\_8-Ks}_{i,t}$, rather than $\text{LN\_VOLUNT\_8-Ks}_{i,t}$ as per Table 9 in the manuscript, because sued firms have a shorter window in which filings can happen because we only count disclosures that occur before the filing. Thus, using the raw number, rather than a number scaled by the days over which it is accumulated (i.e., disclosures on a per-day basis), may indicate that firms with less voluntary disclosure are more likely to be sued, solely due to the shorter period leading to a smaller raw number of filings.

includes firms in the top 25% (50%) of yearly EDGAR views by either top or remaining plaintiffs' lawyers during the prior year. Columns 1 and 3 (2 and 4) present results without (with) year fixed effects. Neither of the variables in *Defense Resources* is significantly associated with litigation. Firms with higher $\text{CAR}_{i,t}$ and $\text{SUED}_{i,t-1}$ are less likely to be sued. Interestingly, we obtain mixed evidence on disclosures. We find evidence that $\text{EARN\_WARN\_ANNCT}_{i,t}$ makes it less likely that a firm will be sued ($p < 0.05$ in the 25% sample), helping reconcile our findings in Table 9 with research that timely disclosures of bad news lower litigation risk (Skinner (1997), Field et al. (2005), and Donelson, McInnis, et al. (2012)). On the other hand, $\text{VOLUNTARY\_8-Ks}_{i,t}$ continues to be positively associated with litigation ($p < 0.1$).[7] Thus, while disclosing bad earnings news has a nuanced relation with plaintiff-lawyer scrutiny and litigation, voluntarily disclosing other bad news results in both additional scrutiny and higher litigation incidence. There is no association between $\text{FORMS\_3-5}_{i,t}$ and the likelihood of being sued. $\text{MAJOR\_RESTATE\_ANNCT}_{i,t}$ is positively associated with litigation ($p < 0.01$), most likely due to the larger damages that can be claimed from stock price drop upon the announcement and its use as an indicator of case merits (Johnson et al. (2007)). This finding also helps explain the prevalence of "stealth" restatements (Files, Swanson, and Tse (2009)). Surprisingly, $\text{ICW\_ANNCT}_{i,t}$ is negatively associated with litigation ($p < 0.1$). In addition to the fact that cases with ICWs do not result in larger settlements, this result may be because material weaknesses are sticky (see Newton et al. (2016)), so plaintiffs' lawyers may have already had that information in the prior year. Finally, several high litigation industry indicators remain significant predictors of being sued ($p < 0.01$), even in this sample of highly scrutinized firms.

One limitation of the prior analysis is that the small sample due to focusing on high-scrutiny firms limits our statistical power. Accordingly, similar to the longer-window analyses in Section II, we examine this analysis over the same extended period. Results are

[7]As shown in the following section, these results are driven by non-earnings information.

shown in Table A.31. While we continue to find that the significant predictors in Table A.31 remain statistically significant, we also identify additional significant determinants in this longer sample period. We find weak evidence that $\text{NY\_INC}_{i,t}$ is positive and significantly related to litigation filings ($p < 0.05$ in the 25% sample). Further, $\text{HIGH\_COMP\_CRO}_{i,t}$ (in the 25% sample), $\text{NON-TIMELY\_FILING\_ANNCT}_{i,t}$ (in the 50% sample), and $\text{FORMS\_3-5}_{i,t}$ are significant and positively associated with litigation ($p < 0.1$).

IX. Different Types of Voluntary 8-K Filings

When examining determinants of plaintiff-lawyer scrutiny in Section IV.A of the manuscript and determinants of litigation, conditional on a firm facing high levels of plaintiff-lawyer scrutiny, in the preceding section, we identify and aggregate voluntary 8-Ks following Bourveau, Lou, and Wang (2018) and He and Plumlee (2020). Specifically, we identify 8-Ks with Item 2.02 (Results of Operations and Financial Condition), 7.01 (Regulation FD Disclosure), or 8.01 (Other Events) as voluntary in nature.

However, we do not expect each type of 8-K filing to be equally associated with scrutiny or as a tiebreaker of litigation among highly scrutinized firms. For example, disclosure of operations (Item 2.02) is likely to attract plaintiff-lawyer attention, as they want to examine if the firm discloses any negative news. Although firms are expected to submit four 8-Ks with Item 2.02 (i.e., one per quarter), firms often submit more. For example, some of these additional 8-Ks we observe inform investors of upcoming conference calls, submit results of subsidiaries, and other financial information. Thus, one would expect the more 8-Ks with Item 2.02 that a firm submits, the more scrutiny it will receive. However, for Items 2.02, the most important characteristic that could serve as a tiebreaker should be the content, rather than the number of forms filed per year. As such, we do not expect the number of 8-Ks with Item 2.02 to be associated with the likelihood of litigation among highly scrutinized firms. On the other hand, disclosure of "Other Events" is likely to both serve as a determinant of views, as plaintiffs' lawyers will want to examine the information disclosed, and a potential litigation tiebreaker from providing material additional information that can serve as a source upon which to build a case. We have no expectations for 8-Ks containing Item 7.01.

In Tables A.32 and A.33, we repeat the analyses performed in Tables 9 in the manuscript and A.30, respectively, after splitting the voluntary disclosure variables into these categories of voluntary 8-Ks. Consistent with our expectations, Table A.32 indicates that the positive relation between the voluntary 8-K filings and plaintiff-lawyer scrutiny reported in Table 9 is driven by 8-Ks that include Items 2.02 and 8.01 ($\text{LN\_8-Ks\_ITEM2.02}_{i,t}$ and $\text{LN\_8-Ks\_ITEM8.01}_{i,t}$, respectively, $p < 0.01$). On the other hand, Table A.33 indicates that, when focusing on the highly scrutinized firms, only 8-Ks filed under Item 8.01 are significantly associated with future litigation and only in the specification in Column 4 ($\text{VOLUNTARY8-Ks\_ITEM8.01}_{i,t}$, $p < 0.1$). Thus it appears that bad earnings news may reduce the incidence of litigation, while other bad news may increase it.

X. Miscellaneous Analyses and Robustness

To be consistent across columns in Table 6 in the manuscript, we use a 3% threshold when predicting $\text{SUED}_{i,t}$ and $\text{SUED\_INV}_{i,t}$. However, it is arguably more appropriate to use a higher threshold when predicting $\text{SUED\_INV}_{i,t}$ due to its greater frequency. As shown in Tables A.34–A.35, we obtain similar inferences if we use a 7% rate based on Table 4 or a 10% rate based on Nelson and Pritchard (2016), respectively. In fact, we find that the sensitivity for our model further outperforms the model from Kim and Skinner (2012) using these higher thresholds.

Because we are unaware of prior empirical research that has examined plaintiff-lawyer investigations, we separately examine the ability of plaintiff-lawyer views to predict these investigations, as opposed to both these investigations and litigation filings. Results are presented in Table A.36. We find similar inferences to Table 6 in the manuscript and note that the model including plaintiff-lawyer views can also predict

INVESTIG_ANNCT$._{i,t}$ with higher precision and sensitivity than the model from Kim and Skinner (2012).

In our analysis on the determinants of plaintiff-lawyer views in Table 9 in the manuscript, we find that prior year returns (CAR$_{i,t-1}$) are negatively associated with current year plaintiff-lawyer views. As we expect plaintiff-lawyer views to respond to unusual market returns in a more timely manner, we also perform quarterly analysis using the same design and controls as Table 8 in the manuscript with the difference that the independent variables are lagged by one quarter and views are contemporaneous. In addition to lagged buy-and-hold abnormal returns, we also add contemporaneous returns. As shown in Table A.37, we find that both prior and current quarter abnormal returns are negatively associated with current plaintiff-lawyer views ($p < 0.01$). Thus, plaintiff-lawyer views both predict and respond to abnormal stock returns in relatively short horizons.

Additionally, in Table 9, we use year t–1 values for many independent variables. As shown in Table A.38, we find similar results if we use values from year t for all variables. Alternatively, we use year t–1 for all independent variables to predict, rather than explain, plaintiff-lawyer views. Results are shown in Table A.39. We continue to find generally similar inferences with some notable changes. For example, we now find that the coefficients on AUDITOR_CHANGE_ANNCT$_{i,t}$ are negative for both types of plaintiff's lawyers ($p < 0.1$), likely due to reduced future scrutiny after a temporary spike in scrutiny around the auditor change announcement. Additionally, we now find that POSITIVE_NON-GAAP_ADJ$_{i,t}$ is positive and significant for plaintiffs' lawyers ($p < 0.01$), indicating that positive non-GAAP adjustments likely result in future scrutiny of the firm. Notably, the R^2 in both models is even lower in this specification as compared to Table 9, likely because temporary increases in plaintiff-lawyer views from some of the bad news events dissipate by the following year.

Finally, in Table 9, EARN_WARN_ANNCT$_{i,t}$ is set to zero for firms missing IBES coverage to avoid substantial sample attrition. Table A.40 shows that we obtain similar results if we also include an indicator variable set to one for firm-years that lack IBES coverage, similar to the approach in prior research (Liljeblom, Pasternack, and Rosenberg (2011), Cassell, Dreher, and Myers (2013)).

References

Ang, A.; A. A. Shtauber; and P. C. Tetlock. "Asset Pricing in the Dark: The Cross-Section of OTC Stocks." *Review of Financial Studies* 26 (2013), 2985–3028.

Arena, M.; B. Wang; and R. Yang. "Securities Litigation and Corporate Tax Avoidance." *Journal of Corporate Finance* 66 (2021), 1–17.

Baker, T., and S. J. Griffith. "Ensuring Corporate Misconduct: How Liability Insurance Undermines Shareholder Litigation.". Chicago, IL: University of Chicago Press, 2010.

Bao, Y.; B. Ke; B. Li; Y. J. Yu; and J. Zhang. "Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach." *Journal of Accounting Research* 58 (2020), 199–235.

Bentley, J. W.; T. E. Christensen; K. H. Gee; and B. C. Whipple. "Disentangling Managers' and Analysts' Non-GAAP Reporting." *Journal of Accounting Research* 56 (2018), 1039–1081.

Billings, M. B., and M. C. Cedergren. "Strategic Silence, Insider Selling and Litigation Risk." *Journal of Accounting and Economics* 59 (2015), 119–142.

Bird, A.; S. A. Karolyi; and T. G. Ruchti. "Information Sharing, Holdup, and External Finance: Evidence from Private Firms." *Review of Financial Studies* 32 (2019), 3075–3104.

Bourveau, T.; Y. Lou; and R. Wang. "Shareholder Litigation and Corporate Disclosure: Evidence from Derivative Lawsuits." *Journal of Accounting Research* 56 (2018), 797–842.

Bozanic, Z.; P. Choudhary; and K. J. Merkley. "Securities Law Expertise and Corporate Disclosure." *The Accounting Review* 94 (2019), 141–172.

Brabec, J.; T. Komárek; V. Franc; and L. Machlica. "On Model Evaluation Under Non-constant Class Imbalance." *Computational Science – International Conference on Computational Science*. Ed. by K. V. et al. Vol. 12140. Springer, Cham, 2020, 74–87.

Brabec, J., and L. Machlica. "Bad Practices in Evaluation Methodology Relevant to Class-Imbalanced Problems." *Critiquing and Correcting Trends in Machine Learning workshop at NeurIPS 2018* (2018). URL: https://arxiv.org/abs/1812.01388.

Brochet, F., and S. Srinivasan. "Accountability of Independent Directors: Evidence from Firms Subject to Securities Litigation." *Journal of Financial Economics* 111 (2014), 430–449.

Brockman, P.; I. K. Khurana; and X. Martin. "Voluntary Disclosures around Share Repurchases." *Journal of Financial Economics* 89 (2008), 175–191.

Brüggemann, U.; A. Kaul; C. Leuz; and I. M. Werner. "The Twilight Zone: OTC Regulatory Regimes and Market Quality." *Review of Financial Studies* 31 (2018), 898–942.

Burbridge, J. B.; L. Maggee; and A. L. Robb. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association* 83 (1988), 123–127.

Callen, J. L., and X. Fang. "Religion and Stock Price Crash Risk." *Journal of Financial and Quantitative Analysis* 50 (2015), 169–195.

Cassell, C. A.; L. M. Dreher; and L. A. Myers. "Reviewing the SEC's Review Process: 10-K Comment Letters and the Cost of Remediation." *The Accounting Review* 88 (2013), 1875–1908.

Dechow, P. M.; R. G. Sloan; and A. P. Sweeney. "Detecting Earnings Management." *The Accounting Review* 70 (2 1995), 193–225.

Dimmock, S. G., and W. C. Gerken. "Predicting Fraud by Investment Managers." *Journal of Financial Economics* 105 (2012), 153–173.

Donelson, D. C.; J. J. Hopkins; and C. G. Yust. "The Role of Directors' and Officers' Insurance in Securities Fraud Class Action Settlements." *Journal of Law and Economics* 58 (2015), 747–778.

Donelson, D. C.; J. J. Hopkins; and C. G. Yust. "The Cost of Disclosure Regulation: Evidence from D&O Insurance and Nonmeritorious Securities Litigation." *Review of Accounting Studies* 23 (2018), 528–588.

Donelson, D. C.; A. Kartapanis; J. M. McInnis; and C. G. Yust. "Measuring Accounting Fraud and Irregularities Using Public and Private Enforcement." *The Accounting Review* 96 (2021), 183–213.

Donelson, D. C.; J. M. McInnis; R. D. Mergenthaler; and Y. Yu. "The Timeliness of Bad Earnings News and Litigation Risk." *The Accounting Review* 87 (2012), 1967–1991.

Donelson, D. C., and C. G. Yust. "Litigation Risk and Agency Costs: Evidence from Nevada Corporate Law." *Journal of Law and Economics* 57 (2014), 747–780.

Ellul, A., and V. Yerramilli. "Stronger Risk Controls, Lower Risk: Evidence from U.S. Bank Holding Companies." *Journal of Finance* 68 (2013), 1757–1803.

Eraker, B., and M. Ready. "Do Investors Overpay for Stocks with Lottery-Like Payoffs? An Examination of the Returns of OTC Stocks." *Journal of Financial Economics* 115 (2015), 486–504.

Fedderke, J. W., and M. Ventoruzzo. "Do Conservative Judges Favor Wall Street? Ideology and the Supreme Court's Securities Regulation Decisions?." *Florida Law Review* 67 (2016), 1211–1280.

Field, L.; M. Lowry; and S. Shu. "Does Disclosure Deter or Trigger Litigation?." *Journal of Accounting and Economics* 39 (2005), 487–507.

Files, R.; E. P. Swanson; and S. Tse. "Stealth Disclosure of Accounting Restatements." *The Accounting Review* 84 (2009), 1495–1520.

Francis, J.; D. Philbrick; and K. Schipper. "Shareholder Litigation and Corporate Disclosures." *Journal of Accounting Research* 32 (1994), 137–164.

Freeman, R. N., and S. Y. Tse. "A Nonlinear Model of Security Price Responses to Unexpected Earnings." *Journal of Accounting Research* 30 (1992), 185–209.

Gande, A., and C. M. Lewis. "Shareholder-Initiated Class Action Lawsuits: Shareholder Wealth Effects and Industry Spillovers." *Journal of Financial and Quantitative Analysis* 44 (2009), 823–850.

Gosnell, T.; A. J. Keown; and J. M. Pinkerton. "Bankruptcy and Insider Trading: Differences Between Exchange-Listed and OTC Firms." *Journal of Finance* 47 (1992), 349–362.

He, J., and M. A. Plumlee. "Measuring Disclosure Using 8-K Filings." *Review of Accounting Studies* 25 (2020), 903–962.

Hennes, K. M.; A. J. Leone; and B. P. Miller. "The Importance of Distinguishing Errors from Irregularities in Restatement Research: The Case of Restatements and CEO/CFO Turnover." *The Accounting Review* 83 (2008), 1487–1519.

Holzman, E.; N. T. Marshall; and B. Schmidt. "Who's on the Hot Seat for an SEC Investigation?." *Unpublished working paper* (2020). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3223815.

Hosmer, D. W., and S. Lemeshow. "Assessing the Fit of the Model.". Applied Logistic Regression. John Wiley & Sons, 2000.

Houston, J. F.; C. Lin; S. Liu; and L. Wei. "Litigation Risk and Voluntary Disclosure: Evidence from Legal Changes." *The Accounting Review* 94 (2019), 247–272.

Huang, A.; K. W. Hui; and R. Z. Li. "Federal Judge Ideology: A New Measure of Ex Ante Litigation Risk." *Journal of Accounting Research* 57 (2019), 431–489.

Hutton, A.; S. Shu; and X. Zheng. "Regulatory Transparency and the Alignment of Private and Public Enforcement: Evidence from the Public Disclosure of SEC Comment Letters." *Journal of Financial Economics* 145 (2022), 297–321.

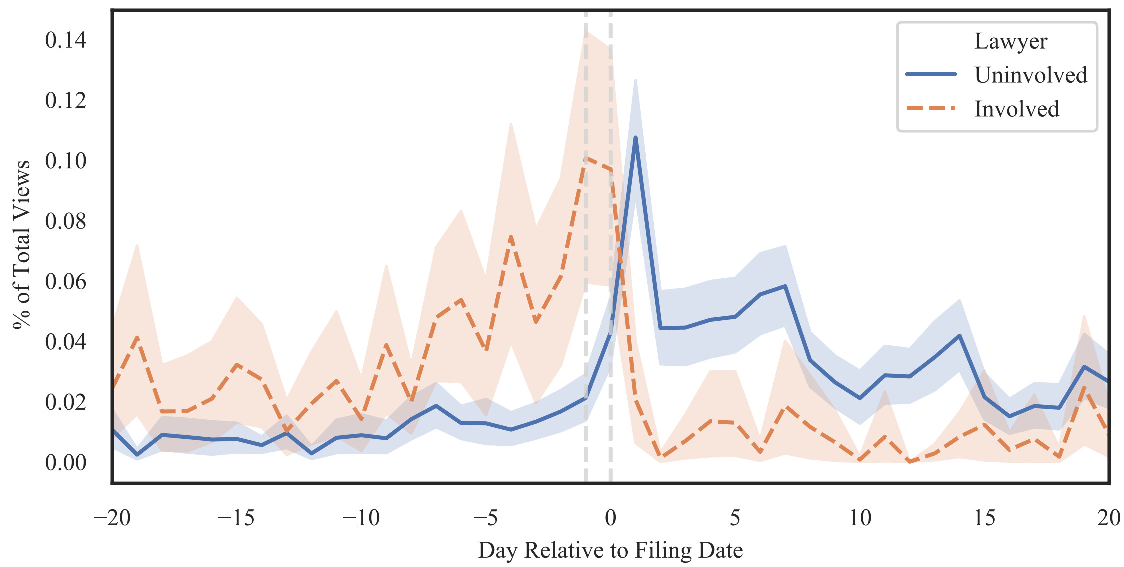
Jiang, J.; K. R. Petroni; and I. Y. Wang. "CFOs and CEOs: Who have the Most Influence on Earnings Management?." *Journal of Financial Economics* 96 (2010), 5137–526.
Johnson, M. F.; K. K. Nelson; and A. C. Pritchard. "Do the Merits Matter More? The Impact of the Private Securities Litigation Reform Act." *Journal of Law, Economics, and Organization* 23 (2007), 627–652.
Johnson, N. L. "Systems of Frequency Curves Generated by Methods of Translation." *Biometrika* 36 (1949), 149–176.
Karmel, R. S. "The Future of Corporate Governance Listing Requirements." *SMU Law Review* 54 (2001), 325–356.
Kim, I., and D. J. Skinner. "Measuring Securities Litigation Risk." *Journal of Accounting and Economics* 53 (2012), 290–310.
Larcker, D. F., and A. A. Zakolyukina. "Detecting Deceptive Discussion in Conference Calls." *Journal of Accounting Research* 50 (2012), 495–540.
Liljeblom, E.; D. Pasternack; and M. Rosenberg. "What Determines Stock Option Contract Design." *Journal of Financial Economics* 102 (2011), 293–316.
Nelson, K. K., and A. C. Pritchard. "Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors." *Journal of Empirical Legal Studies* 13 (2016), 266–297.
Newton, N. J.; J. S. Persellin; D. Wang; and M. S. Wilkins. "Internal Control Opinion Shopping and Audit Market Competition." *The Accounting Review* 91 (2016), 603–623.
Pagach, D., and R. Warr. "The Characteristics of Firms that Hire Chief Risk Officers." *Journal of Risk and Insurance* 78 (2011), 185–211.
Rogers, J. L.; A. Van Buskirk; and S. L. Zechman. "Disclosure Tone and Shareholder Litigation." *The Accounting Review* 86 (2011), 2155–2183.
Ryan, J. "Using the EDGAR Log File Data Set." *Unpublished working paper* (2017).
Saito, T., and M. Rehmsmeier. "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets." *PLoS ONE* 10 (2015), 1–21.
Skinner, D. J. "Earnings Disclosures and Stockholder Lawsuits." *Journal of Accounting and Economics* 23 (1997), 249–282.
Swamidass, S. J.; C.-A. Azencott; K. Daily; and P. Baldi. "A CROC Stronger than ROC: Measuring, Visualizing and Optimizing Early Retrieval." *Bioinformatics* 26 (2010), 1348–1356.
Tan, C. E. L., and S. M. Young. "An Analysis of "Little r" Restatements." *Accounting Horizons* 29 (2015), 667–693.

FIGURE A.1

Percent of Daily Views around Relevant Litigation Dates—
Examining Views for Firms Involved and Not Involved in the Litigation Filing

Graphs A and B of Figure A.1 present the average percent of daily views in the 20 days before and after the filing date (Graph A) and class period end (Graph B) as per the first identified securities complaint (FIC). The percentages are calculated by summing up views within each securities class action, separately for firms involved and not involved, and scaling daily views by the sum. As such, by construction, we only include cases for which there was at least one view by the respective lawyers. The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval calculated using bootstrap resampling. Refer to Table A.2 Panel A for the case-level sample composition.

Graph A. Filing Date



Graph B. Class Period End

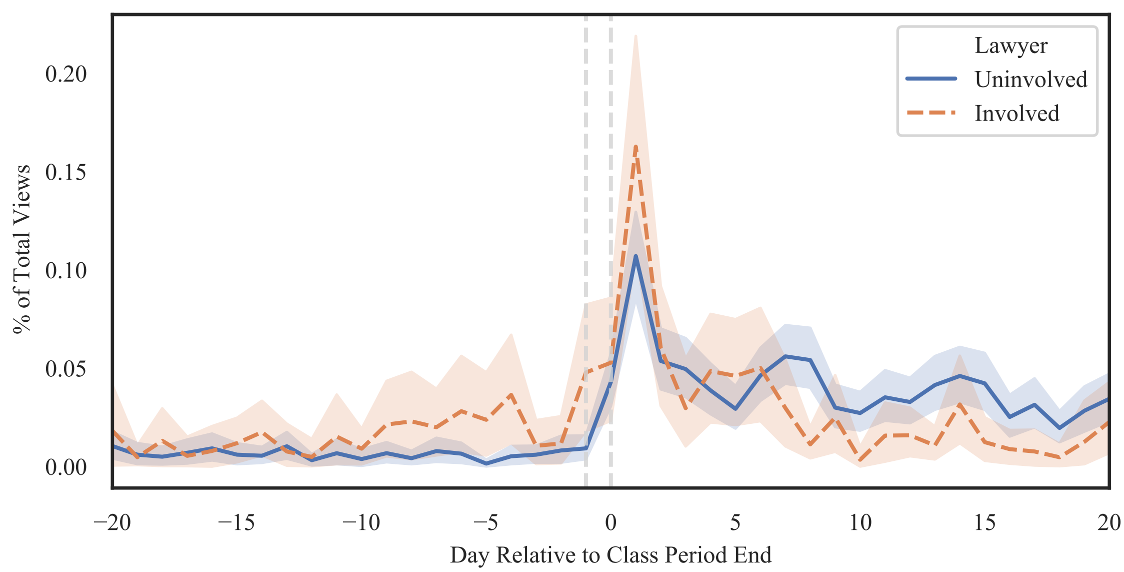
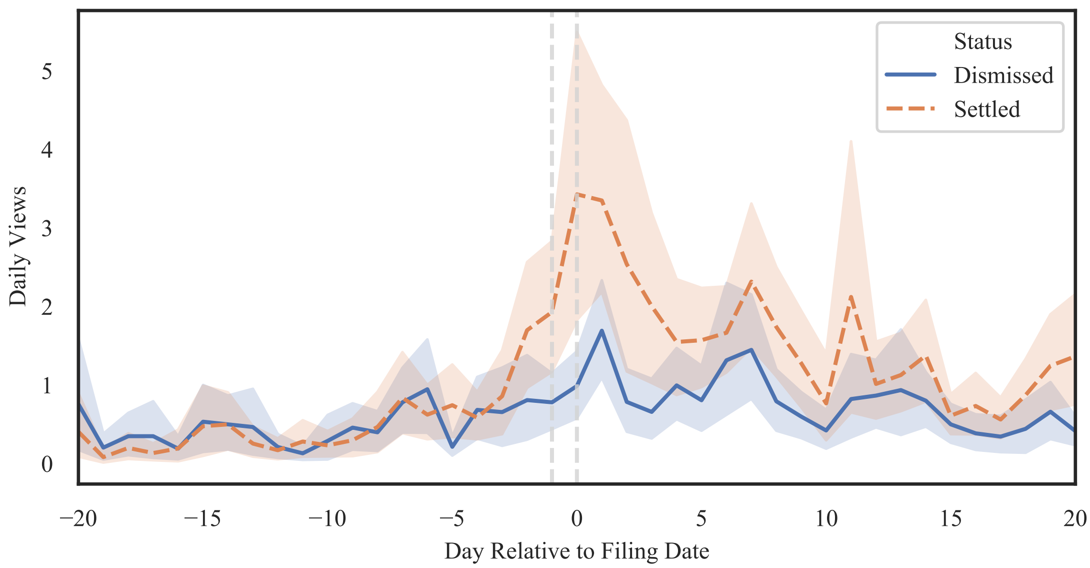


FIGURE A.2

Daily Views of EDGAR Forms around Relevant Litigation Dates for Settled versus Dismissed Cases

Graphs A and B of Figure A.2 present the average daily views in the 20 days before and after the filing date (Graph A) and class period end (Graph B) as per the first identified securities complaint (FIC) for settled versus dismissed litigation. The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval. Refer to Table A.2 Panel A for the case-level sample composition.

Graph A. Filing Date



Graph B. Class Period End

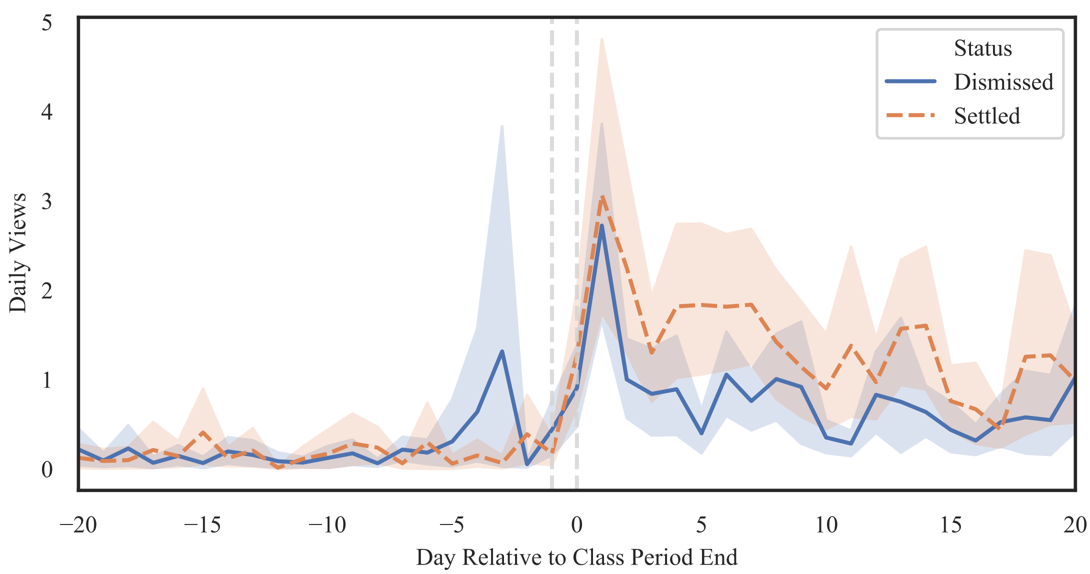
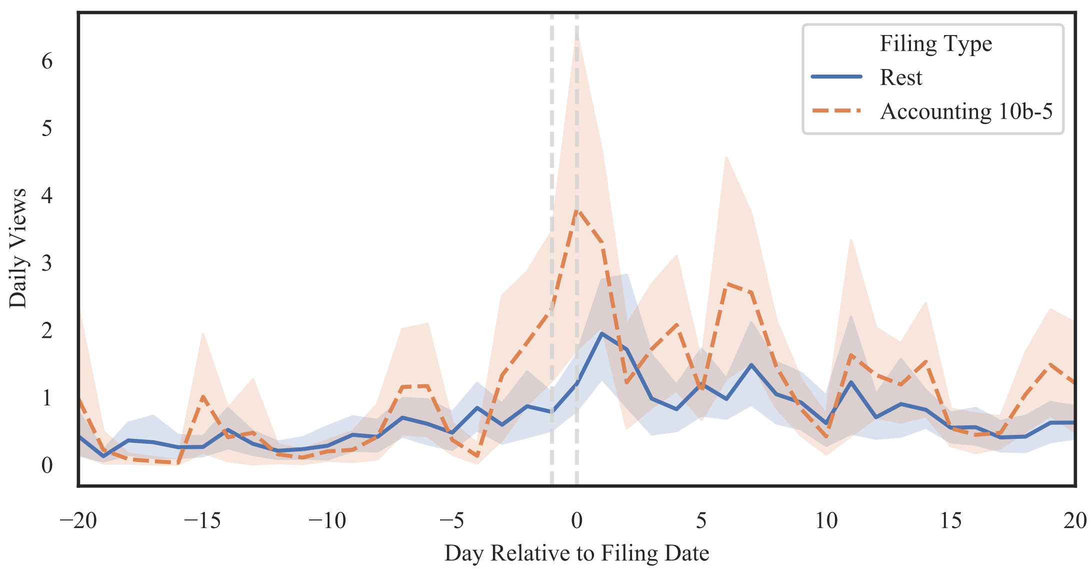


FIGURE A.3

Daily Views of EDGAR Forms around Relevant Litigation Dates for Accounting 10b-5 versus Non-Accounting 10b-5 Cases

Graphs A and B of Figure A.3 present the average daily views in the 20 days before and after the filing date (Graph A) and class period end (Graph B) as per the first identified securities complaint (FIC) for accounting 10b-5 cases versus other cases. The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval. Refer to Table A.2 Panel A for the case-level sample composition.

Graph A. Filing Date



Graph B. Class Period End

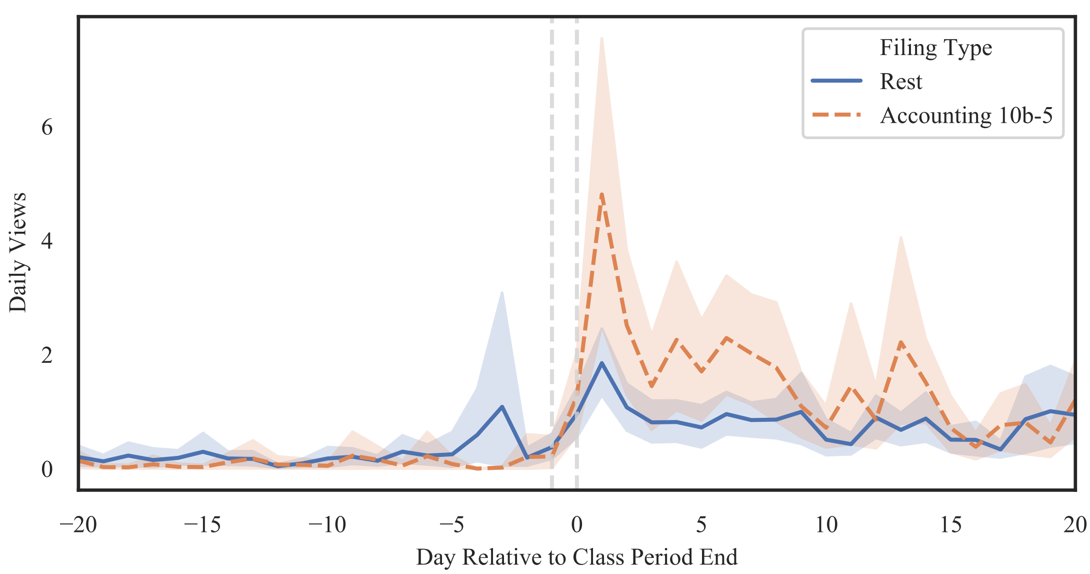
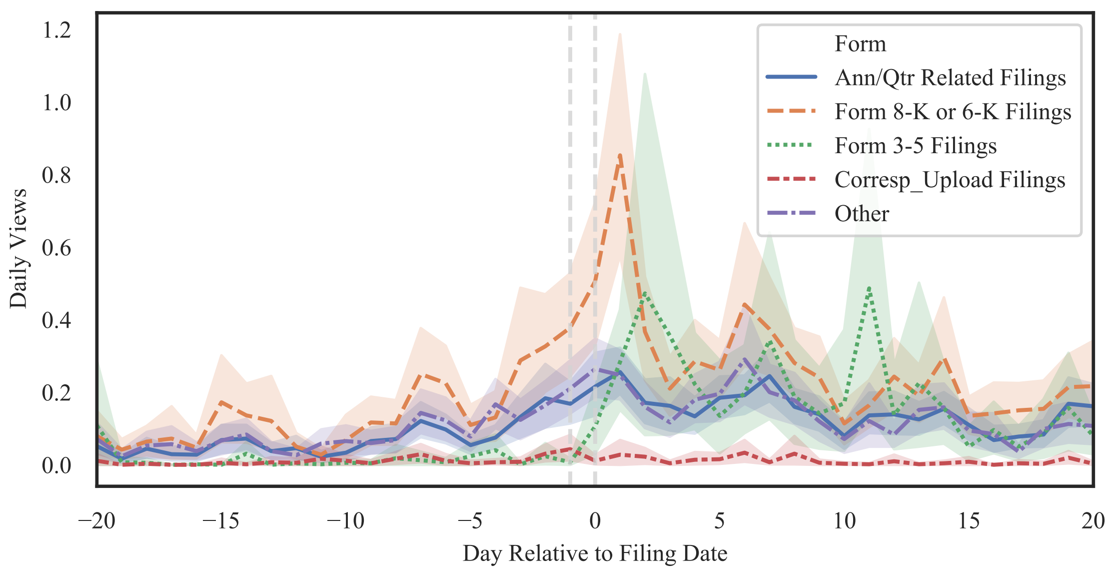


FIGURE A.4

Daily Views of EDGAR Forms around Relevant Litigation Dates by SEC Filing Type

Graphs A and B of Figure A.4 present the average daily views in the 20 days before and after the filing date (Graph A) and class period end (Graph B) as per the first identified securities complaint (FIC). To account for the fact that some SEC filings have multiple files submitted by firms (e.g., an 8-K that includes both earnings announcements and financial exhibits), we include only one view per IP-CIK-Date-Accession Number (i.e., overall SEC filing). The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval. Refer to Table A.2 Panel A for the case-level sample composition.

Graph A. Filing Date



Graph B. Class Period End

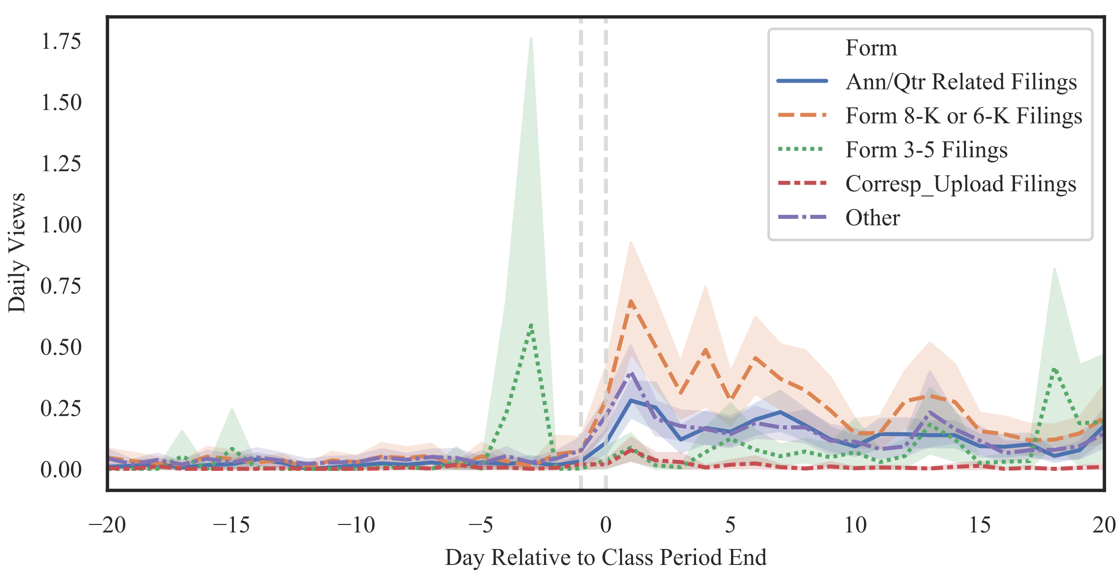


FIGURE A.5

Daily Views of EDGAR Forms around the Filing Date for the Top Five Most Viewed Types of 8-Ks

Figure A.5 presents the average daily views for the top 5 most viewed 8-K item codes in the 20 days before and after the filing date as per the first identified securities complaint (FIC). To account for the fact that some SEC filings have multiple files submitted by firms (e.g., an 8-K that includes both earnings announcements and financial exhibits), we include only one view per IP-CIK-Date-Accession Number (i.e., overall SEC filing). We note that each 8-K may contain multiple item codes. The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval. Refer to Table A.2 Panel A for the case-level sample composition.

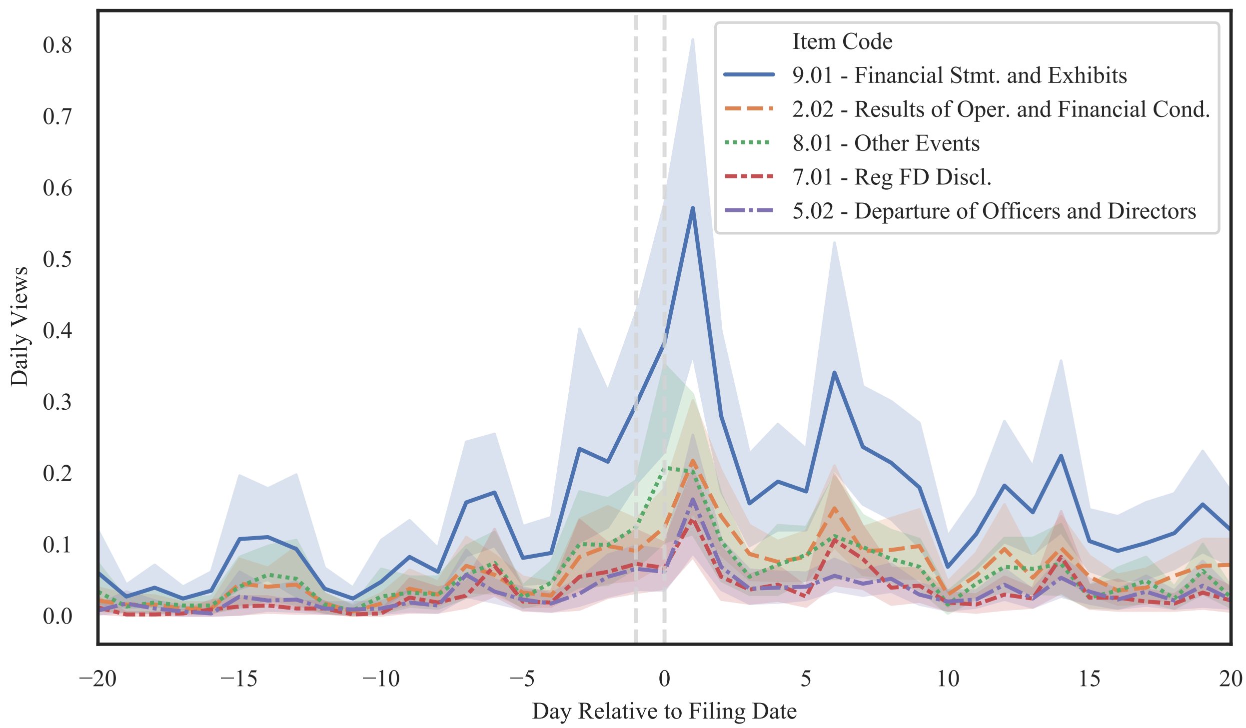
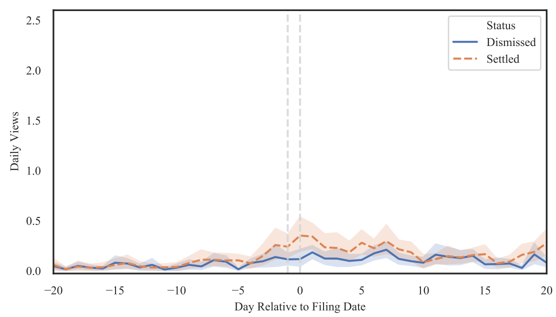


FIGURE A.6

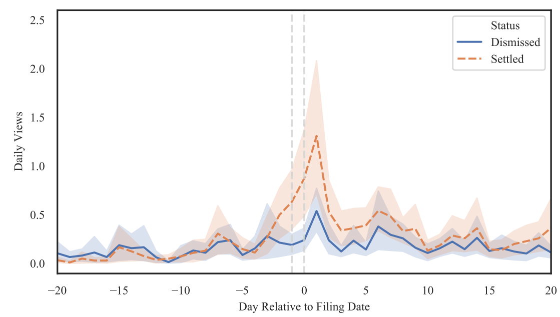
Daily Views of EDGAR Forms around the Filing Date by SEC Filing Type for Settled versus Dismissed Cases

Graphs A through E of Figure A.6 present the average daily views in the 20 days before and after the filing date as per the first identified securities complaint (FIC) for settled versus dismissed cases based on the type of SEC filing viewed. To account for the fact that some SEC filings have multiple files submitted by firms (e.g., an 8-K that includes both earnings announcements and financial exhibits), we include only one view per IP-CIK-Date-Accession Number (i.e., overall SEC filing). The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval. Refer to Table A.2 Panel A for the case-level sample composition.

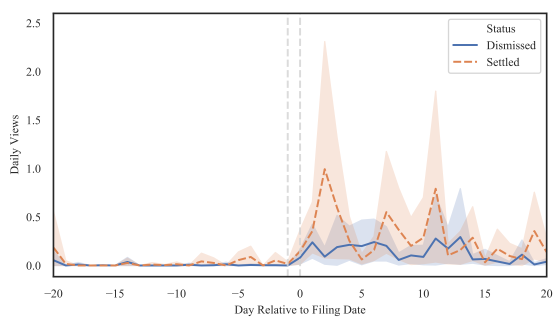
Graph A. Annual/Quarterly Filings



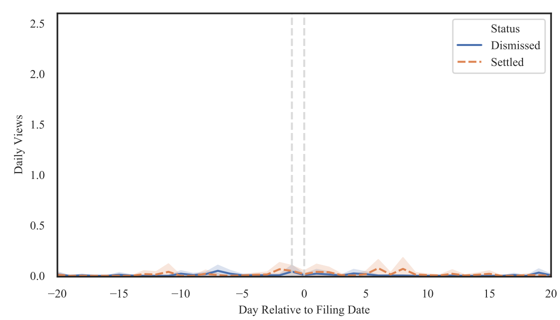
Graph B. 8-Ks and 6-Ks



Graph C. Insider Trading



Graph D. SEC Correspondence



Graph E. Other

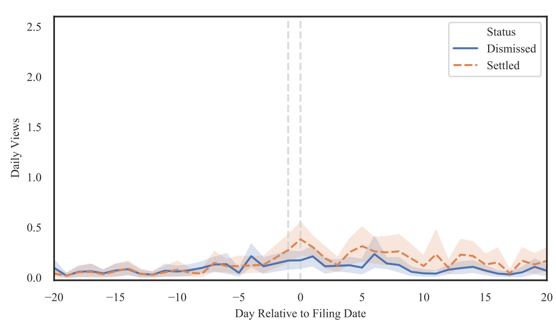
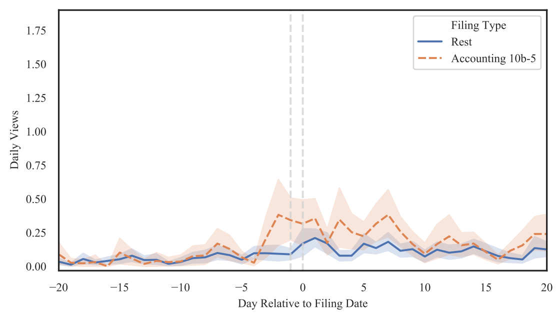


FIGURE A.7

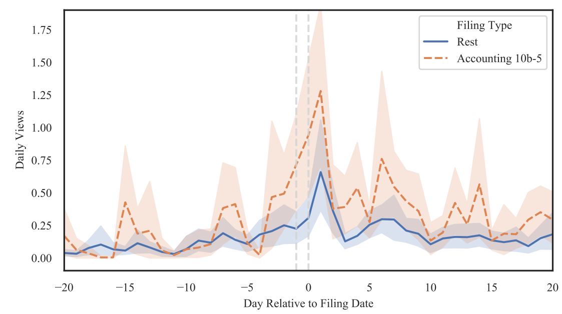
Daily Views of EDGAR Forms around the Filing Date by SEC Filing Type for Accounting 10b-5 versus Non-Accounting 10b-5 Cases

Graphs A through E of Figure A.7 present the average daily views in the 20 days before and after the filing date as per the first identified securities complaint (FIC) for litigation filings that include accounting 10b-5 cases versus other cases based on the type of SEC filing viewed. To account for the fact that some SEC filings have multiple files submitted by firms (e.g., an 8-K that includes both earnings announcements and financial exhibits), we include only one view per IP-CIK-Date-Accession Number (i.e., overall SEC filing). The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval. Refer to Table A.2 Panel A for the case-level sample composition.

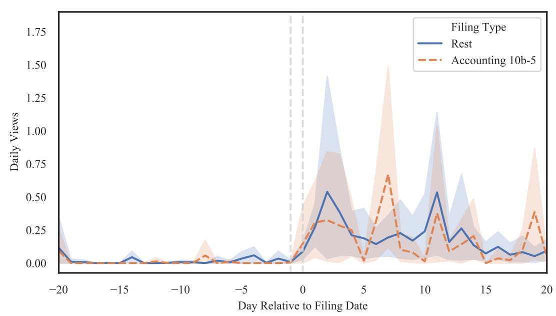
Graph A. Annual/Quarterly Filings



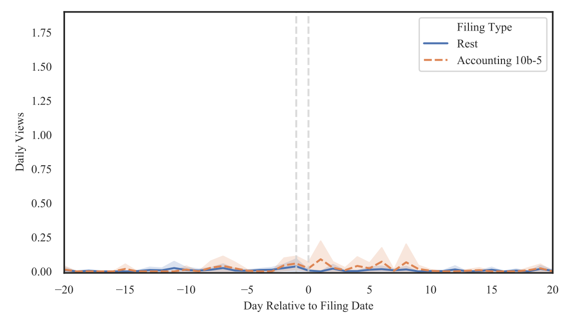
Graph B. 8-Ks and 6-Ks



Graph C. Insider Trading



Graph D. SEC Correspondence



Graph E. Other

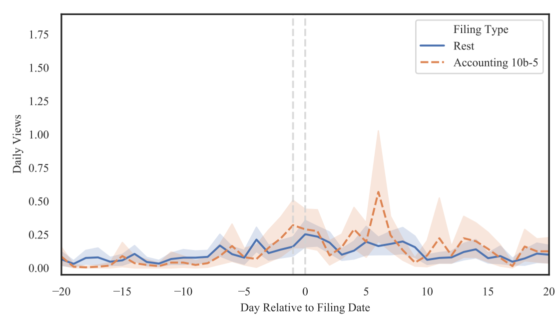
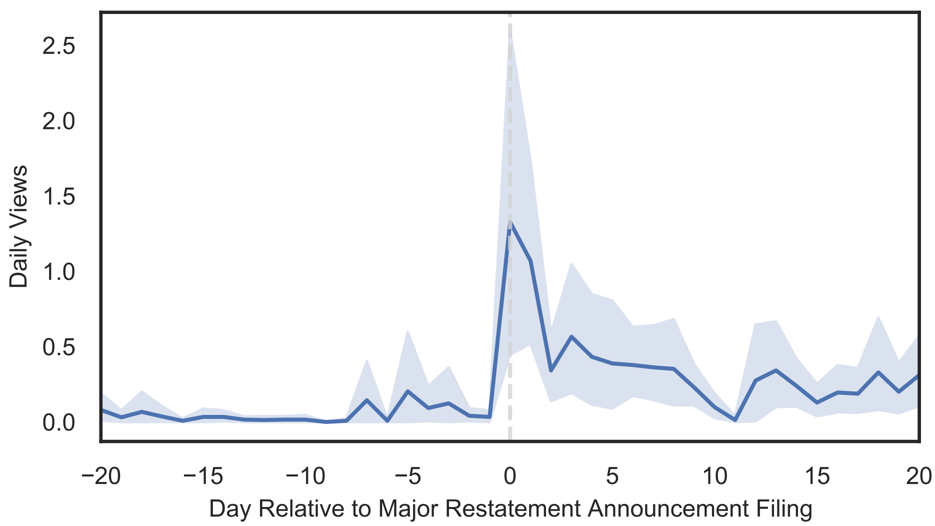


FIGURE A.8

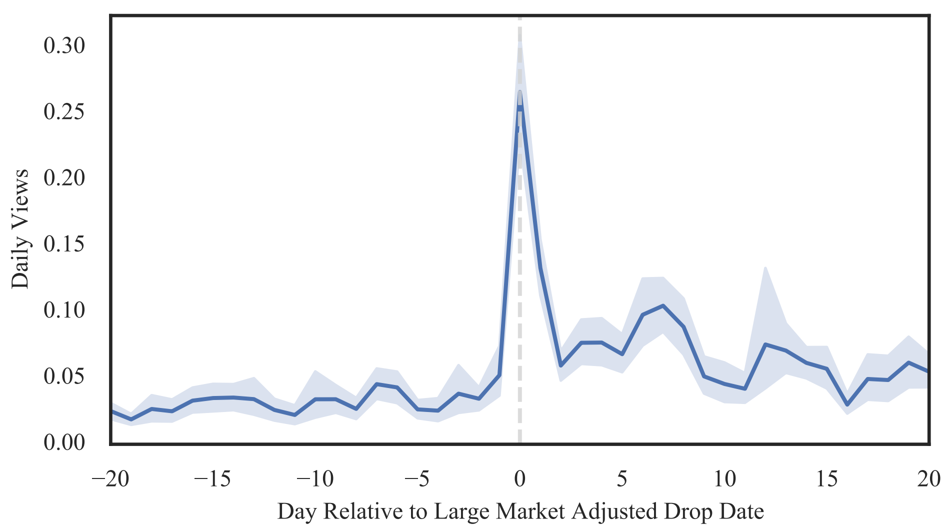
Daily Views around Major Corporate Events

Graphs A through D of Figure A.8 present daily views, using 20 days before and after: 1) major restatement announcements (Graph A), 2) at least –10% market-adjusted return dates (Graph B), 3) earnings announcements (Graph C), and 4) announcements of internal control weaknesses (ICWs) under Sarbanes-Oxley Act (SOX) Sections 302 or 404 (Graph D). For ICWs filed pursuant to Section 404, we require that the opinion is issued by the auditor, rather than management. The vertical grey lines are at days -1 and 0. The shaded areas present a 95% confidence interval calculated using bootstrap resampling. To be consistent with our case level analyses figures, the sample includes events with filing dates 01/01/2012 through 12/31/2016, which can be linked to Compustat and have available CRSP and CIK identifiers, but no additional data requirements are imposed. That is, other than the period adjustment, the starting population for this test is similar in nature to Table A.2 Panel B before requiring Kim and Skinner (2012) variables and accrual-related data.

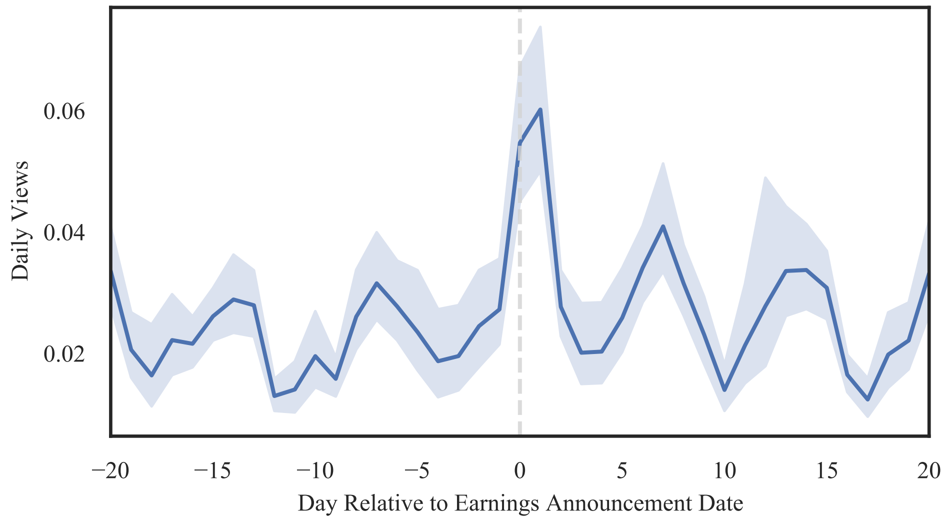
Graph A. Major Restatement Announcements



Graph B. At Least –10% Market-Adjusted Return Dates



Graph C. Earnings Announcements



Graph D. Internal Control Weakness Announcements

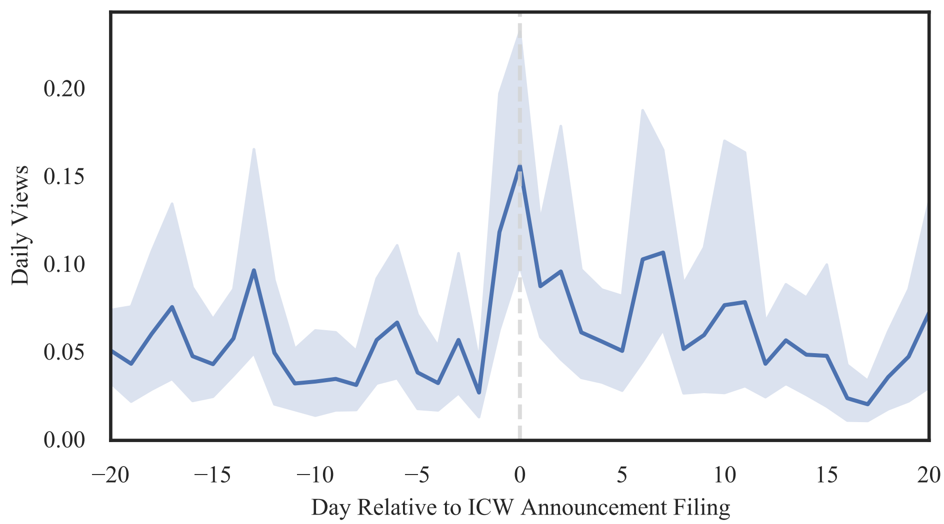
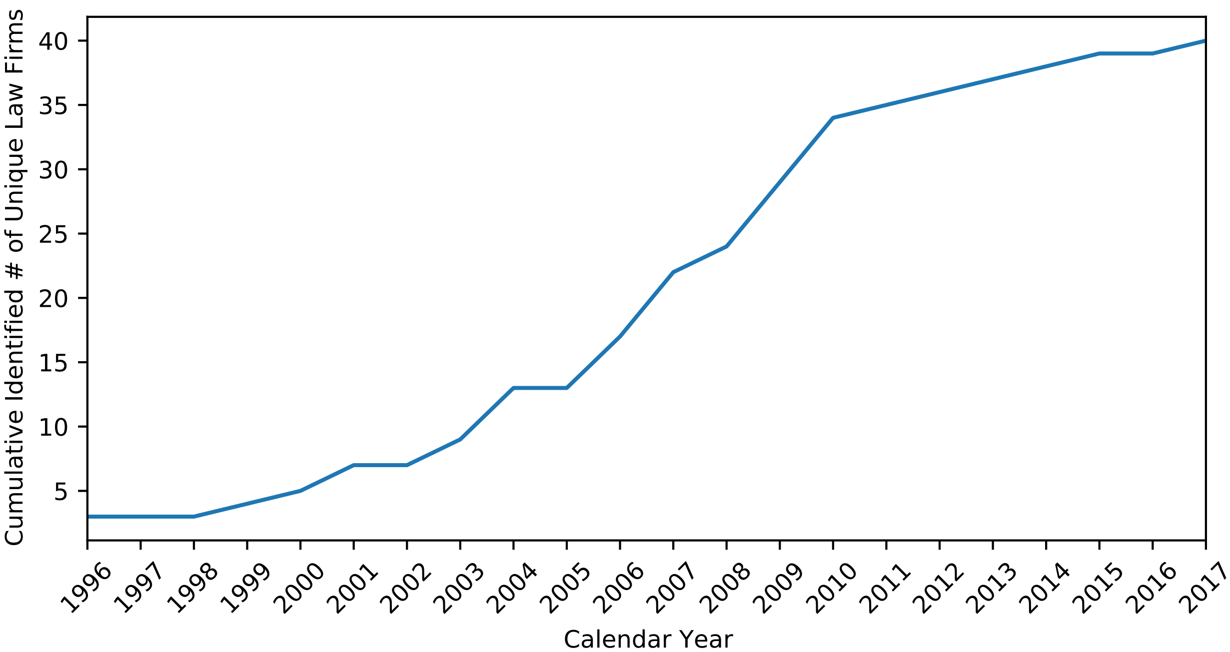


FIGURE A.9

Plaintiff-lawyer IP Addresses and Views Over Time

Graphs A through C of Figure A.9 present how the ability to identify plaintiff-lawyer IP addresses and views change over time. Graph A shows the cumulative number of unique plaintiff lawyers by year for which we are able to identify at least one IP address. Graph B of Figure A.9 presents the percentage of securities class actions by year for which we can identify the IP address for at least one of the plaintiffs' lawyers involved in the initial complaint. Graph C of Figure A.9 presents the total number of plaintiff-lawyer views by year.

Graph A. Number of Identified Unique Plaintiff Law Firms with IP Addresses



Graph B. Percent of Securities Class Actions with Identified Plaintiff-Lawyer IP Address

Graph C. Total Yearly Plaintiff-Lawyer Views

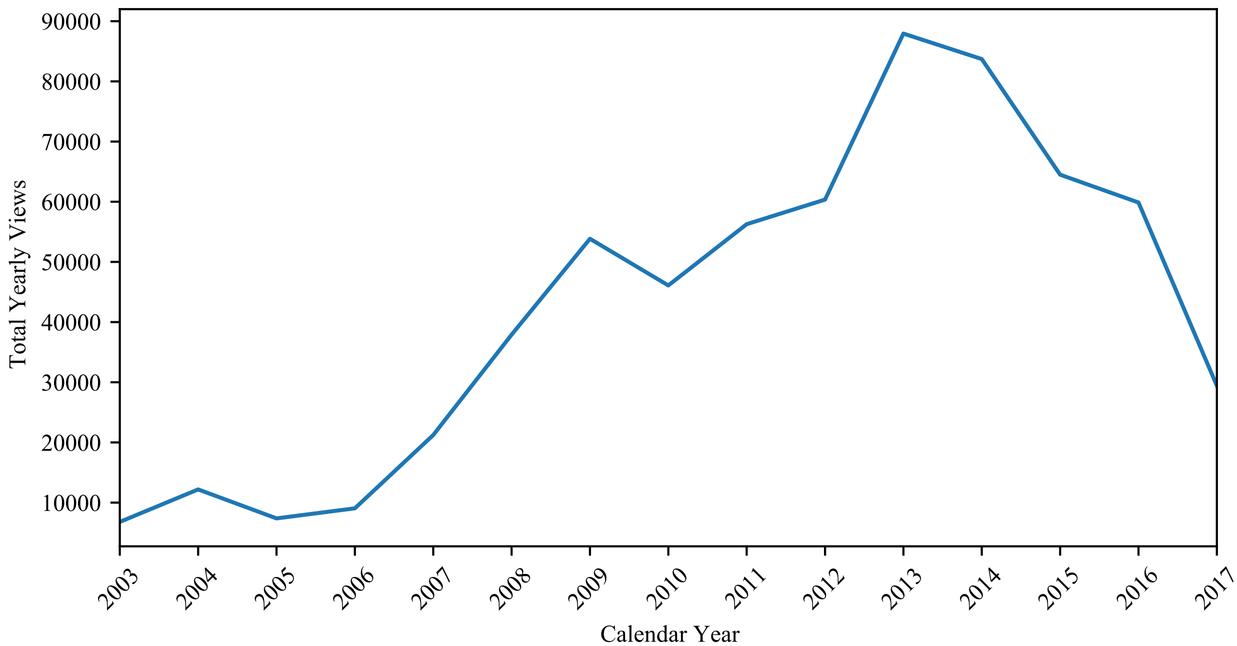
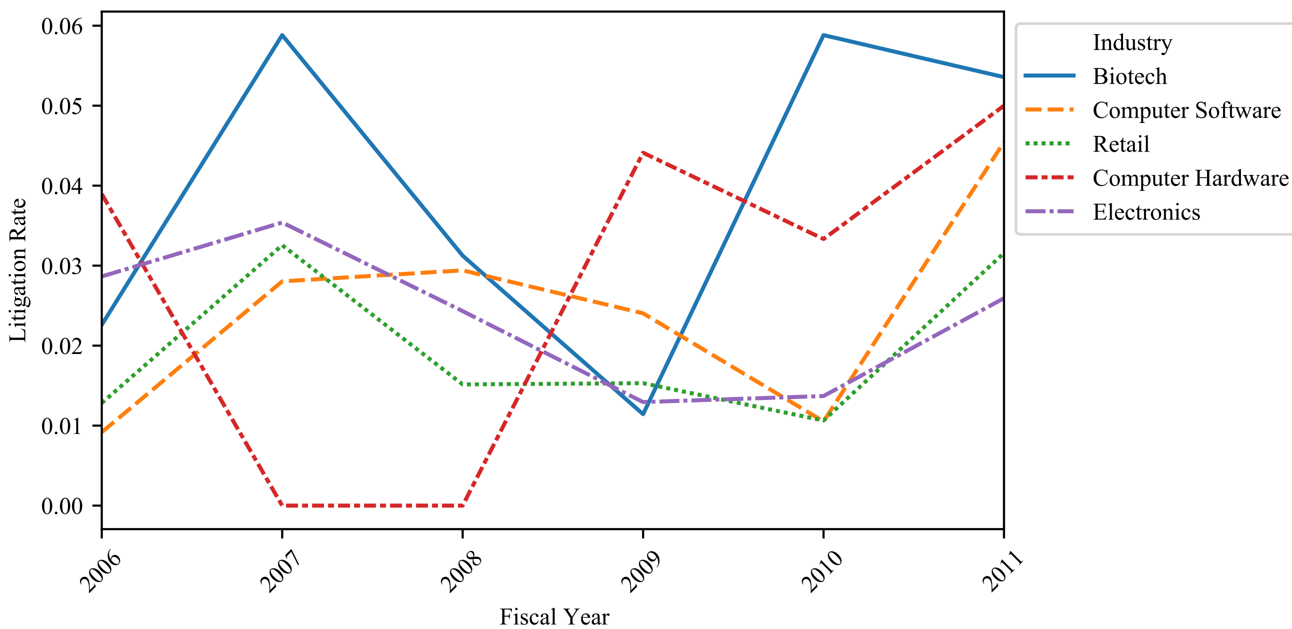


FIGURE A.10

Litigation by High Risk Industries Prior to the Sample Period

Graphs A and B of Figure A.10 present litigation rates across industries identified by Brochet and Srinivasan (2014) during the period preceding our main sample (i.e., fiscal years 2006-2011). Graph A presents litigation rates by industry over time, and Figure Graph B by industry over the whole period.

Graph A. Yearly Litigation Filings



Graph B. Filings by Industry

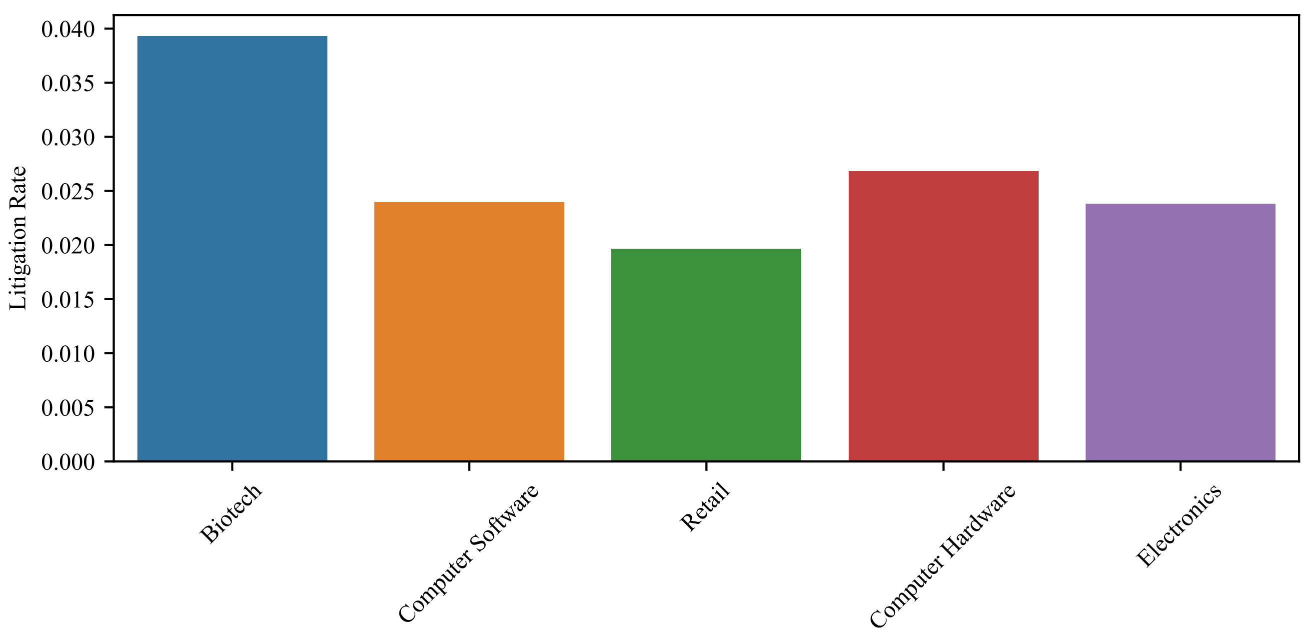


FIGURE A.11

ROC Curve for Models 1 and 2 from Table 6 Panel A

Figure A.11 presents ROC curves for Columns 1 and 2 from Table 6 Panel A.

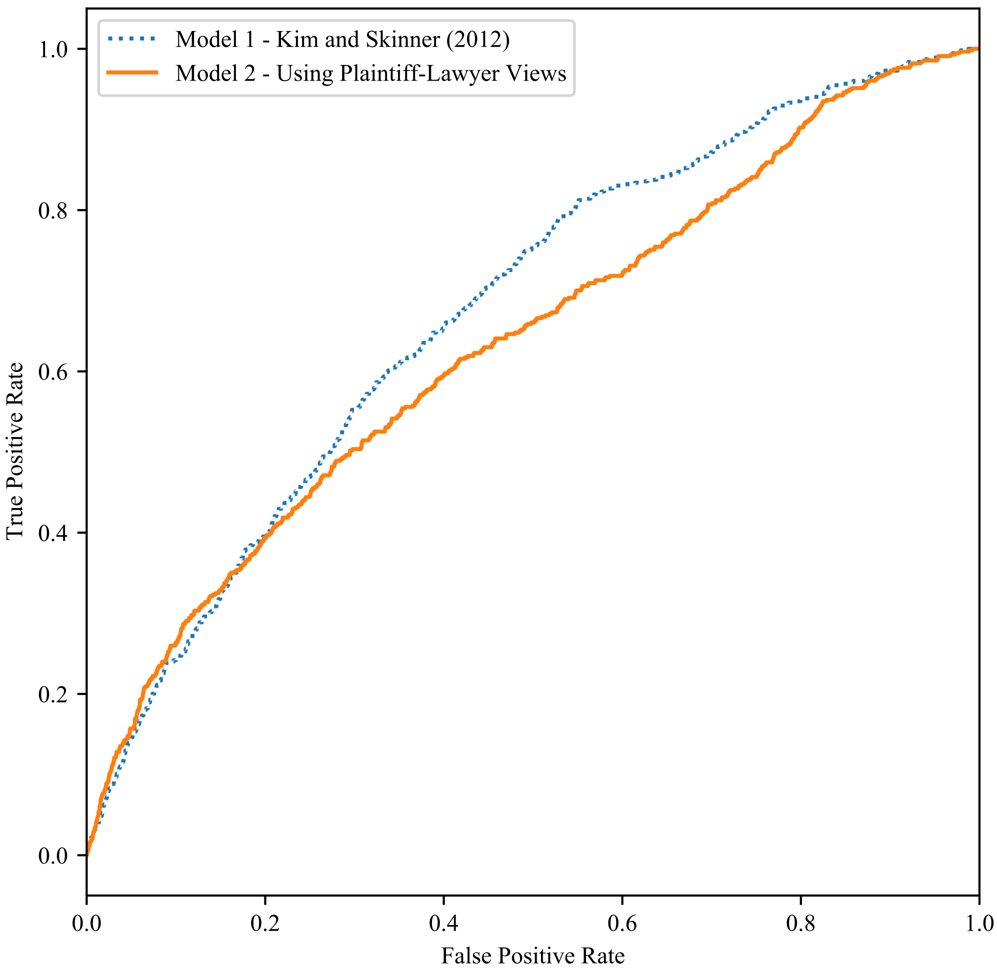


FIGURE A.12

ROC Curve for Models 4 and 5 from Table 6 Panel A

Figure A.12 presents ROC curves for Columns 4 and 5 from Table 6 Panel A.

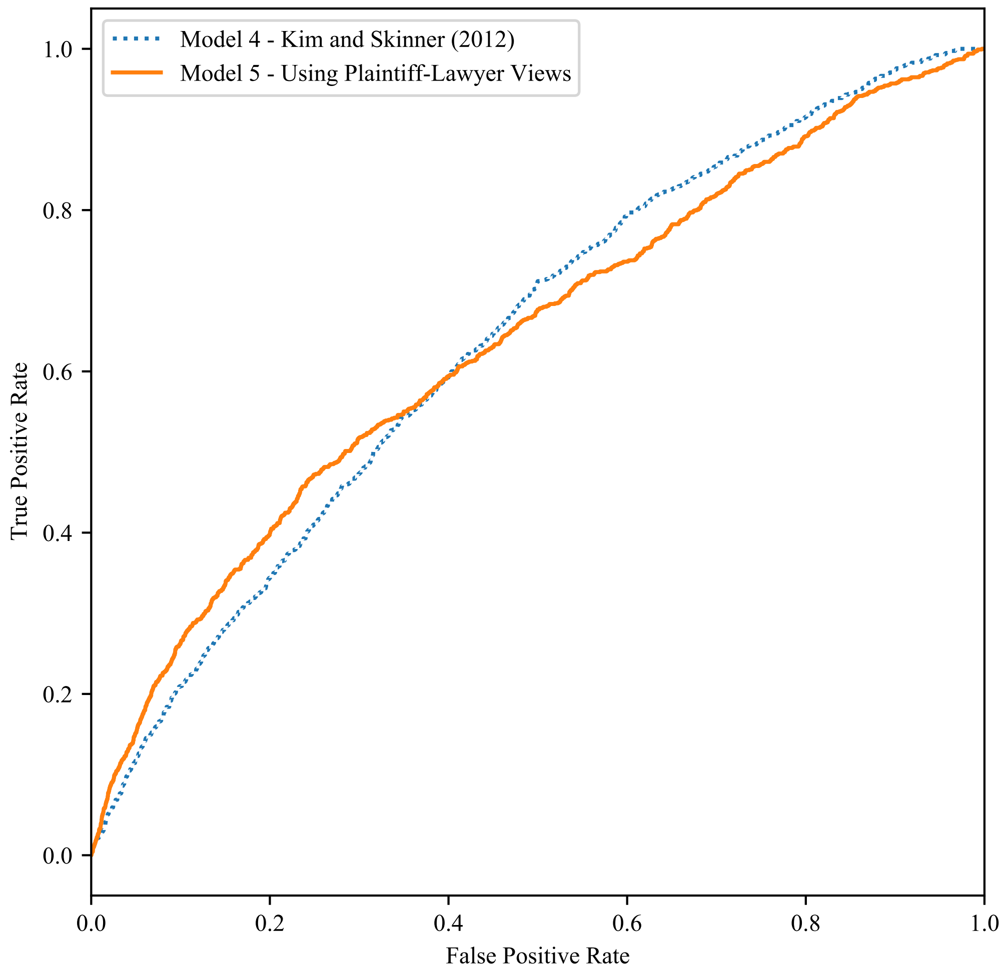


TABLE A.1

Most Active Plaintiffs' Law Firms

Table A.1 lists the name and number of cases that plaintiffs' law firms participated in from 01/01/2012 to 12/31/2016 to provide insight into the plaintiffs' law firms that were most active over the sample period. No other filters are applied (e.g., there is no requirement that the case can be linked to Compustat and CRSP). If a law firm changes names (e.g., due to the addition of a new partner), it is only listed and counted once. We identify IPs by searching for plaintiffs' law firms ranked in the top 50 in terms of their involvement in securities cases, based on either first identified or reference complaints for the period starting Jan. 1, 2001, up to March 11, 2020, and that were active until at least 2006. We then search for registered IPs for these plaintiffs' law firms. We indicate plaintiffs' law firms for which at least one IP was identified through this process. Because we collect IPs based on the most active plaintiffs' law firms over a longer time period, we note that we have collected IPs for additional firms not presented in the table. For brevity, we only tabulate plaintiffs' law firms that were involved in either more than: 1) 30 first identified complaints (FIC) or 2) 30 reference complaints (REF).

Combined Name	FIC	REF	IPs Identified Indicator
Pomerantz LLP	188	109	1
Robbins Geller Rudman & Dowd LLP	186	166	1
The Rosen Law Firm	143	112	1
Glancy Binkow & GoldBerg LLP	68	26	1
Glancy Prongay & Murray LLP	46	56	0
Bronstein, Gewirtz & Grossman, LLC	43	23	0
Levi & Korsinsky	43	53	1
Law Offices of Howard G. Smith	39	5	0
Faruqi & Faruqi LLP	32	21	1
Bernstein Litowitz Berger & Grossmann LLP	21	58	1
Labaton Sucharow LLP	14	65	1
Kessler Topaz Meltzer & Check, LLP	6	34	1

TABLE A.2
Sample Construction

Table A.2 presents the sample construction. Panel A presents the case-level sample construction in identifying securities class actions filed between 01/01/2012 and 12/31/2016 for which we can identify IPs for at least one of the plaintiffs' lawyers involved in the first identified complaint (FIC). For 198 of the cases, we either cannot identify a registered IP address for the plaintiffs' lawyers participating in the filing or did not search for the IPs for those plaintiffs' law firms as they did not participate in a large number of securities litigation cases as per our screening criteria. The majority of the excluded cases are due to three plaintiffs' law firms for which we are unable to locate a registered IP address. We also exclude 11 cases for which the registered plaintiffs' law firms' IP that we identified never accessed EDGAR. Finally, we exclude three cases because the filing date of the securities class action precedes the registration date of the IP, so we cannot ensure that any EDGAR views from the IP are from the plaintiffs' lawyers. Panel B presents the construction of the firm-year level sample used in the analyses. The final sample consists of all firm-years with fiscal years between 2012 and 2016 that pass the usual screening criteria, including required Compustat and CRSP data. Panel C presents the construction of the firm-quarter level sample used in the stock market outcome analyses. The final sample consists of all firm-quarters with fiscal years between 2012 and 2016 that pass the usual screening criteria, including required Compustat and CRSP data.

Panel A. Securities Class Action Sample Construction

Description	Obs
Cases filed between 2012 and 2016	963
Less missing Compustat identifier or CIK	31
	932
Less cases for which the IPs were not identified, or not searched for, for any of the lawyers involved in the FIC	198
Less cases for which identified IPs never accessed EDGAR since registerd (using all EDGAR log files)	11
Less cases for which earliest identified IP was registered after the filing date	3
Final sample size	720

Panel B. Firm-Year Sample Construction

Description	Obs
Firm-years with fiscal years 2012 to 2016	47,143
Less firm-years missing CIK info	8,610
Less firm-years missing Kim and Skinner (2012) related data	17,115
Less firm-years missing accruals related data and other Compustat data	4,239
Final sample size	17,179

Panel C. Firm-Quarter Sample Construction

Description	Obs
Firm-quarters with fiscal years 2012 to 2016	189,446
Less firm-quarters missing CIK info	33,231
Less firm-quarters missing Compustat or returns related data	65,079
Final sample size	91,136

TABLE A.3

Performance Evaluation Metrics

Table A.3 presents explanatory information regarding the intuition and identification of false positives and negatives and true positives and negatives in our setting (Panel A) and the calculation of the performance evaluation metrics other than AUC using those classifications (Panel B). The actual condition is identified by whether a firm has realized litigation in the following year as measured using either $\text{SUED}_{i,t}$ or $\text{SUED\_INV}_{i,t}$ as a proxy for high litigation risk; firm-years that subsequently have realized litigation risk are classified as high litigation risk, while those that do not are classified as low litigation risk. The predicted condition is identified by whether the predicted litigation risk for the firm-year using the relevant model is in the top 3% of observations; these firm-years are classified as high litigation risk, while others are classified as low litigation risk. The false negative rate can be calculated as 1 - sensitivity.

Panel A. Classification Matrix

	Predicted Condition	
	Low Litigation Risk	High Litigation Risk
Actual Condition		
Low Litigation Risk	True Negative (TN)	False Positive (FP)
High Litigation Risk	False Negative (FN)	True Positive (TP)

Panel B. Metric Calculations

Performance Metrics:	
Sensitivity	TP / (TP + FN)
Precision	TP / (TP + FP)
Specificity	TN / (TN + FP)

TABLE A.4

Predicting Realized Litigation Risk over a Longer Period

Table A.4 presents results similar to Columns 1–3 of Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings in the firm-year sample from 2008–2016. Column 1 presents results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Column 2 presents results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Column 3 presents results based on combining the two groups of variables other than $FPS_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2	3
	$SUED_{i,t}$	$SUED_{i,t}$	$SUED_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$		0.25***	0.21***
		(0.04)	(0.05)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$		0.21***	0.17***
		(0.03)	(0.03)
$FPS_{i,t}$	0.15*		
	(0.08)		
$LN\_ASSETS_{i,t-1}$	0.16***	0.09***	0.13***
	(0.02)	(0.02)	(0.02)
$SALES\_GR_{i,t-1}$	0.90***		0.96***
	(0.13)		(0.13)
$CAR_{i,t-1}$	-0.29***		-0.22***
	(0.07)		(0.08)
$RETURN\_SKEW_{i,t-1}$	-0.09*		-0.08
	(0.05)		(0.05)
$RETURN\_VOL_{i,t-1}$	3.01***		2.45***
	(0.46)		(0.47)
$SHARE\_TURN_{i,t-1}$	0.10***		0.09***
	(0.01)		(0.01)
$BIOTECH_{i,t}$		0.77***	0.74***
		(0.12)	(0.12)
$COMP\_HARDWARE_{i,t}$		0.43**	0.36
		(0.21)	(0.23)
$ELECTRONICS_{i,t}$		-0.22	-0.34**
		(0.14)	(0.14)
$RETAIL_{i,t}$		-0.20	-0.25
		(0.17)	(0.16)
$COMP\_SOFTWARE_{i,t}$		0.28**	0.24*
		(0.14)	(0.14)
INTERCEPT	-5.48***	-4.44***	-5.32***
	(0.17)	(0.12)	(0.17)
N	31,625	31,625	31,625
Pseudo R^2	0.040	0.033	0.058

Table A.4, Continued

In-sample:			
Pred. Posit.	949	949	949
Correct Pred. Posit.	82	95	105
Precision	0.086	0.100	0.111
Sensitivity	0.092	0.107	0.118
False Negative Rate	0.908	0.893	0.882
Specificity	0.972	0.972	0.973
AUC	0.680	0.645	0.713
Out-of-sample:			
Precision	0.086	0.097	0.104
Sensitivity	0.092	0.103	0.111
False Negative Rate	0.908	0.897	0.889
Specificity	0.972	0.972	0.972
AUC	0.677	0.640	0.704

TABLE A.5

Examining Future Market Outcomes over a Longer Period

Table A.5 presents results similar to Table 8 examining whether current quarter's plaintiff-lawyer views are associated with future market outcomes in the firm-quarter sample from 2008–2016 by estimating equation (B) in the manuscript. Panel A examines future next quarter's buy-and-hold abnormal returns using a Fama-French 4 factor model to calculate expected returns, and Panel B examines next quarter's daily return volatility. Columns 1, 3, and 5 examine total plaintiff-lawyer views ($LN\_VIEWS_{i,t}$); Columns 2, 4, and 6 examine disaggregated plaintiff-lawyer views ($TOP\_PLF\_LN\_VIEWS_{i,t}$ and $REM\_PLF\_LN\_VIEWS_{i,t}$). Columns 1–2 examine all firm-quarters. Columns 3–4 (5–6) exclude quarters in which litigation is filed (class period end occurs for subsequent litigation). For readability, we scale plaintiff-lawyer views by 100. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Refer to Table A.41 for variable definitions.

Panel A. Future Market Returns

	1 $BHAR_{i,t+1}$	2 $BHAR_{i,t+1}$	3 $BHAR_{i,t+1}$	4 $BHAR_{i,t+1}$	5 $BHAR_{i,t+1}$	6 $BHAR_{i,t+1}$
LN_VIEWS (/100)$_{i,t}$	-0.32***		-0.32***		-0.33***	
	(0.08)		(0.08)		(0.08)	
TOP_PLF_LN_VIEWS (/100)$_{i,t}$		-0.70***		-0.76***		-0.77***
		(0.20)		(0.20)		(0.20)
REM_PLF_LN_VIEWS (/100)$_{i,t}$		-0.16*		-0.17*		-0.17*
		(0.09)		(0.09)		(0.09)
LN_MVE$_{i,t}$	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LN_BOOK-TO-MARKET$_{i,t}$	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LN_TURNOVER$_{i,t}$	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ALPHA$_{i,t}$	-61.44***	-61.47***	-61.52***	-61.53***	-61.51***	-61.53***
	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)
INSTIT_OWN$_{i,t}$	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
NASDAQ$_{i,t}$	0.00***	0.00***	0.00***	0.00***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
INTERCEPT	-0.10***	-0.10***	-0.11***	-0.11***	-0.11***	-0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
FF48 FE	Yes	Yes	Yes	Yes	Yes	Yes
Fyear x Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
N	167,358	167,358	166,183	166,183	166,593	166,593
R^2	0.209	0.209	0.209	0.209	0.209	0.209

Panel B. Future Return Volatility

	1 RETURN_VOL$_{i,t+1}$	2 RETURN_VOL$_{i,t+1}$	3 RETURN_VOL$_{i,t+1}$	4 RETURN_VOL$_{i,t+1}$	5 RETURN_VOL$_{i,t+1}$	6 RETURN_VOL$_{i,t+1}$
LN_VIEWS (/100)$_{i,t}$	0.14***		0.13***		0.13***	
	(0.01)		(0.01)		(0.01)	
TOP_PLF_LN_VIEWS (/100)$_{i,t}$		0.16***		0.17***		0.17***
		(0.02)		(0.02)		(0.02)
REM_PLF_LN_VIEWS (/100)$_{i,t}$		0.11***		0.11***		0.11***
		(0.01)		(0.01)		(0.01)
LN_MVE$_{i,t}$	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LN_BOOK-TO-MARKET$_{i,t}$	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LN_TURNOVER$_{i,t}$	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ALPHA$_{i,t}$	-1.18***	-1.18***	-1.17***	-1.17***	-1.17***	-1.17***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)
INSTIT_OWN$_{i,t}$	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
NASDAQ$_{i,t}$	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
INTERCEPT	0.08***	0.08***	0.08***	0.08***	0.08***	0.08***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
FF48 FE	Yes	Yes	Yes	Yes	Yes	Yes
Fyear x Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
N	167,358	167,358	166,183	166,183	166,593	166,593
R^2	0.492	0.492	0.492	0.492	0.492	0.492

TABLE A.6

Persistence of Litigation Filings and Major Restatements

Table A.6 presents results examining the persistence of litigation filings (Panel A) and major restatements (Panel B) in consecutive years in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. Refer to Table A.41 for variable definitions.

Panel A. Persistence of Litigation Filings

	$\mathbf{SUED}_{i,t}$	
	0	1
$\mathbf{SUED}_{i,t-1}$		
0	16,181	537
1	444	17

Panel B. Persistence of Major Restatements

	$\mathbf{MAJOR\_RESTATE\_ANNCT}_{i,t}$	
	0	1
$\mathbf{MAJOR\_RESTATE\_ANNCT}_{i,t-1}$		
0	16,607	268
1	281	23

TABLE A.7

Persistence of Plaintiff-Lawyer Views

Table A.7 presents results on the persistence of plaintiff-lawyer scrutiny in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. We concentrate on observations with at least one view from plaintiffs' lawyers in the current year and then independently sort current plaintiff-lawyer views (LN_VIEWS$_{i,t}$) and prior plaintiff-lawyer year views (LN_VIEWS$_{i,t-1}$) into terciles. Larger concentration of observations on the diagonal implies higher persistence.

	Ln_VIEWS$_{i,t}$ Tercile		
	Low	Medium	High
Ln_VIEWS$_{i,t-1}$ Tercile			
Low	1,850	1,084	765
Medium	636	488	403
High	502	621	1,202

TABLE A.8

Predicting Realized Litigation Risk Controlling for Prior Year Litigation Filings

Table A.8 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equations (3) and (3) in the manuscript to include litigation during year t–1 ($\text{SUED}_{i,t-1}$). Columns 1 and 4 present results based on estimating equation (3) and $\text{SUED}_{i,t-1}$ using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) using plaintiff-lawyer EDGAR views and $\text{SUED}_{i,t-1}$. Columns 3 and 6 present results based on combining the two groups of variables other than $\text{FPS}_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Tabble A.41 for variable definitions.

	1 $\text{SUED}_{i,t}$	2 $\text{SUED}_{i,t}$	3 $\text{SUED}_{i,t}$	4 $\text{SUED\_INV}_{i,t}$	5 $\text{SUED\_INV}_{i,t}$	6 $\text{SUED\_INV}_{i,t}$
$\text{TOP\_PLF\_LN\_VIEWS}_{i,t-1}$		0.36***	0.31***		0.29***	0.26***
		(0.05)	(0.05)		(0.04)	(0.04)
$\text{REM\_PLF\_LN\_VIEWS}_{i,t-1}$		0.22***	0.20***		0.20***	0.18***
		(0.04)	(0.04)		(0.03)	(0.03)
$\text{FPS}_{i,t}$	0.19*			0.25***		
	(0.10)			(0.07)		
$\text{LN\_ASSETS}_{i,t-1}$	0.15***	0.02	0.09***	0.14***	0.03**	0.08***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
$\text{SALES\_GR}_{i,t-1}$	0.88***		0.93***	0.62***		0.66***
	(0.18)		(0.18)	(0.13)		(0.13)
$\text{CAR}_{i,t-1}$	-0.23**		-0.18*	0.01		0.05
	(0.10)		(0.10)	(0.07)		(0.07)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.15**		-0.12*	-0.09**		-0.08*
	(0.06)		(0.06)	(0.04)		(0.04)
$\text{RETURN\_VOL}_{i,t-1}$	4.87***		3.96***	4.01***		3.33***
	(0.72)		(0.74)	(0.55)		(0.56)
$\text{SHARE\_TURN}_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
$\text{SUED}_{i,t-1}$	-0.52*	-1.29***	-1.65***	0.65***	-0.08	-0.28
	(0.28)	(0.30)	(0.32)	(0.15)	(0.18)	(0.18)
$\text{BIOTECH}_{i,t}$		0.80***	0.72***		0.75***	0.65***
		(0.16)	(0.16)		(0.12)	(0.12)
$\text{COMP\_HARDWARE}_{i,t}$		0.55**	0.51*		0.61***	0.59***
		(0.27)	(0.27)		(0.19)	(0.19)
$\text{ELECTRONICS}_{i,t}$		-0.31	-0.38*		-0.13	-0.17
		(0.19)	(0.20)		(0.13)	(0.13)
$\text{RETAIL}_{i,t}$		-0.21	-0.23		-0.01	-0.04
		(0.21)	(0.21)		(0.16)	(0.16)
$\text{COMP\_SOFTWARE}_{i,t}$		0.20	0.12		0.09	0.04
		(0.19)	(0.19)		(0.14)	(0.14)
INTERCEPT	-5.39***	-3.89***	-5.08***	-4.33***	-3.16***	-4.04***
	(0.23)	(0.15)	(0.23)	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.039	0.036	0.064	0.032	0.035	0.050

Table A.8, Continued

In-sample:						
Pred. Posit.	516	516	516	516	516	516
Correct Pred. Posit.	48	64	72	89	114	117
Precision	0.093	0.124	0.140	0.172	0.221	0.227
Sensitivity	0.087	0.116	0.130	0.077	0.098	0.101
False Negative Rate	0.913	0.884	0.870	0.923	0.902	0.899
Specificity	0.972	0.973	0.973	0.973	0.975	0.975
AUC	0.673	0.637	0.708	0.645	0.637	0.680
Out-of-sample:						
Precision	0.085	0.115	0.131	0.162	0.219	0.206
Sensitivity	0.080	0.108	0.123	0.072	0.098	0.092
False Negative Rate	0.920	0.892	0.877	0.928	0.902	0.908
Specificity	0.971	0.972	0.973	0.973	0.975	0.974
AUC	0.666	0.629	0.694	0.641	0.633	0.669

TABLE A.9

Predicting Realized Litigation Risk excluding Serial Litigation

Table A.9 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after excluding from the sample firm-years that face a lawsuit both in year t and t–1. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than $\text{FPS}_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $\text{SUED}_{i,t}$	2 $\text{SUED}_{i,t}$	3 $\text{SUED}_{i,t}$	4 $\text{SUED\_INV}_{i,t}$	5 $\text{SUED\_INV}_{i,t}$	6 $\text{SUED\_INV}_{i,t}$
$\text{TOP\_PLF\_LN\_VIEWS}_{i,t-1}$		0.25***	0.18***		0.27***	0.22***
		(0.05)	(0.05)		(0.03)	(0.04)
$\text{REM\_PLF\_LN\_VIEWS}_{i,t-1}$		0.15***	0.11***		0.18***	0.16***
		(0.04)	(0.04)		(0.03)	(0.03)
$\text{FPS}_{i,t}$	0.18*			0.26***		
	(0.10)			(0.08)		
$\text{LN\_ASSETS}_{i,t-1}$	0.14***	0.04**	0.11***	0.14***	0.03**	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
$\text{SALES\_GR}_{i,t-1}$	0.93***		0.96***	0.65***		0.68***
	(0.18)		(0.18)	(0.13)		(0.13)
$\text{CAR}_{i,t-1}$	-0.19*		-0.13	-0.02		0.06
	(0.10)		(0.10)	(0.07)		(0.07)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.11*		-0.08	-0.10**		-0.06
	(0.06)		(0.06)	(0.04)		(0.04)
$\text{RETURN\_VOL}_{i,t-1}$	4.34***		3.39***	4.25***		3.17***
	(0.71)		(0.74)	(0.55)		(0.57)
$\text{SHARE\_TURN}_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
$\text{BIOTECH}_{i,t}$		0.79***	0.71***		0.74***	0.65***
		(0.15)	(0.15)		(0.12)	(0.12)
$\text{COMP\_HARDWARE}_{i,t}$		0.45	0.42		0.57***	0.55***
		(0.28)	(0.27)		(0.20)	(0.20)
$\text{ELECTRONICS}_{i,t}$		-0.28	-0.35*		-0.12	-0.16
		(0.19)	(0.19)		(0.13)	(0.13)
$\text{RETAIL}_{i,t}$		-0.14	-0.17		0.01	-0.01
		(0.21)	(0.21)		(0.16)	(0.16)
$\text{COMP\_SOFTWARE}_{i,t}$		0.26	0.18		0.11	0.06
		(0.18)	(0.18)		(0.14)	(0.14)
INTERCEPT	-5.28***	-3.98***	-5.07***	-4.36***	-3.17***	-4.03***
	(0.22)	(0.15)	(0.22)	(0.17)	(0.11)	(0.16)
N	17,162	17,162	17,162	17,162	17,162	17,162
Pseudo R^2	0.037	0.025	0.050	0.028	0.031	0.046
In-sample:						
Pred. Posit.	515	515	515	515	515	515
Correct Pred. Posit.	46	53	50	76	103	106

Table A.9, Continued

Precision	0.089	0.103	0.097	0.148	0.200	0.206
Sensitivity	0.086	0.099	0.093	0.066	0.090	0.092
False Negative Rate	0.914	0.901	0.907	0.934	0.910	0.908
Specificity	0.972	0.972	0.972	0.973	0.974	0.974
AUC	0.674	0.624	0.699	0.637	0.633	0.677
Out-of-sample:						
Precision	0.081	0.087	0.087	0.150	0.190	0.190
Sensitivity	0.078	0.084	0.084	0.068	0.086	0.086
False Negative Rate	0.922	0.916	0.916	0.932	0.914	0.914
Specificity	0.971	0.971	0.971	0.972	0.974	0.974
AUC	0.669	0.617	0.687	0.633	0.630	0.666

TABLE A.10

Predicting Realized Litigation Risk excluding Firm-Years Sued in Year t–1

Table A.10 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after excluding from the sample firm-years that faced a lawsuit in year t–1. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than $\text{FPS}_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $\text{SUED}_{i,t}$	2 $\text{SUED}_{i,t}$	3 $\text{SUED}_{i,t}$	4 $\text{SUED\_INV}_{i,t}$	5 $\text{SUED\_INV}_{i,t}$	6 $\text{SUED\_INV}_{i,t}$
$\text{TOP\_PLF\_LN\_VIEWS}_{i,t-1}$		0.39***	0.33***		0.33***	0.29***
		(0.05)	(0.05)		(0.04)	(0.04)
$\text{REM\_PLF\_LN\_VIEWS}_{i,t-1}$		0.23***	0.21***		0.22***	0.20***
		(0.04)	(0.04)		(0.03)	(0.03)
$\text{FPS}_{i,t}$	0.19*			0.27***		
	(0.10)			(0.08)		
$\text{LN\_ASSETS}_{i,t-1}$	0.15***	0.01	0.08***	0.14***	0.03**	0.08***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
$\text{SALES\_GR}_{i,t-1}$	0.95***		1.01***	0.61***		0.65***
	(0.18)		(0.18)	(0.14)		(0.14)
$\text{CAR}_{i,t-1}$	-0.25**		-0.19*	-0.01		0.05
	(0.11)		(0.11)	(0.08)		(0.08)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.15**		-0.12*	-0.09*		-0.07
	(0.07)		(0.06)	(0.05)		(0.05)
$\text{RETURN\_VOL}_{i,t-1}$	4.99***		4.03***	4.06***		3.24***
	(0.74)		(0.76)	(0.57)		(0.58)
$\text{SHARE\_TURN}_{i,t-1}$	0.09***		0.09***	0.07***		0.07***
	(0.01)		(0.01)	(0.01)		(0.01)
$\text{BIOTECH}_{i,t}$		0.79***	0.71***		0.77***	0.67***
		(0.16)	(0.17)		(0.13)	(0.13)
$\text{COMP\_HARDWARE}_{i,t}$		0.43	0.38		0.61***	0.57***
		(0.29)	(0.28)		(0.21)	(0.21)
$\text{ELECTRONICS}_{i,t}$		-0.29	-0.36*		-0.10	-0.16
		(0.20)	(0.20)		(0.13)	(0.14)
$\text{RETAIL}_{i,t}$		-0.18	-0.21		-0.06	-0.09
		(0.22)	(0.22)		(0.16)	(0.16)
$\text{COMP\_SOFTWARE}_{i,t}$		0.23	0.14		0.15	0.09
		(0.19)	(0.20)		(0.15)	(0.15)
INTERCEPT	-5.40***	-3.84***	-5.06***	-4.41***	-3.19***	-4.09***
	(0.23)	(0.15)	(0.23)	(0.17)	(0.11)	(0.17)
N	16,718	16,718	16,718	16,718	16,718	16,718
Pseudo R^2	0.043	0.038	0.068	0.030	0.033	0.051
In-sample:						
Pred. Posit.	502	502	502	502	502	502
Correct Pred. Posit.	47	64	73	78	97	117

Table A.10, Continued

Precision	0.094	0.127	0.145	0.155	0.193	0.233
Sensitivity	0.088	0.119	0.136	0.072	0.089	0.108
False Negative Rate	0.912	0.881	0.864	0.928	0.911	0.892
Specificity	0.972	0.973	0.973	0.973	0.974	0.975
AUC	0.682	0.639	0.715	0.639	0.630	0.678
Out-of-sample:						
Precision	0.094	0.133	0.137	0.151	0.200	0.204
Sensitivity	0.089	0.127	0.130	0.071	0.094	0.096
False Negative Rate	0.911	0.873	0.870	0.929	0.906	0.904
Specificity	0.971	0.973	0.973	0.972	0.974	0.974
AUC	0.676	0.633	0.703	0.634	0.623	0.665

TABLE A.11

Predicting Realized Litigation Risk excluding Firm-Years with Major Bad News Events in Year *t*–1

Table A.11 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after excluding firm-year with any major bad events in year *t*–1. We define major bad events as litigation filings; earnings warnings; or announcement of major restatements, internal control weaknesses, CEO turnover, CFO turnover, non-timely filings, or auditor changes. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than $FPS_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED_{i,t}$	4 $SUED\_INV_{i,t}$	5 $SUED\_INV_{i,t}$	6 $SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$		0.37***	0.33***		0.32***	0.29***
		(0.07)	(0.08)		(0.06)	(0.06)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$		0.27***	0.25***		0.22***	0.20***
		(0.06)	(0.05)		(0.04)	(0.04)
$FPS_{i,t}$	0.07			0.29***		
	(0.15)			(0.11)		
$LN\_ASSETS_{i,t-1}$	0.10***	-0.00	0.04	0.09***	0.01	0.04*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
$SALES\_GR_{i,t-1}$	0.91***		1.00***	0.63***		0.69***
	(0.25)		(0.25)	(0.19)		(0.19)
$CAR_{i,t-1}$	-0.02		0.02	0.17		0.20*
	(0.14)		(0.15)	(0.10)		(0.10)
$RETURN\_SKEW_{i,t-1}$	-0.13		-0.11	-0.05		-0.04
	(0.09)		(0.09)	(0.06)		(0.06)
$RETURN\_VOL_{i,t-1}$	3.93***		2.95***	2.87***		2.09***
	(0.98)		(1.01)	(0.75)		(0.76)
$SHARE\_TURN_{i,t-1}$	0.10***		0.10***	0.08***		0.08***
	(0.02)		(0.02)	(0.01)		(0.01)
$BIOTECH_{i,t}$		0.71***	0.58**		0.88***	0.75***
		(0.22)	(0.23)		(0.17)	(0.17)
$COMP\_HARDWARE_{i,t}$		0.35	0.23		0.73**	0.64**
		(0.46)	(0.44)		(0.30)	(0.30)
$ELECTRONICS_{i,t}$		-0.31	-0.46		-0.18	-0.32
		(0.28)	(0.29)		(0.20)	(0.21)
$RETAIL_{i,t}$		-0.67	-0.73		-0.13	-0.18
		(0.48)	(0.48)		(0.25)	(0.25)
$COMP\_SOFTWARE_{i,t}$		-0.06	-0.20		0.28	0.17
		(0.30)	(0.30)		(0.20)	(0.19)
INTERCEPT	-5.04***	-3.77***	-4.71***	-4.07***	-3.15***	-3.81***
	(0.30)	(0.21)	(0.30)	(0.21)	(0.15)	(0.21)
N	9,705	9,705	9,705	9,705	9,705	9,705
Pseudo R^2	0.038	0.032	0.063	0.028	0.027	0.047

Table A.11, Continued

In-sample:						
Pred. Posit.	292	292	292	292	292	292
Correct Pred. Posit.	26	34	39	40	52	61
Precision	0.089	0.116	0.134	0.137	0.178	0.209
Sensitivity	0.093	0.121	0.139	0.071	0.092	0.108
False Negative Rate	0.907	0.879	0.861	0.929	0.908	0.892
Specificity	0.972	0.973	0.973	0.972	0.974	0.975
AUC	0.676	0.616	0.704	0.641	0.614	0.670
Out-of-sample:						
Precision	0.083	0.100	0.120	0.133	0.163	0.177
Sensitivity	0.089	0.107	0.128	0.071	0.087	0.094
False Negative Rate	0.911	0.893	0.872	0.929	0.913	0.906
Specificity	0.971	0.971	0.972	0.972	0.973	0.973
AUC	0.665	0.596	0.683	0.633	0.609	0.654

TABLE A.12

Predicting Realized Litigation Risk using Plaintiff-Lawyer Views that are Orthogonal to Major Bad News Events

Table A.12 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equation (3) in the manuscript to use views that are orthogonal to major bad news events. Specifically, we regress TOP_PLF_LN_VIEWS$_{i,t-1}$ and REM_PLF_LN_VIEWS$_{i,t-1}$ on indicator variables set to one for litigation filings; earnings warnings; and announcement of major restatements, internal control weaknesses, CEO turnover, CFO turnover, non-timely filings, or auditor changes; and a count of large daily negative market-adjusted returns (i.e., $< -10\%$). All variables are measured with respect to year t–1 to be consistent with plaintiff-lawyer views. We then obtain the residuals from these regressions (TOP_PLF_LN_VIEWS-ORTHOGONAL$_{i,t-1}$ and REM_PLF_LN_VIEWS-ORTHOGONAL$_{i,t-1}$) and use them in our prediction model. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating the modified equation (3) based on plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than FPS$_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, views are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 SUED$_{i,t}$	2 SUED$_{i,t}$	3 SUED$_{i,t}$	4 SUED_INV$_{i,t}$	5 SUED_INV$_{i,t}$	6 SUED_INV$_{i,t}$
TOP_PLF_LN_VIEWS-ORTHOGONAL$_{i,t-1}$		0.36***	0.30***		0.30***	0.26***
		(0.05)	(0.05)		(0.04)	(0.04)
REM_PLF_LN_VIEWS-ORTHOGONAL$_{i,t-1}$		0.22***	0.19***		0.19***	0.17***
		(0.04)	(0.04)		(0.03)	(0.03)
FPS$_{i,t}$	0.19*			0.26***		
	(0.10)			(0.08)		
LN_ASSETS$_{i,t-1}$	0.15***	0.03	0.09***	0.14***	0.04**	0.09***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
SALES_GR$_{i,t-1}$	0.88***		0.93***	0.62***		0.66***
	(0.18)		(0.18)	(0.13)		(0.13)
CAR$_{i,t-1}$	-0.21**		-0.15	-0.03		0.01
	(0.10)		(0.10)	(0.07)		(0.07)
RETURN_SKEW$_{i,t-1}$	-0.13**		-0.11*	-0.11***		-0.10**
	(0.06)		(0.06)	(0.04)		(0.04)

Table A.12, Continued

RETURN_VOL$_{i,t-1}$	4.63***		3.59***	4.40***		3.56***
	(0.71)		(0.74)	(0.55)		(0.57)
SHARE_TURN$_{i,t-1}$	0.09***		0.08***	0.06***		0.06***
	(0.01)		(0.01)	(0.01)		(0.01)
BIOTECH$_{i,t}$		0.80***	0.71***		0.78***	0.68***
		(0.16)	(0.16)		(0.13)	(0.13)
COMP_HARDWARE$_{i,t}$		0.57**	0.53**		0.66***	0.62***
		(0.27)	(0.27)		(0.19)	(0.19)
ELECTRONICS$_{i,t}$		-0.30	-0.37*		-0.13	-0.17
		(0.19)	(0.19)		(0.13)	(0.13)
RETAIL$_{i,t}$		-0.18	-0.22		0.01	-0.01
		(0.21)	(0.21)		(0.16)	(0.16)
COMP_SOFTWARE$_{i,t}$		0.21	0.13		0.12	0.06
		(0.19)	(0.19)		(0.15)	(0.14)
INTERCEPT	-5.35***	-3.74***	-4.87***	-4.39***	-3.00***	-3.96***
	(0.23)	(0.16)	(0.23)	(0.17)	(0.12)	(0.17)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.038	0.036	0.062	0.029	0.029	0.047
In-sample:						
Pred. Posit.	516	516	516	516	516	516
Correct Pred. Posit.	49	60	68	79	102	112
Precision	0.095	0.116	0.132	0.153	0.198	0.217
Sensitivity	0.088	0.108	0.123	0.068	0.088	0.096
False Negative Rate	0.912	0.892	0.877	0.932	0.912	0.904
Specificity	0.972	0.973	0.973	0.973	0.974	0.975
AUC	0.674	0.636	0.708	0.638	0.622	0.672
Out-of-sample:						
Precision	0.085	0.117	0.123	0.146	0.198	0.200
Sensitivity	0.080	0.110	0.116	0.065	0.089	0.089
False Negative Rate	0.920	0.890	0.884	0.935	0.911	0.911
Specificity	0.971	0.972	0.973	0.972	0.974	0.974
AUC	0.669	0.628	0.696	0.635	0.617	0.660

TABLE A.13

Determinants of Plaintiff-Lawyer Views excluding Variables Associated with Bad News

Table A.13 presents results similar in nature to Appendix C estimating equation (A) in the manuscript using OLS after also excluding bad news related variables defined as variables under *Accounting Events*, *Personnel Events*, and *Disclosure* categories. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1
	$\text{LN\_VIEWS}_{i,t}$
Earnings Characteristics	
$\text{POSITIVE\_NON-GAAP\_ADJ}_{i,t}$	0.17***
	(0.03)
$\text{POSITIVE\_DISC\_ACCR}_{i,t-1}$	-0.06***
	(0.02)
Visibility	
$\text{LN\_ASSETS}_{i,t-1}$	0.20***
	(0.01)
$\text{LN\_AGE}_{i,t-1}$	-0.03
	(0.02)
Complexity	
$\text{MULTI-SEGMENTS}_{i,t-1}$	-0.07**
	(0.03)
$\text{MULTINATIONAL}_{i,t-1}$	0.15***
	(0.03)
$\text{LOSS}_{i,t-1}$	0.22***
	(0.03)
$\text{NO\_DIVIDEND\_PAID}_{i,t-1}$	0.17***
	(0.03)
External Monitors	
$\text{BIG4}_{i,t-1}$	-0.16***
	(0.03)
$\text{INSTIT\_OWN}_{i,t-1}$	0.20***
	(0.04)
High Risk Industries	
$\text{BIOTECH}_{i,t}$	0.59***
	(0.07)
$\text{COMP\_HARDWARE}_{i,t}$	0.26**
	(0.11)
$\text{ELECTRONICS}_{i,t}$	-0.03
	(0.05)
$\text{RETAIL}_{i,t}$	0.29***
	(0.06)
$\text{COMP\_SOFTWARE}_{i,t}$	0.15**
	(0.06)
INTERCEPT	-0.73***
	(0.07)
Year FE	No
N	17,179
R^2	0.122

TABLE A.14

Predicting Realized Litigation Risk using Predicted Plaintiff-Lawyer Views excluding Variables Associated with Bad News

Table A.14 presents results similar in nature to Table 10 examining *predicted* plaintiff-lawyer views as a proxy for ex ante litigation risk after replacing predicted plaintiff-lawyer views obtained from using coefficients from Appendix C to using coefficients from Table A.13. Columns 1–2 and 5–6 use realized litigation filings ($SUED_{i,t}$), whereas Columns 3–4 use realized litigation filings supplemented with investigation announcements ($SUED\_INV_{i,t}$). We use a later (an earlier) time period in Columns 1–4 (5–6) to assess the predictive ability of our model in periods where it was not estimated. Columns 1–4 use fiscal years 2017 up to those ending on 12/31/2019; Columns 5–6 use fiscal years 2007 to 2011. We do not use investigation announcements for the earlier period as those announcements were not common prior to 2012. Columns 1, 3, and 5 use predicted ex ante litigation risk from Kim and Skinner (2012). Columns 2, 4, and 6 use predicted plaintiff-lawyer scrutiny obtained using Table A.13. All models are estimated using logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED\_INV_{i,t}$	4 $SUED\_INV_{i,t}$	5 $SUED_{i,t}$	6 $SUED_{i,t}$
$PRED\_WITHOUT\_BAD\_NEWS\_LN\_VIEWS_{i,t-1}$		1.40***		1.42***		0.88***
		(0.22)		(0.17)		(0.22)
$EX\_ANTE\_LIT\_RISK_{i,t}$	0.18***		0.18***		0.11***	
	(0.02)		(0.02)		(0.02)	
$FPS_{i,t}$	0.29**		0.36***			
	(0.11)		(0.09)			
$LN\_ASSETS_{i,t-1}$	0.13***	-0.07*	0.15***	-0.06*	0.23***	0.11***
	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.04)
$BIOTECH_{i,t}$		0.15		0.24	1.06***	0.64***
		(0.23)		(0.18)	(0.17)	(0.21)
$COMP\_HARDWARE_{i,t}$		-0.07		-0.07	0.25	-0.08
		(0.40)		(0.32)	(0.39)	(0.40)
$ELECTRONICS_{i,t}$		0.10		-0.19	0.12	0.11
		(0.22)		(0.21)	(0.19)	(0.19)
$RETAIL_{i,t}$		-0.36		-0.31	0.04	-0.14
		(0.25)		(0.20)	(0.23)	(0.23)
$COMP\_SOFTWARE_{i,t}$		-0.27		-0.06	0.46**	0.24
		(0.23)		(0.19)	(0.23)	(0.23)

Table A.14, Continued

INTERCEPT	-3.85***	-3.73***	-3.35***	-3.23***	-5.47***	-5.35***
	(0.18)	(0.20)	(0.14)	(0.16)	(0.19)	(0.21)
N	9,133	9,133	9,133	9,133	17,065	17,065
Pseudo R^2	0.033	0.031	0.038	0.040	0.041	0.039
In-sample:						
Pred. Posit.	274	274	274	274	512	512
Correct Pred. Posit.	35	44	60	79	45	49
Precision	0.128	0.161	0.219	0.288	0.088	0.096
Sensitivity	0.076	0.095	0.073	0.096	0.116	0.127
False Negative Rate	0.924	0.905	0.927	0.904	0.884	0.873
Specificity	0.972	0.973	0.974	0.977	0.972	0.972
AUC	0.649	0.634	0.647	0.645	0.667	0.659
Out-of-sample:						
Precision	0.114	0.150	0.221	0.279	0.085	0.092
Sensitivity	0.069	0.091	0.075	0.094	0.113	0.124
False Negative Rate	0.931	0.909	0.925	0.906	0.887	0.876
Specificity	0.971	0.973	0.974	0.976	0.971	0.972
AUC	0.645	0.627	0.645	0.642	0.656	0.650

TABLE A.15

Can Plaintiff-Lawyer Views Predict Case Outcomes using Arctan Transformation?

Table A.15 presents results similar in nature to Tables 2 and 3 after partitioning $\text{LN\_VIEWS}_{i,[Class\ End,\ Filing-1]}$ into $\text{TOP\_PLF\_LN\_VIEWS}_{i,[Class\ End,\ Filing-1]}$ and $\text{REM\_PLF\_LN\_VIEWS}_{i,[Class\ End,\ Filing-1]}$ and adjusting equation (F) in the manuscript to use views transformed using arctan (inverse of the tangent), instead of a natural log transformation. Panel A (B) of the table examines whether total (disaggregated) plaintiff-lawyer EDGAR views starting on the class period end date and ending the day prior to the filing date can predict outcomes associated with case merits (i.e., the case settles in Columns 1–2 and has larger settlement amounts in Column 3) using the full population of securities class actions with available data from 2012–2016. That is, the starting population for this test is the 932 cases shown in Table A.2 Panel A before requiring data for control variables, excluding ongoing cases given the examination of case outcomes. Columns 1–2 (3–4) use a logistic (OLS) regression. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

Panel A. All Views

	1 $\text{SETTLED}_{i,t}$	2 $\text{SETTLED}_{i,t}$	3 $\text{LN\_SETTLEM}_{i,t}$
$\text{ARCTAN\_VIEWS}_{i,[Class\ End,\ Filing-1]}$	0.33***	0.08***	0.27***
	(0.12)	(0.03)	(0.10)
$\text{LN\_DAMAGES}_{i,t}$	0.06*	0.01**	0.13***
	(0.03)	(0.01)	(0.04)
$\text{LN\_MVE}_{i,t}$	-0.07	-0.01	0.42***
	(0.05)	(0.01)	(0.06)
$\text{ROA}_{i,t}$	0.01	-0.00	-0.51***
	(0.29)	(0.07)	(0.19)
$\text{TOBINSQ}_{i,t}$	0.00	0.00	-0.01
	(0.03)	(0.01)	(0.03)
$\text{INSTIT\_OWN}_{i,t}$	-0.28	-0.06	0.43*
	(0.24)	(0.06)	(0.22)
INTERCEPT	-0.66	0.37	10.00***
	(1.30)	(0.30)	(0.63)
Filing Year FE	Yes	Yes	Yes
Circuit FE	Yes	Yes	Yes
N	768	768	317
AUC	0.656		
Pseudo R^2	0.052		
R^2		0.068	0.513

Panel B. Top versus Remaining Views

	1 SETTLED$_{i,t}$	2 SETTLED$_{i,t}$	3 LN_SETTLEM$_{i,t}$
TOP_PLF_ARCTAN_VIEWS$_{i,[Class\ End,\ Filing-1]}$	0.41***	0.09***	0.38***
	(0.13)	(0.03)	(0.10)
REM_PLF_ARCTAN_VIEWS$_{i,[Class\ End,\ Filing-1]}$	0.01	0.00	0.05
	(0.12)	(0.03)	(0.10)
LN_DAMAGES$_{i,t}$	0.05*	0.01**	0.12***
	(0.03)	(0.01)	(0.04)
LN_MVE$_{i,t}$	-0.07	-0.01	0.41***
	(0.05)	(0.01)	(0.06)
ROA$_{i,t}$	-0.02	-0.01	-0.52***
	(0.29)	(0.07)	(0.19)
TOBINSQ$_{i,t}$	-0.00	-0.00	-0.00
	(0.03)	(0.01)	(0.02)
INSTIT_OWN$_{i,t}$	-0.24	-0.05	0.43**
	(0.24)	(0.06)	(0.22)
INTERCEPT	-0.25	0.46	10.61***
	(1.43)	(0.35)	(0.69)
Filing Year FE	Yes	Yes	Yes
Circuit FE	Yes	Yes	Yes
N	768	768	317
AUC	0.662		
Pseudo R^2	0.055		
R^2		0.071	0.527

TABLE A.16
Predicting Realized Litigation Risk using Arctan Transformation

Table A.16 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equation (3) in the manuscript to use views transformed using arctan (inverse of the tangent) instead of a natural log transformation. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating the modified equation (3) using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than $FPS_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED_{i,t}$	4 $SUED\_INV_{i,t}$	5 $SUED\_INV_{i,t}$	6 $SUED\_INV_{i,t}$
$TOP\_PLF\_ARCTAN\_VIEWS_{i,t-1}$		0.59***	0.43***		0.58***	0.48***
		(0.09)	(0.09)		(0.06)	(0.07)
$REM\_PLF\_ARCTAN\_VIEWS_{i,t-1}$		0.38***	0.32***		0.40***	0.36***
		(0.08)	(0.08)		(0.06)	(0.06)
$FPS_{i,t}$	0.19*			0.26***		
	(0.10)			(0.08)		
$LN\_ASSETS_{i,t-1}$	0.15***	0.03	0.10***	0.14***	0.03*	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$SALES\_GR_{i,t-1}$	0.88***		0.91***	0.62***		0.66***
	(0.18)		(0.18)	(0.13)		(0.13)
$CAR_{i,t-1}$	-0.21**		-0.14	-0.03		0.04
	(0.10)		(0.10)	(0.07)		(0.07)
$RETURN\_SKEW_{i,t-1}$	-0.13**		-0.08	-0.11***		-0.07
	(0.06)		(0.06)	(0.04)		(0.04)
$RETURN\_VOL_{i,t-1}$	4.63***		3.47***	4.40***		3.27***
	(0.71)		(0.75)	(0.55)		(0.56)
$SHARE\_TURN_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
$BIOTECH_{i,t}$		0.79***	0.71***		0.75***	0.65***
		(0.15)	(0.16)		(0.12)	(0.12)
$COMP\_HARDWARE_{i,t}$		0.54**	0.52**		0.61***	0.59***
		(0.26)	(0.26)		(0.18)	(0.18)
$ELECTRONICS_{i,t}$		-0.31*	-0.38*		-0.14	-0.18
		(0.19)	(0.19)		(0.13)	(0.13)
$RETAIL_{i,t}$		-0.21	-0.23		-0.04	-0.05
		(0.21)	(0.21)		(0.16)	(0.16)
$COMP\_SOFTWARE_{i,t}$		0.23	0.15		0.10	0.06
		(0.18)	(0.18)		(0.14)	(0.14)
INTERCEPT	-5.35***	-3.99***	-5.09***	-4.39***	-3.19***	-4.08***
	(0.23)	(0.15)	(0.23)	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.038	0.032	0.056	0.029	0.034	0.050
In-sample:						
Pred. Posit.	516	516	516	516	516	516
Correct Pred. Posit.	49	54	62	79	100	116
Precision	0.095	0.105	0.120	0.153	0.194	0.225
Sensitivity	0.088	0.097	0.112	0.068	0.086	0.100
False Negative Rate	0.912	0.903	0.888	0.932	0.914	0.900

Table A.16, Continued

Specificity	0.972	0.972	0.973	0.973	0.974	0.975
AUC	0.674	0.634	0.705	0.638	0.636	0.678
Out-of-sample:						
Precision	0.085	0.096	0.092	0.146	0.185	0.190
Sensitivity	0.080	0.090	0.087	0.065	0.083	0.085
False Negative Rate	0.920	0.910	0.913	0.935	0.917	0.915
Specificity	0.971	0.972	0.972	0.972	0.974	0.974
AUC	0.669	0.628	0.693	0.635	0.633	0.668

TABLE A.17
Can Plaintiff-Lawyer Views Predict Case Outcomes using Arcsinh Transformation?

Table A.17 presents results similar in nature to Tables 2 and 3 after partitioning LN_VIEWS$_{i,[Class\ End,\ Filing-1]}$ into TOP_PLF_LN_VIEWS$_{i,[Class\ End,\ Filing-1]}$ and REM_PLF_LN_VIEWS$_{i,[Class\ End,\ Filing-1]}$ and adjusting equation (F) in the manuscript to use views transformed using arcsinh (inverse hyperbolic sine), instead of a natural log transformation. Panel A (B) of the table examines whether total (disaggregated) plaintiff-lawyer EDGAR views starting on the class period end date and ending the day prior to the filing date can predict outcomes associated with case merits (i.e., the case settles in Columns 1–2 and has larger settlement amounts in Column 3) using the full population of securities class actions with available data from 2012–2016. That is, the starting population for this test is the 932 cases shown in Table A.2 Panel A before requiring data for control variables, excluding ongoing cases given the examination of case outcomes. Columns 1–2 (3–4) use a logistic (OLS) regression. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

Panel A. All Views

	1 SETTLED$_{i,t}$	2 SETTLED$_{i,t}$	3 LN_SETTLEM$_{i,t}$
ARCSINH_VIEWS$_{i,[Class\ End,\ Filing-1]}$	0.11***	0.03***	0.14***
	(0.04)	(0.01)	(0.04)
LN_DAMAGES$_{i,t}$	0.06*	0.01**	0.13***
	(0.03)	(0.01)	(0.03)
LN_MVE$_{i,t}$	-0.07	-0.02	0.41***
	(0.05)	(0.01)	(0.06)
ROA$_{i,t}$	0.00	-0.00	-0.52***
	(0.30)	(0.07)	(0.19)
TOBINSQ$_{i,t}$	0.00	0.00	-0.00
	(0.03)	(0.01)	(0.02)
INSTIT_OWN$_{i,t}$	-0.27	-0.06	0.39*
	(0.24)	(0.06)	(0.22)
INTERCEPT	-0.70	0.36	9.88***
	(1.29)	(0.30)	(0.59)
Filing Year FE	Yes	Yes	Yes
Circuit FE	Yes	Yes	Yes
N	768	768	317
AUC	0.656		
Pseudo R^2	0.051		
R^2		0.067	0.526

Panel B. Top versus Remaining Views

	1 $\text{SETTLED}_{i,t}$	2 $\text{SETTLED}_{i,t}$	3 $\text{LN\_SETTLEM}_{i,t}$
$\text{TOP\_PLF\_ARCSINH\_VIEWS}_{i,[Class\ End,\ Filing-1]}$	0.10**	0.02**	0.15***
	(0.05)	(0.01)	(0.04)
$\text{REM\_PLF\_ARCSINH\_VIEWS}_{i,[Class\ End,\ Filing-1]}$	0.06	0.01	0.06
	(0.05)	(0.01)	(0.04)
$\text{LN\_DAMAGES}_{i,t}$	0.06*	0.01**	0.13***
	(0.03)	(0.01)	(0.03)
$\text{LN\_MVE}_{i,t}$	-0.07	-0.02	0.41***
	(0.05)	(0.01)	(0.06)
$\text{ROA}_{i,t}$	-0.01	-0.01	-0.52***
	(0.30)	(0.07)	(0.19)
$\text{TOBINSQ}_{i,t}$	0.00	0.00	-0.00
	(0.03)	(0.01)	(0.02)
$\text{INSTIT\_OWN}_{i,t}$	-0.24	-0.06	0.42*
	(0.24)	(0.06)	(0.22)
INTERCEPT	-0.51	0.41	10.39***
	(1.37)	(0.33)	(0.65)
Filing Year FE	Yes	Yes	Yes
Circuit FE	Yes	Yes	Yes
N	768	768	317
AUC	0.657		
Pseudo R^2	0.050		
R^2		0.066	0.531

TABLE A.18
Predicting Realized Litigation Risk using Arcsinh Transformation

Table A.18 presents results similar in nature to Table 6 Panel A in the manuscript examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equation (3) to use views transformed using arcsinh (inverse hyperbolic sine) instead of a natural log transformation. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating the modified equation (3) using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than $FPS_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED_{i,t}$	4 $SUED\_INV_{i,t}$	5 $SUED\_INV_{i,t}$	6 $SUED\_INV_{i,t}$
$TOP\_PLF\_ARCSINH\_VIEWS_{i,t-1}$		0.23***	0.17***		0.24***	0.20***
		(0.04)	(0.04)		(0.03)	(0.03)
$REM\_PLF\_ARCSINH\_VIEWS_{i,t-1}$		0.15***	0.12***		0.17***	0.14***
		(0.03)	(0.03)		(0.02)	(0.02)
$FPS_{i,t}$	0.19*			0.26***		
	(0.10)			(0.08)		
$LN\_ASSETS_{i,t-1}$	0.15***	0.04**	0.11***	0.14***	0.03**	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
$SALES\_GR_{i,t-1}$	0.88***		0.91***	0.62***		0.66***
	(0.18)		(0.18)	(0.13)		(0.13)
$CAR_{i,t-1}$	-0.21**		-0.13	-0.03		0.06
	(0.10)		(0.10)	(0.07)		(0.07)
$RETURN\_SKEW_{i,t-1}$	-0.13**		-0.09	-0.11***		-0.07
	(0.06)		(0.06)	(0.04)		(0.04)
$RETURN\_VOL_{i,t-1}$	4.63***		3.47***	4.40***		3.21***
	(0.71)		(0.74)	(0.55)		(0.56)
$SHARE\_TURN_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
$BIOTECH_{i,t}$		0.80***	0.72***		0.74***	0.65***
		(0.15)	(0.16)		(0.12)	(0.12)
$COMP\_HARDWARE_{i,t}$		0.54**	0.52**		0.61***	0.59***
		(0.25)	(0.25)		(0.18)	(0.18)
$ELECTRONICS_{i,t}$		-0.31	-0.37*		-0.13	-0.17
		(0.19)	(0.19)		(0.13)	(0.13)
$RETAIL_{i,t}$		-0.19	-0.21		-0.02	-0.04
		(0.21)	(0.21)		(0.16)	(0.16)
$COMP\_SOFTWARE_{i,t}$		0.22	0.15		0.09	0.05
		(0.18)	(0.18)		(0.14)	(0.14)
INTERCEPT	-5.35***	-3.98***	-5.07***	-4.39***	-3.16***	-4.04***
	(0.23)	(0.15)	(0.22)	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.038	0.032	0.055	0.029	0.035	0.050
In-sample:						
Pred. Posit.	516	516	516	516	516	516
Correct Pred. Posit.	49	58	61	79	110	119
Precision	0.095	0.112	0.118	0.153	0.213	0.231
Sensitivity	0.088	0.105	0.110	0.068	0.095	0.102
False Negative Rate	0.912	0.895	0.890	0.932	0.905	0.898

Table A.18, Continued

Specificity	0.972	0.972	0.973	0.973	0.975	0.975
AUC	0.674	0.635	0.707	0.638	0.638	0.682
Out-of-sample:						
Precision	0.085	0.115	0.104	0.146	0.213	0.210
Sensitivity	0.080	0.108	0.098	0.065	0.095	0.094
False Negative Rate	0.920	0.892	0.902	0.935	0.905	0.906
Specificity	0.971	0.972	0.972	0.972	0.974	0.974
AUC	0.669	0.626	0.696	0.635	0.633	0.670

TABLE A.19

Predicting Realized Litigation Risk using Total Plaintiff-Lawyer Views

Table A.19 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equation (3) in the manuscript to use all views combined ($\text{LN\_VIEWS}_{i,t-1}$), rather than splitting them into $\text{TOP\_PLF\_LN\_VIEWS}_{i,t-1}$ and $\text{REM\_PLF\_LN\_VIEWS}_{i,t-1}$. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) modified to use combined plaintiff-lawyer EDGAR views ($\text{LN\_VIEWS}_{i,t-1}$). Columns 3 and 6 present results based on combining the two groups of variables other than $\text{FPS}_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $\text{SUED}_{i,t}$	2 $\text{SUED}_{i,t}$	3 $\text{SUED}_{i,t}$	4 $\text{SUED\_INV}_{i,t}$	5 $\text{SUED\_INV}_{i,t}$	6 $\text{SUED\_INV}_{i,t}$
$\text{LN\_VIEWS}_{i,t-1}$		0.29***	0.23***		0.31***	0.26***
		(0.03)	(0.03)		(0.02)	(0.02)
$\text{FPS}_{i,t}$	0.19*			0.26***		
	(0.10)			(0.08)		
$\text{LN\_ASSETS}_{i,t-1}$	0.15***	0.03*	0.10***	0.14***	0.03*	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
$\text{SALES\_GR}_{i,t-1}$	0.88***		0.91***	0.62***		0.66***
	(0.18)		(0.18)	(0.13)		(0.13)
$\text{CAR}_{i,t-1}$	-0.21**		-0.15	-0.03		0.04
	(0.10)		(0.10)	(0.07)		(0.07)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.13**		-0.09	-0.11***		-0.07*
	(0.06)		(0.06)	(0.04)		(0.04)
$\text{RETURN\_VOL}_{i,t-1}$	4.63***		3.57***	4.40***		3.33***
	(0.71)		(0.73)	(0.55)		(0.56)
$\text{SHARE\_TURN}_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
$\text{BIOTECH}_{i,t}$		0.75***	0.68***		0.70***	0.62***
		(0.15)	(0.15)		(0.12)	(0.12)
$\text{COMP\_HARDWARE}_{i,t}$		0.56**	0.52**		0.62***	0.59***
		(0.26)	(0.26)		(0.19)	(0.18)
$\text{ELECTRONICS}_{i,t}$		-0.32*	-0.39**		-0.15	-0.18
		(0.19)	(0.19)		(0.13)	(0.13)
$\text{RETAIL}_{i,t}$		-0.17	-0.20		-0.01	-0.03
		(0.21)	(0.21)		(0.16)	(0.16)
$\text{COMP\_SOFTWARE}_{i,t}$		0.23	0.16		0.10	0.06
		(0.18)	(0.18)		(0.14)	(0.14)
INTERCEPT	-5.35***	-3.99***	-5.10***	-4.39***	-3.17***	-4.07***
	(0.23)	(0.15)	(0.22)	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.038	0.031	0.056	0.029	0.035	0.050
In-sample:						
Pred. Posit.	516	516	516	516	516	516

Table A.19, Continued

Correct Pred. Posit.	49	53	62	79	106	109
Precision	0.095	0.103	0.120	0.153	0.205	0.211
Sensitivity	0.088	0.096	0.112	0.068	0.091	0.094
False Negative Rate	0.912	0.904	0.888	0.932	0.909	0.906
Specificity	0.972	0.972	0.973	0.973	0.974	0.975
AUC	0.674	0.635	0.704	0.638	0.637	0.679
Out-of-sample:						
Precision	0.085	0.092	0.100	0.146	0.196	0.202
Sensitivity	0.080	0.087	0.094	0.065	0.088	0.090
False Negative Rate	0.920	0.913	0.906	0.935	0.912	0.910
Specificity	0.971	0.972	0.972	0.972	0.974	0.974
AUC	0.669	0.628	0.693	0.635	0.633	0.669

TABLE A.20

Predicting Realized Litigation Risk using a Single High-Risk Industry Indicator Variable in All Models

Table A.20 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equation (3) in the manuscript to use $\text{FPS}_{i,t}$ instead of the industry indicator variables per Brochet and Srinivasan (2014). Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating the modified equation (3) using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $\text{SUED}_{i,t}$	2 $\text{SUED}_{i,t}$	3 $\text{SUED}_{i,t}$	4 $\text{SUED\_INV}_{i,t}$	5 $\text{SUED\_INV}_{i,t}$	6 $\text{SUED\_INV}_{i,t}$
$\text{TOP\_PLF\_LN\_VIEWS}_{i,t-1}$		0.27***	0.21***		0.28***	0.23***
		(0.05)	(0.05)		(0.03)	(0.04)
$\text{REM\_PLF\_LN\_VIEWS}_{i,t-1}$		0.19***	0.15***		0.21***	0.18***
		(0.04)	(0.04)		(0.03)	(0.03)
$\text{FPS}_{i,t}$	0.19*	0.18*	0.13	0.26***	0.24***	0.20***
	(0.10)	(0.09)	(0.10)	(0.08)	(0.07)	(0.07)
$\text{LN\_ASSETS}_{i,t-1}$	0.15***	0.03	0.10***	0.14***	0.03*	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
$\text{SALES\_GR}_{i,t-1}$	0.88***		0.90***	0.62***		0.65***
	(0.18)		(0.18)	(0.13)		(0.13)
$\text{CAR}_{i,t-1}$	-0.21**		-0.12	-0.03		0.07
	(0.10)		(0.10)	(0.07)		(0.07)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.13**		-0.09	-0.11***		-0.07*
	(0.06)		(0.06)	(0.04)		(0.04)
$\text{RETURN\_VOL}_{i,t-1}$	4.63***		3.77***	4.40***		3.48***
	(0.71)		(0.72)	(0.55)		(0.56)
$\text{SHARE\_TURN}_{i,t-1}$	0.09***		0.07***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
INTERCEPT	-5.35***	-3.92***	-5.05***	-4.39***	-3.14***	-4.05***
	(0.23)	(0.15)	(0.22)	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.038	0.025	0.050	0.029	0.031	0.046
In-sample:						
Pred. Posit.	516	516	516	516	516	516
Correct Pred. Posit.	49	50	52	79	101	110
Precision	0.095	0.097	0.101	0.153	0.196	0.213
Sensitivity	0.088	0.090	0.094	0.068	0.087	0.095
False Negative Rate	0.912	0.910	0.906	0.932	0.913	0.905
Specificity	0.972	0.972	0.972	0.973	0.974	0.975
AUC	0.674	0.621	0.701	0.638	0.630	0.674
Out-of-sample:						
Precision	0.085	0.100	0.100	0.146	0.192	0.208

Table A.20, Continued

Sensitivity	0.080	0.094	0.094	0.065	0.086	0.093
False Negative Rate	0.920	0.906	0.906	0.935	0.914	0.907
Specificity	0.971	0.972	0.972	0.972	0.974	0.974
AUC	0.669	0.618	0.696	0.635	0.629	0.670

TABLE A.21

Predicting Realized Litigation Risk using Multiple Industry Indicator Variables in All Models

Table A.21 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after adjusting equation (3) in the manuscript to use individual industry indicator variables per Brochet and Srinivasan (2014). Columns 1 and 4 present results based on estimating the modified equation (3) using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2	3	4	5	6
	$\text{SUED}_{i,t}$	$\text{SUED}_{i,t}$	$\text{SUED}_{i,t}$	$\text{SUED\_INV}_{i,t}$	$\text{SUED\_INV}_{i,t}$	$\text{SUED\_INV}_{i,t}$
$\text{TOP\_PLF\_LN\_VIEWS}_{i,t-1}$		0.28***	0.21***		0.28***	0.24***
		(0.05)	(0.05)		(0.03)	(0.04)
$\text{REM\_PLF\_LN\_VIEWS}_{i,t-1}$		0.17***	0.13***		0.19***	0.17***
		(0.04)	(0.04)		(0.03)	(0.03)
$\text{BIOTECH}_{i,t}$	0.84***	0.81***	0.73***	0.78***	0.75***	0.66***
	(0.16)	(0.15)	(0.16)	(0.13)	(0.12)	(0.12)
$\text{COMP\_HARDWARE}_{i,t}$	0.62**	0.55**	0.53**	0.69***	0.62***	0.59***
	(0.27)	(0.25)	(0.25)	(0.19)	(0.18)	(0.18)
$\text{ELECTRONICS}_{i,t}$	-0.39**	-0.31	-0.37*	-0.19	-0.13	-0.17
	(0.20)	(0.19)	(0.19)	(0.14)	(0.13)	(0.13)
$\text{RETAIL}_{i,t}$	-0.11	-0.18	-0.21	0.07	-0.01	-0.03
	(0.22)	(0.21)	(0.21)	(0.16)	(0.16)	(0.16)
$\text{COMP\_SOFTWARE}_{i,t}$	0.23	0.22	0.15	0.14	0.09	0.05
	(0.19)	(0.18)	(0.18)	(0.15)	(0.14)	(0.14)
$\text{LN\_ASSETS}_{i,t-1}$	0.16***	0.04**	0.11***	0.15***	0.03**	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
$\text{SALES\_GR}_{i,t-1}$	0.89***		0.91***	0.63***		0.66***
	(0.18)		(0.18)	(0.13)		(0.13)
$\text{CAR}_{i,t-1}$	-0.22**		-0.13	-0.04		0.06
	(0.10)		(0.10)	(0.07)		(0.07)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.13**		-0.09	-0.11**		-0.07
	(0.06)		(0.06)	(0.04)		(0.04)
$\text{RETURN\_VOL}_{i,t-1}$	4.35***		3.50***	4.15***		3.23***
	(0.72)		(0.74)	(0.56)		(0.56)
$\text{SHARE\_TURN}_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
INTERCEPT	-5.39***	-3.98***	-5.08***	-4.39***	-3.17***	-4.04***
	(0.23)	(0.15)	(0.22)	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.045	0.031	0.055	0.034	0.035	0.050
In-sample:						
Pred. Posit.	516	516	516	516	516	516

Table A.21, Continued

Correct Pred. Posit.	54	59	59	86	111	117
Precision	0.105	0.114	0.114	0.167	0.215	0.227
Sensitivity	0.097	0.106	0.106	0.074	0.095	0.101
False Negative Rate	0.903	0.894	0.894	0.926	0.905	0.899
Specificity	0.972	0.973	0.973	0.973	0.975	0.975
AUC	0.684	0.635	0.707	0.648	0.638	0.681
Out-of-sample:						
Precision	0.092	0.112	0.100	0.154	0.217	0.208
Sensitivity	0.087	0.105	0.094	0.069	0.097	0.093
False Negative Rate	0.913	0.895	0.906	0.931	0.903	0.907
Specificity	0.972	0.972	0.972	0.973	0.975	0.974
AUC	0.675	0.626	0.696	0.642	0.634	0.670

TABLE A.22

Predicting Realized Litigation Risk Relative to using Litigation Industry Indicator Variable(s)

Table A.22 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. Columns 1 and 4 present results based on only estimating a model containing $FPS_{i,t}$ and $LN\_ASSETS_{i,t-1}$. Columns 2 and 5 present results based on only estimating a model containing individual high litigation industries indicator variables as per equation (3) from Brochet and Srinivasan (2014) and $LN\_ASSETS_{i,t-1}$. Columns 3 and 6 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED_{i,t}$	4 $SUED\_INV_{i,t}$	5 $SUED\_INV_{i,t}$	6 $SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$			0.28***			0.28***
			(0.05)			(0.03)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$			0.17***			0.19***
			(0.04)			(0.03)
$FPS_{i,t}$	0.27***			0.33***		
	(0.10)			(0.08)		
$BIOTECH_{i,t}$		0.98***	0.81***		0.92***	0.75***
		(0.16)	(0.15)		(0.13)	(0.12)
$COMP\_HARDWARE_{i,t}$		0.69**	0.55**		0.74***	0.62***
		(0.27)	(0.25)		(0.19)	(0.18)
$ELECTRONICS_{i,t}$		-0.28	-0.31		-0.11	-0.13
		(0.20)	(0.19)		(0.14)	(0.13)
$RETAIL_{i,t}$		-0.07	-0.18		0.10	-0.01
		(0.22)	(0.21)		(0.16)	(0.16)
$COMP\_SOFTWARE_{i,t}$		0.34*	0.22		0.21	0.09
		(0.19)	(0.18)		(0.15)	(0.14)
$LN\_ASSETS_{i,t-1}$	0.09***	0.10***	0.04**	0.09***	0.09***	0.03**
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
INTERCEPT	-4.08***	-4.16***	-3.98***	-3.30***	-3.34***	-3.17***
	(0.16)	(0.16)	(0.15)	(0.12)	(0.12)	(0.11)
N	17,179	17,179	17,179	17,179	17,179	17,179
Pseudo R^2	0.005	0.012	0.031	0.006	0.011	0.035
In-sample:						
Pred. Posit.	516	516	516	516	516	516
Correct Pred. Posit.	36	46	59	53	82	111
Precision	0.070	0.089	0.114	0.103	0.159	0.215
Sensitivity	0.065	0.083	0.106	0.046	0.071	0.095
False Negative Rate	0.935	0.917	0.894	0.954	0.929	0.905
Specificity	0.971	0.972	0.973	0.971	0.973	0.975
AUC	0.550	0.574	0.635	0.559	0.573	0.638
Out-of-sample:						
Precision	0.059	0.087	0.112	0.100	0.156	0.217

Table A.22, Continued

Sensitivity	0.056	0.081	0.105	0.045	0.070	0.097
False Negative Rate	0.944	0.919	0.895	0.955	0.930	0.903
Specificity	0.970	0.971	0.972	0.971	0.973	0.975
AUC	0.549	0.563	0.626	0.558	0.569	0.634

TABLE A.23
Predicting Realized Litigation Risk Relative to Brochet and Srinivasan (2014)

Table A.23 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–2) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 3–4) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. Columns 1 and 3 present results based on estimating the Brochet and Srinivasan (2014) model, modifying it to use lagged values. Columns 2 and 4 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer views. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED\_INV_{i,t}$	4 $SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$		0.28***		0.28***
		(0.05)		(0.03)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$		0.17***		0.19***
		(0.04)		(0.03)
$LN\_MVE_{v1\ i,t-1}$	0.34***		0.28***	
	(0.03)		(0.02)	
$BHR_{i,t-1}$	0.15*		0.19***	
	(0.09)		(0.06)	
$RETURN\_VOL_{v1\ i,t-1}$	35.82***		30.26***	
	(4.09)		(3.07)	
$RETURN\_SKEW_{v1\ i,t-1}$	-0.17***		-0.14***	
	(0.04)		(0.03)	
$SHARE\_TURN_{v1\ i,t-1}$	23.30***		16.38***	
	(4.01)		(3.25)	
$BETA_{i,t-1}$	0.11		0.09*	
	(0.07)		(0.06)	
$BIOTECH_{i,t}$	0.61***	0.81***	0.61***	0.75***
	(0.17)	(0.15)	(0.13)	(0.12)
$COMP\_HARDWARE_{i,t}$	0.51*	0.55**	0.62***	0.62***
	(0.26)	(0.25)	(0.19)	(0.18)
$ELECTRONICS_{i,t}$	-0.46**	-0.31	-0.23*	-0.13
	(0.20)	(0.19)	(0.14)	(0.13)
$RETAIL_{i,t}$	-0.13	-0.18	0.05	-0.01
	(0.21)	(0.21)	(0.16)	(0.16)
$COMP\_SOFTWARE_{i,t}$	0.16	0.22	0.08	0.09
	(0.18)	(0.18)	(0.14)	(0.14)
$LN\_ASSETS_{i,t-1}$		0.04**		0.03**
		(0.02)		(0.01)
INTERCEPT	-9.54***	-3.98***	-7.65***	-3.17***
	(0.50)	(0.15)	(0.36)	(0.11)
N	17,179	17,179	17,179	17,179
Pseudo R^2	0.069	0.031	0.052	0.035
In-sample:				
Pred. Posit.	516	516	516	516
Correct Pred. Posit.	60	59	106	111
Precision	0.116	0.114	0.205	0.215

Table A.23, Continued

Sensitivity	0.108	0.106	0.091	0.095
False Negative Rate	0.892	0.894	0.909	0.905
Specificity	0.973	0.973	0.974	0.975
AUC	0.720	0.635	0.677	0.638
Out-of-sample:				
Precision	0.108	0.112	0.198	0.217
Sensitivity	0.101	0.105	0.089	0.097
False Negative Rate	0.899	0.895	0.911	0.903
Specificity	0.972	0.972	0.974	0.975
AUC	0.715	0.626	0.674	0.634

TABLE A.24

Predicting Realized Litigation Risk Relative to using LIBERAL_COURT$_{i,t-1}$

Table A.24 presents results examining proxies for ex ante litigation risk using realized litigation filings (Panel A) and realized litigation filings supplemented with announcements of investigations by law firm (Panel B) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. The sample size decreases relative to Table 6 Panel A due to the additional data requirements from Huang et al. (2019). Column 1 presents results using only LIBERAL_COURT$_{i,t-1}$, FPS$_{i,t}$, and LN_ASSETS$_{i,t-1}$. Column 2 presents results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Column 3 presents results combining these two models and estimating the equation. Column 4 presents results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Column 5 augments this model with LIBERAL_COURT$_{i,t-1}$ and estimates the equation. Column 6 combines the variables in Columns 2 and 4 other than FPS$_{i,t}$ and estimates the equation. Column 7 augments this model with LIBERAL_COURT$_{i,t-1}$ and estimates the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Kim and Skinner (2012) and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are indeed true positives. Sensitivity is calculated as the percent of true positive cases correctly identified by the model. Specificity is calculated as the percent of true negative cases correctly identified by the model. False Negative Rate is calculated as 1 minus Sensitivity. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

Panel A. Sued

	1 SUED$_{i,t}$	2 SUED$_{i,t}$	3 SUED$_{i,t}$	4 SUED$_{i,t}$	5 SUED$_{i,t}$	6 SUED$_{i,t}$	7 SUED$_{i,t}$
TOP_PLF_LN_VIEWS$_{i,t-1}$				0.30***	0.29***	0.23***	0.22***
				(0.05)	(0.05)	(0.05)	(0.05)
REM_PLF_LN_VIEWS$_{i,t-1}$				0.16***	0.16***	0.11***	0.10**
				(0.04)	(0.04)	(0.04)	(0.04)
LIBERAL_COURT$_{i,t-1}$	0.80***		0.64***		0.57***		0.49**
	(0.23)		(0.23)		(0.22)		(0.22)
FPS$_{i,t}$	0.22**	0.18	0.15				
	(0.11)	(0.11)	(0.11)				
LN_ASSETS$_{i,t-1}$	0.09***	0.16***	0.16***	0.03	0.04*	0.11***	0.11***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
SALES_GR$_{i,t-1}$		0.80***	0.80***			0.84***	0.84***
		(0.20)	(0.20)			(0.20)	(0.20)
CAR$_{i,t-1}$		-0.20*	-0.21*			-0.11	-0.11
		(0.12)	(0.12)			(0.12)	(0.12)
RETURN_SKEW$_{i,t-1}$		-0.17**	-0.17**			-0.12*	-0.12*
		(0.07)	(0.07)			(0.07)	(0.07)
RETURN_VOL$_{i,t-1}$		5.17***	5.18***			3.88***	3.90***
		(0.80)	(0.79)			(0.83)	(0.83)
SHARE_TURN$_{i,t-1}$		0.10***	0.10***			0.09***	0.09***
		(0.02)	(0.02)			(0.02)	(0.02)
BIOTECH$_{i,t}$				0.77***	0.75***	0.64***	0.62***
				(0.18)	(0.17)	(0.18)	(0.18)
COMP_HARDWARE$_{i,t}$				0.69***	0.62**	0.64**	0.58**
				(0.25)	(0.25)	(0.25)	(0.25)
ELECTRONICS$_{i,t}$				-0.40*	-0.45**	-0.43*	-0.47**
				(0.23)	(0.23)	(0.23)	(0.23)
RETAIL$_{i,t}$				-0.20	-0.18	-0.22	-0.20
				(0.22)	(0.22)	(0.22)	(0.22)
COMP_SOFTWARE$_{i,t}$				0.18	0.13	0.15	0.11
				(0.19)	(0.19)	(0.20)	(0.20)

Table A.24, Continued

INTERCEPT	-4.46***	-5.48***	-5.80***	-3.93***	-4.21***	-5.13***	-5.38***
	(0.21)	(0.25)	(0.28)	(0.16)	(0.19)	(0.25)	(0.28)
N	14,193	14,193	14,193	14,193	14,193	14,193	14,193
Pseudo R^2	0.008	0.044	0.046	0.033	0.034	0.059	0.061
In-sample:							
Pred. Posit.	426	426	426	426	426	426	426
Correct Pred. Posit.	27	37	42	46	46	46	47
Precision	0.063	0.087	0.099	0.108	0.108	0.108	0.110
Sensitivity	0.059	0.081	0.092	0.101	0.101	0.101	0.103
False Negative Rate	0.941	0.919	0.908	0.899	0.899	0.899	0.897
Specificity	0.971	0.972	0.972	0.972	0.972	0.972	0.972
AUC	0.573	0.687	0.690	0.644	0.650	0.719	0.719
Out-of-sample:							
Precision	0.063	0.091	0.093	0.107	0.107	0.109	0.107
Sensitivity	0.059	0.085	0.088	0.101	0.101	0.103	0.100
False Negative Rate	0.941	0.915	0.912	0.899	0.899	0.897	0.900
Specificity	0.971	0.972	0.972	0.972	0.972	0.972	0.972
AUC	0.569	0.682	0.684	0.636	0.640	0.707	0.707

Panel B. Sued or Investigated

	1	2	3	4	5	6	7
	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$				0.29***	0.28***	0.24***	0.23***
				(0.04)	(0.04)	(0.04)	(0.04)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$				0.17***	0.17***	0.14***	0.13***
				(0.03)	(0.03)	(0.03)	(0.03)
$LIBERAL\_COURT_{i,t-1}$	0.71***		0.60***		0.52***		0.46***
	(0.17)		(0.17)		(0.17)		(0.17)
$FPS_{i,t}$	0.29***	0.26***	0.23***				
	(0.08)	(0.08)	(0.08)				
$LN\_ASSETS_{i,t-1}$	0.11***	0.18***	0.19***	0.05***	0.06***	0.12***	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$SALES\_GR_{i,t-1}$		0.48***	0.48***			0.53***	0.53***
		(0.15)	(0.15)			(0.15)	(0.15)
$CAR_{i,t-1}$		-0.06	-0.06			0.04	0.04
		(0.08)	(0.08)			(0.08)	(0.08)
$RETURN\_SKEW_{i,t-1}$		-0.12***	-0.12***			-0.08*	-0.08*
		(0.05)	(0.05)			(0.04)	(0.04)
$RETURN\_VOL_{i,t-1}$		5.37***	5.37***			4.08***	4.10***
		(0.62)	(0.61)			(0.63)	(0.63)
$SHARE\_TURN_{i,t-1}$		0.07***	0.06***			0.05***	0.05***
		(0.01)	(0.01)			(0.01)	(0.01)
$BIOTECH_{i,t}$				0.78***	0.76***	0.64***	0.63***
				(0.14)	(0.14)	(0.14)	(0.14)
$COMP\_HARDWARE_{i,t}$				0.69***	0.62***	0.64***	0.59***
				(0.19)	(0.19)	(0.19)	(0.19)
$ELECTRONICS_{i,t}$				-0.08	-0.13	-0.09	-0.13
				(0.14)	(0.14)	(0.14)	(0.14)
$RETAIL_{i,t}$				-0.05	-0.03	-0.06	-0.04
				(0.16)	(0.17)	(0.17)	(0.17)
$COMP\_SOFTWARE_{i,t}$				-0.03	-0.07	-0.03	-0.07
				(0.15)	(0.15)	(0.16)	(0.16)
INTERCEPT	-3.73***	-4.69***	-4.98***	-3.21***	-3.47***	-4.28***	-4.51***
	(0.15)	(0.19)	(0.20)	(0.12)	(0.15)	(0.19)	(0.20)
N	14,193	14,193	14,193	14,193	14,193	14,193	14,193
Pseudo R^2	0.011	0.036	0.038	0.035	0.037	0.053	0.054
In-sample:							
Pred. Posit.	426	426	426	426	426	426	426
Correct Pred. Posit.	47	72	68	91	89	89	93
Precision	0.110	0.169	0.160	0.214	0.209	0.209	0.218

Table A.24, Continued

Sensitivity	0.046	0.071	0.067	0.090	0.088	0.088	0.092
False Negative Rate	0.954	0.929	0.933	0.910	0.912	0.912	0.908
Specificity	0.971	0.973	0.973	0.975	0.974	0.974	0.975
AUC	0.583	0.652	0.656	0.644	0.649	0.688	0.690
Out-of-sample:							
Precision	0.107	0.172	0.174	0.212	0.219	0.216	0.221
Sensitivity	0.045	0.073	0.074	0.090	0.093	0.092	0.094
False Negative Rate	0.955	0.927	0.926	0.910	0.907	0.908	0.906
Specificity	0.971	0.973	0.973	0.974	0.975	0.974	0.975
AUC	0.581	0.648	0.651	0.638	0.644	0.677	0.679

TABLE A.25

Predicting Realized Litigation Risk without Requiring Accruals Variables

Table A.25 presents results similar to those presented in Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–3) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 4–6) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B but with no additional data requirements other than the Kim and Skinner (2012) variables included in the model. That is, it does not require data to calculate accruals (POSITIVE_DISC_ACCR$_{i,t-1}$) examined in Table 9. Columns 1 and 4 present results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Columns 2 and 5 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Columns 3 and 6 present results based on combining the two groups of variables other than FPS$_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2	3	4	5	6
	SUED$_{i,t}$	SUED$_{i,t}$	SUED$_{i,t}$	SUED_INV$_{i,t}$	SUED_INV$_{i,t}$	SUED_INV$_{i,t}$
TOP_PLF_LN_VIEWS$_{i,t-1}$		0.30***	0.22***		0.30***	0.24***
		(0.04)	(0.05)		(0.03)	(0.03)
REM_PLF_LN_VIEWS$_{i,t-1}$		0.18***	0.13***		0.21***	0.18***
		(0.04)	(0.04)		(0.02)	(0.02)
FPS$_{i,t}$	0.30***			0.35***		
	(0.10)			(0.07)		
LN_ASSETS$_{i,t-1}$	0.15***	0.03	0.10***	0.13***	0.02	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
SALES_GR$_{i,t-1}$	0.99***		1.02***	0.68***		0.71***
	(0.17)		(0.18)	(0.13)		(0.13)
CAR$_{i,t-1}$	-0.21**		-0.14	-0.06		0.04
	(0.10)		(0.10)	(0.07)		(0.07)
RETURN_SKEW$_{i,t-1}$	-0.16***		-0.11*	-0.12***		-0.07*
	(0.06)		(0.06)	(0.04)		(0.04)
RETURN_VOL$_{i,t-1}$	5.13***		3.95***	4.77***		3.55***
	(0.67)		(0.70)	(0.52)		(0.52)
SHARE_TURN$_{i,t-1}$	0.09***		0.08***	0.06***		0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
BIOTECH$_{i,t}$		0.94***	0.80***		0.84***	0.71***
		(0.15)	(0.16)		(0.12)	(0.12)
COMP_HARDWARE$_{i,t}$		0.68***	0.62**		0.69***	0.64***
		(0.25)	(0.25)		(0.18)	(0.18)
ELECTRONICS$_{i,t}$		-0.16	-0.27		-0.03	-0.10
		(0.19)	(0.19)		(0.13)	(0.13)
RETAIL$_{i,t}$		0.07	-0.00		0.16	0.12
		(0.18)	(0.18)		(0.14)	(0.14)
COMP_SOFTWARE$_{i,t}$		0.37**	0.26		0.17	0.10
		(0.18)	(0.18)		(0.14)	(0.14)
INTERCEPT	-5.51***	-4.07***	-5.23***	-4.47***	-3.18***	-4.10***
	(0.22)	(0.14)	(0.21)	(0.16)	(0.10)	(0.15)
N	21,394	21,394	21,394	21,394	21,394	21,394
Pseudo R^2	0.044	0.034	0.060	0.032	0.037	0.053

Table A.25, Continued

In-sample:						
Pred. Posit.	642	642	642	642	642	642
Correct Pred. Posit.	57	70	73	98	138	138
Precision	0.089	0.109	0.114	0.153	0.215	0.215
Sensitivity	0.093	0.114	0.119	0.073	0.103	0.103
Specificity	0.972	0.972	0.973	0.973	0.975	0.975
AUC	0.691	0.637	0.721	0.648	0.637	0.685
Out-of-sample:						
Precision	0.082	0.103	0.100	0.149	0.212	0.206
Sensitivity	0.087	0.109	0.106	0.072	0.103	0.100
Specificity	0.971	0.972	0.972	0.972	0.974	0.974
AUC	0.687	0.631	0.712	0.646	0.636	0.677

TABLE A.26

Predicting Realized Litigation Risk with Minimal Data Requirements

Table A.26 presents results based on estimating equation (3) in the manuscript using logistic estimation, similar to those presented in Columns 2 and 4 of Table 6 Panel A, in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B but with no additional data requirements other than the variables included in the model. That is, it only requires firm size and industry (i.e., it does not require accruals or the remaining Kim and Skinner (2012) variables). Columns 1 (2) present results using realized litigation filings (realized litigation filings, supplemented with plaintiff-lawyer investigation announcements). Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$	0.29***	0.30***
	(0.05)	(0.03)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$	0.21***	0.25***
	(0.03)	(0.02)
$BIOTECH_{i,t}$	1.18***	1.08***
	(0.11)	(0.09)
$COMP\_HARDWARE_{i,t}$	0.82***	0.85***
	(0.25)	(0.20)
$ELECTRONICS_{i,t}$	0.34**	0.34***
	(0.16)	(0.13)
$RETAIL_{i,t}$	0.54***	0.47***
	(0.15)	(0.12)
$COMP\_SOFTWARE_{i,t}$	0.39**	0.27*
	(0.18)	(0.15)
$LN\_ASSETS_{i,t-1}$	0.12***	0.11***
	(0.01)	(0.01)
INTERCEPT	-5.03***	-4.22***
	(0.11)	(0.08)
N	31,860	31,860
Pseudo R^2	0.054	0.062
In-sample:		
Pred. Posit.	956	956
Correct Pred. Posit.	89	176
Precision	0.093	0.184
Sensitivity	0.128	0.121
Specificity	0.972	0.974
AUC	0.694	0.698
Out-of-sample:		
Precision	0.090	0.184
Sensitivity	0.124	0.122
Specificity	0.972	0.974

Table A.26, Continued

AUC	0.690	0.697

TABLE A.27

Predicting Realized Litigation Risk using Lagged Variables from the Table 9 Determinants Model

Table A.27 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–2) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 3–4) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B but uses lagged variables from the determinants model (i.e., equation (A) in the manuscript) instead of the variables from Kim and Skinner (2012). Columns 1 and 3 present results based on estimating the lagged model from equation (A) in the manuscript to predict realized litigation risk. Columns 2 and 4 present results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2	3	4
	$SUED_{i,t}$	$SUED_{i,t}$	$SUED\_INV_{i,t}$	$SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t-1}$		0.26***		0.27***
		(0.05)		(0.04)
$REM\_PLF\_LN\_VIEWS_{i,t-1}$		0.19***		0.21***
		(0.04)		(0.03)
$AUDITOR\_CHANGE\_ANNCT_{i,t-1}$	-0.09		0.05	
	(0.23)		(0.15)	
$MAJOR\_RESTATE\_ANNCT_{i,t-1}$	-0.44		-0.21	
	(0.36)		(0.23)	
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t-1}$	0.32		0.36**	
	(0.23)		(0.16)	
$ICW\_ANNCT_{i,t-1}$	0.45**		0.48***	
	(0.21)		(0.15)	
$CEO\_TURNOVER_{i,t-1}$	0.07		-0.01	
	(0.15)		(0.11)	
$CFO\_TURNOVER_{i,t-1}$	0.01		0.09	
	(0.14)		(0.09)	
$LN\_VOLUNT\_8\text{-}Ks_{i,t-1}$	0.14**		0.18***	
	(0.06)		(0.05)	
$EARN\_WARN\_ANNCT_{i,t-1}$	0.13		0.01	
	(0.12)		(0.09)	
$POSITIVE\_NON\text{-}GAAP\_ADJ_{i,t-1}$	0.31***		0.20**	
	(0.12)		(0.08)	
$POSITIVE\_DISC\_ACCR_{i,t-2}$	0.10		0.08	
	(0.09)		(0.07)	
$LN\_ASSETS_{i,t-2}$	0.22***		0.19***	
	(0.03)		(0.02)	
$LN\_AGE_{i,t-2}$	-0.27***		-0.16***	
	(0.07)		(0.05)	
$MULTI\text{-}SEGMENTS_{i,t-2}$	-0.26***		-0.18**	
	(0.10)		(0.07)	
$MULTINATIONAL_{i,t-2}$	-0.07		0.02	
	(0.11)		(0.08)	
$LOSS_{i,t-2}$	0.32***		0.36***	
	(0.11)		(0.08)	

Table A.27, Continued

NO_DIVIDEND_PAID$_{i,t-2}$	0.59***		0.53***	
	(0.11)		(0.08)	
BIG4$_{i,t-2}$	-0.07		-0.13	
	(0.14)		(0.11)	
INSTIT_OWN$_{i,t-2}$	0.02		0.42***	
	(0.16)		(0.13)	
BIOTECH$_{i,t-1}$	1.04***	0.96***	0.90***	0.81***
	(0.17)	(0.15)	(0.13)	(0.13)
COMP_HARDWARE$_{i,t-1}$	0.68**	0.65**	0.65***	0.65***
	(0.30)	(0.26)	(0.20)	(0.19)
ELECTRONICS$_{i,t-1}$	-0.26	-0.24	-0.12	-0.10
	(0.20)	(0.19)	(0.14)	(0.13)
RETAIL$_{i,t-1}$	0.09	-0.04	0.14	0.04
	(0.22)	(0.21)	(0.17)	(0.16)
COMP_SOFTWARE$_{i,t-1}$	0.24	0.29	-0.01	0.01
	(0.21)	(0.20)	(0.17)	(0.16)
LN_ASSETS$_{i,t-1}$		0.05**		0.04**
		(0.02)		(0.02)
INTERCEPT	-5.15***	-4.15***	-4.64***	-3.27***
	(0.29)	(0.16)	(0.22)	(0.12)
N	16,158	16,158	16,158	16,158
Pseudo R^2	0.043	0.037	0.042	0.038
In-sample:				
Pred. Posit.	485	485	485	485
Correct Pred. Posit.	42	53	84	107
Precision	0.087	0.109	0.173	0.221
Sensitivity	0.085	0.107	0.080	0.102
False Negative Rate	0.915	0.893	0.920	0.898
Specificity	0.972	0.972	0.973	0.975
AUC	0.676	0.653	0.661	0.649
Out-of-sample:				
Precision	0.063	0.112	0.155	0.208
Sensitivity	0.063	0.111	0.072	0.097
False Negative Rate	0.937	0.889	0.928	0.903
Specificity	0.971	0.972	0.973	0.974
AUC	0.656	0.647	0.649	0.645

TABLE A.28

Predicting Realized Litigation Risk using Contemporaneous Variables from the Table 9 Determinants Model

Table A.28 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using realized litigation filings (Columns 1–2) and realized litigation filings, supplemented with plaintiff-lawyer investigation announcements (Columns 3–4) in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B but uses variables from the determinants model (i.e., equation (A) in the manuscript) instead of the variables from Kim and Skinner (2012). Columns 1 and 3 present results based on estimating the model from equation (A) in the manuscript to predict realized litigation risk. Columns 2 and 4 present results based on estimating equation (3) in the manuscript using contemporaneous plaintiff-lawyer EDGAR views. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED\_INV_{i,t}$	4 $SUED\_INV_{i,t}$
$TOP\_PLF\_LN\_VIEWS_{i,t}$		0.77***		0.65***
		(0.04)		(0.03)
$REM\_PLF\_LN\_VIEWS_{i,t}$		0.91***		0.76***
		(0.03)		(0.02)
$AUDITOR\_CHANGE\_ANNCT_{i,t}$	-0.17		-0.14	
	(0.20)		(0.15)	
$MAJOR\_RESTATE\_ANNCT_{i,t}$	1.51***		0.98***	
	(0.19)		(0.16)	
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t}$	0.95***		0.79***	
	(0.15)		(0.12)	
$ICW\_ANNCT_{i,t}$	0.23		0.48***	
	(0.17)		(0.12)	
$CEO\_TURNOVER_{i,t}$	0.68***		0.43***	
	(0.12)		(0.09)	
$CFO\_TURNOVER_{i,t}$	0.34***		0.17**	
	(0.12)		(0.09)	
$LN\_VOLUNT\_8\text{-}Ks_{i,t}$	0.08		0.18***	
	(0.06)		(0.05)	
$EARN\_WARN\_ANNCT_{i,t}$	0.22*		0.02	
	(0.11)		(0.09)	
$POSITIVE\_NON\text{-}GAAP\_ADJ_{i,t}$	0.17		0.07	
	(0.11)		(0.08)	
$POSITIVE\_DISC\_ACCR_{i,t-1}$	-0.09		0.02	
	(0.10)		(0.07)	
$LN\_ASSETS_{i,t-1}$	0.24***	-0.22***	0.21***	-0.15***
	(0.03)	(0.03)	(0.02)	(0.02)
$LN\_AGE_{i,t-1}$	-0.37***		-0.24***	
	(0.06)		(0.05)	
$MULTI\text{-}SEGMENTS_{i,t-1}$	-0.33***		-0.22***	
	(0.10)		(0.07)	
$MULTINATIONAL_{i,t-1}$	-0.00		0.07	
	(0.11)		(0.08)	

Table A.28, Continued

$\text{LOSS}_{i,t-1}$	0.25**		0.35***	
	(0.11)		(0.08)	
$\text{NO\_DIVIDEND\_PAID}_{i,t-1}$	0.51***		0.51***	
	(0.10)		(0.08)	
$\text{BIG4}_{i,t-1}$	0.03		-0.09	
	(0.14)		(0.10)	
$\text{INSTIT\_OWN}_{i,t-1}$	0.08		0.57***	
	(0.16)		(0.12)	
$\text{BIOTECH}_{i,t}$	0.88***	0.09	0.84***	0.23
	(0.18)	(0.17)	(0.14)	(0.15)
$\text{COMP\_HARDWARE}_{i,t}$	0.54*	0.17	0.64***	0.34
	(0.30)	(0.34)	(0.21)	(0.26)
$\text{ELECTRONICS}_{i,t}$	-0.27	-0.26	-0.09	-0.13
	(0.19)	(0.23)	(0.13)	(0.14)
$\text{RETAIL}_{i,t}$	0.02	-0.55**	0.14	-0.26
	(0.22)	(0.23)	(0.17)	(0.19)
$\text{COMP\_SOFTWARE}_{i,t}$	0.22	-0.20	0.13	-0.20
	(0.19)	(0.27)	(0.16)	(0.18)
INTERCEPT	-5.09***	-3.94***	-4.74***	-2.93***
	(0.29)	(0.16)	(0.22)	(0.11)
N	17,179	17,179	17,179	17,179
Pseudo R^2	0.094	0.412	0.074	0.279
In-sample:				
Pred. Posit.	516	516	516	516
Correct Pred. Posit.	99	268	149	341
Precision	0.192	0.519	0.289	0.661
Sensitivity	0.179	0.484	0.128	0.293
False Negative Rate	0.821	0.516	0.872	0.707
Specificity	0.975	0.985	0.977	0.989
AUC	0.728	0.924	0.699	0.834
Out-of-sample:				
Precision	0.185	0.512	0.287	0.650
Sensitivity	0.173	0.480	0.128	0.291
False Negative Rate	0.827	0.520	0.872	0.709
Specificity	0.974	0.985	0.977	0.989
AUC	0.715	0.923	0.692	0.833

TABLE A.29

Using Predicted Plaintiff-Lawyer Views versus Predicted Litigation Filings to Predict Realized Litigation Risk

Table A.29 presents results similar in nature to Table 10 examining *predicted* plaintiff-lawyer views as a proxy for ex ante litigation risk after replacing predicted ex ante litigation risk from Kim and Skinner (2012) with predicted lagged litigation probability as estimated in Column 1 of Table A.28 using logistic or OLS estimation. Thus, rather than the benchmark for PRED_LN_VIEWS$_{i,t-1}$ being EX_ANTE_LIT_RISK$_{i,t}$, the benchmarks are PRED_LITIG_PROB-DETERM_MODEL$_{Logit\ i,t-1}$ and PRED_LITIG_PROB-DETERM_MODEL$_{OLS\ i,t-1}$. Panel A presents results using fiscal years 2007 to 2011 to predict realized litigation filings. Panel B presents results using fiscal years 2017 up to those ending on 12/31/2019 to predict realized litigation filings, supplemented with plaintiff-lawyer investigation announcements. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

Panel A. 2007-2011

	1 SUED$_{i,t}$	2 SUED$_{i,t}$	3 SUED$_{i,t}$
PRED_LN_VIEWS$_{i,t-1}$			0.91***
			(0.14)
PRED_LITIG_PROB-DETERM_MODEL$_{Logit\ i,t-1}$		4.60***	
		(0.73)	
PRED_LITIG_PROB-DETERM_MODEL$_{OLS\ i,t-1}$	5.31***		
	(0.82)		
BIOTECH$_{i,t}$	0.94***	0.94***	0.65***
	(0.18)	(0.18)	(0.19)
COMP_HARDWARE$_{i,t}$	0.11	0.09	-0.09
	(0.39)	(0.40)	(0.39)
ELECTRONICS$_{i,t}$	0.16	0.16	0.08
	(0.19)	(0.19)	(0.19)
RETAIL$_{i,t}$	0.08	0.07	-0.18
	(0.23)	(0.23)	(0.23)
COMP_SOFTWARE$_{i,t}$	0.35	0.32	0.20
	(0.23)	(0.23)	(0.23)
LN_ASSETS$_{i,t-1}$	0.24***	0.24***	0.11***
	(0.02)	(0.02)	(0.03)
INTERCEPT	-5.80***	-5.72***	-5.44***
	(0.20)	(0.19)	(0.20)
N	17,065	17,065	17,065
Pseudo R^2	0.042	0.042	0.046
In-sample:			
Pred. Posit.	512	512	512
Correct Pred. Posit.	48	45	51
Precision	0.094	0.088	0.100
Sensitivity	0.124	0.116	0.132
False Negative Rate	0.876	0.884	0.868
Specificity	0.972	0.972	0.972
AUC	0.666	0.667	0.674

Table A.29, Continued

Out-of-sample:			
Precision	0.088	0.083	0.098
Sensitivity	0.119	0.111	0.132
False Negative Rate	0.881	0.889	0.868
Specificity	0.972	0.971	0.972
AUC	0.657	0.657	0.666

Panel B. 2017-2019

	1 $\text{SUED}_{i,t}$	2 $\text{SUED}_{i,t}$	3 $\text{SUED}_{i,t}$	4 $\text{SUED\_INV}_{i,t}$	5 $\text{SUED\_INV}_{i,t}$	6 $\text{SUED\_INV}_{i,t}$
$\text{PRED\_LN\_VIEWS}_{i,t-1}$			0.93***			1.08***
			(0.14)			(0.12)
$\text{PRED\_LITIG\_PROB-DETERM\_MODEL}_{Logit\ i,t-1}$		8.12***			9.22***	
		(1.19)			(1.07)	
$\text{PRED\_LITIG\_PROB-DETERM\_MODEL}_{OLS\ i,t-1}$	8.53***			10.26***		
	(1.27)			(1.11)		
$\text{BIOTECH}_{i,t}$	0.65***	0.61***	0.43**	0.69***	0.69***	0.44***
	(0.18)	(0.18)	(0.20)	(0.15)	(0.15)	(0.16)
$\text{COMP\_HARDWARE}_{i,t}$	0.23	0.20	0.10	0.20	0.18	0.06
	(0.39)	(0.39)	(0.39)	(0.31)	(0.32)	(0.32)
$\text{ELECTRONICS}_{i,t}$	0.24	0.24	0.11	-0.03	-0.03	-0.19
	(0.22)	(0.22)	(0.22)	(0.21)	(0.21)	(0.21)
$\text{RETAIL}_{i,t}$	0.04	0.04	-0.22	0.11	0.11	-0.20
	(0.24)	(0.24)	(0.25)	(0.19)	(0.19)	(0.20)
$\text{COMP\_SOFTWARE}_{i,t}$	-0.02	-0.05	-0.17	0.17	0.15	0.01
	(0.22)	(0.22)	(0.22)	(0.18)	(0.18)	(0.18)
$\text{LN\_ASSETS}_{i,t-1}$	0.13***	0.12***	0.00	0.14***	0.14***	-0.00
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
INTERCEPT	-4.23***	-4.18***	-3.86***	-3.78***	-3.70***	-3.33***
	(0.19)	(0.19)	(0.19)	(0.15)	(0.15)	(0.15)
N	9,133	9,133	9,133	9,133	9,133	9,133
Pseudo R^2	0.027	0.027	0.030	0.039	0.038	0.043
In-sample:						
Pred. Posit.	274	274	274	274	274	274
Correct Pred. Posit.	39	36	44	68	69	71
Precision	0.142	0.131	0.161	0.248	0.252	0.259
Sensitivity	0.084	0.078	0.095	0.082	0.084	0.086
False Negative Rate	0.916	0.922	0.905	0.918	0.916	0.914
Specificity	0.973	0.973	0.973	0.975	0.975	0.976
AUC	0.621	0.622	0.628	0.639	0.639	0.648
Out-of-sample:						
Precision	0.143	0.129	0.146	0.250	0.239	0.264
Sensitivity	0.086	0.078	0.089	0.085	0.081	0.090
False Negative Rate	0.914	0.922	0.911	0.915	0.919	0.910
Specificity	0.972	0.972	0.972	0.975	0.974	0.975
AUC	0.612	0.614	0.621	0.634	0.634	0.644

TABLE A.30
Determinants of Litigation Filings among Highly Scrutinized Firms

Table A.30 presents results based on estimating equation (VIII) using logistic estimation examining determinants of litigation filings among firms that are highly scrutinized by plaintiffs' lawyers during the prior year in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. Highly scrutinized firms are defined as firms in the top 25% (50%) of non-zero yearly views from either the top or remaining plaintiffs' lawyers in Columns 1–2 (3–4). To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED_{i,t}$	4 $SUED_{i,t}$
Defense Resources				
$TOP\_SECURITIES\_LAW\_FIRM_{i,t}$	0.16	0.15	0.11	0.09
	(0.22)	(0.22)	(0.17)	(0.17)
$HIGH\_COMP\_GC_{i,t}$	-0.19	-0.17	-0.17	-0.16
	(0.21)	(0.21)	(0.16)	(0.16)
Litigation Proceeds				
$CAR_{i,t}$	-1.16***	-1.15***	-1.08***	-1.09***
	(0.25)	(0.24)	(0.18)	(0.18)
$NY\_INC_{i,t}$	0.44	0.39	0.35	0.33
	(0.58)	(0.58)	(0.45)	(0.45)
Risk Environment				
$HIGH\_COMP\_CRO_{i,t}$	0.39	0.40	0.30	0.28
	(0.48)	(0.49)	(0.43)	(0.43)
$RISK\_MNGT\_COMMITTEE_{i,t}$	0.22	0.18	0.25	0.22
	(0.36)	(0.36)	(0.25)	(0.25)
$SUED_{i,t-1}$	-1.10***	-1.15***	-1.00***	-1.08***
	(0.37)	(0.38)	(0.33)	(0.34)
Voluntary Disclosures				
$VOLUNTARY\_8\text{-}Ks_{i,t}$	11.84*	12.82*	13.31***	14.28***
	(7.09)	(7.21)	(5.15)	(5.23)
$EARN\_WARN\_ANNCT_{i,t}$	-0.58**	-0.60***	-0.25	-0.23
	(0.23)	(0.23)	(0.17)	(0.17)
Insider Trading				
$FORMS\_3\text{-}5_{i,t}$	0.32	0.55	0.44	0.64
	(1.34)	(1.37)	(0.99)	(1.00)
Accounting Events				
$MAJOR\_RESTATE\_ANNCT_{i,t}$	2.96***	3.12***	2.18***	2.31***
	(0.64)	(0.67)	(0.49)	(0.50)
$ICW\_ANNCT_{i,t}$	-1.29**	-1.49**	-1.16**	-1.25**
	(0.64)	(0.67)	(0.49)	(0.50)
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t}$	-0.37	-0.32	0.67*	0.65*
	(0.59)	(0.60)	(0.35)	(0.35)
Controls				
$LN\_ASSETS_{i,t}$	-0.04	-0.04	-0.01	-0.02
	(0.06)	(0.06)	(0.05)	(0.04)
$BIOTECH_{i,t}$	1.11***	1.03***	1.00***	0.91***
	(0.33)	(0.34)	(0.27)	(0.27)
$COMP\_HARDWARE_{i,t}$	1.25**	1.26**	1.19***	1.16***
	(0.52)	(0.52)	(0.38)	(0.38)
$ELECTRONICS_{i,t}$	0.51	0.57	0.26	0.23
	(0.44)	(0.44)	(0.33)	(0.33)
$RETAIL_{i,t}$	0.43	0.34	0.31	0.24
	(0.36)	(0.36)	(0.27)	(0.27)
$COMP\_SOFTWARE_{i,t}$	0.83**	0.81**	0.73**	0.73**
	(0.37)	(0.37)	(0.30)	(0.30)
INTERCEPT	-2.83***	-3.00***	-3.36***	-3.41***
	(0.47)	(0.55)	(0.36)	(0.40)

Table A.30, Continued

Year FE	No	Yes	No	Yes
N	1,698	1,698	3,501	3,501
Pseudo R^2	0.106	0.118	0.076	0.087

TABLE A.31

Determinants of Litigation Filings among Highly Scrutinized Firms over a Longer Period

Table A.31 presents results similar to Table A.30 based on estimating equation (VIII) using logistic estimation examining determinants of litigation filings among firms that are highly scrutinized by plaintiffs' lawyers during the prior year in the firm-year sample from 2008–2016. Highly scrutinized firms are defined as firms in the top 25% (50%) of non-zero yearly views from either the top or remaining plaintiffs' lawyers in Columns 1–2 (3–4). To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2	3	4
	$SUED_{i,t}$	$SUED_{i,t}$	$SUED_{i,t}$	$SUED_{i,t}$
Defense Resources				
$TOP\_SECURITIES\_LAW\_FIRM_{i,t}$	0.21	0.21	0.12	0.11
	(0.19)	(0.20)	(0.15)	(0.15)
$HIGH\_COMP\_GC_{i,t}$	-0.06	-0.06	-0.11	-0.10
	(0.18)	(0.19)	(0.14)	(0.14)
Litigation Proceeds				
$CAR_{i,t}$	-1.24***	-1.25***	-1.05***	-1.08***
	(0.22)	(0.22)	(0.14)	(0.15)
$NY\_INC_{i,t}$	0.72**	0.67*	0.37	0.36
	(0.35)	(0.35)	(0.31)	(0.31)
Risk Environment				
$HIGH\_COMP\_CRO_{i,t}$	0.69*	0.71*	0.49	0.47
	(0.39)	(0.40)	(0.35)	(0.35)
$RISK\_MNGT\_COMMITTEE_{i,t}$	0.09	0.11	0.14	0.14
	(0.33)	(0.32)	(0.22)	(0.22)
$SUED_{i,t-1}$	-1.00***	-1.05***	-0.91***	-0.97***
	(0.33)	(0.35)	(0.29)	(0.30)
Voluntary Disclosures				
$VOLUNTARY\_8\text{-}Ks_{i,t}$	8.70	9.21	13.93***	14.54***
	(6.11)	(6.23)	(4.28)	(4.35)
$EARN\_WARN\_ANNCT_{i,t}$	-0.78***	-0.81***	-0.28*	-0.27*
	(0.21)	(0.21)	(0.15)	(0.15)
Insider Trading				
$FORMS\_3\text{-}5_{i,t}$	2.10**	2.29**	1.13	1.30*
	(0.97)	(0.98)	(0.74)	(0.75)
Accounting Events				
$MAJOR\_RESTATE\_ANNCT_{i,t}$	2.30***	2.40***	1.74***	1.82***
	(0.52)	(0.54)	(0.37)	(0.38)
$ICW\_ANNCT_{i,t}$	-0.83*	-1.01**	-0.76**	-0.81**
	(0.46)	(0.50)	(0.38)	(0.38)
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t}$	-0.14	-0.12	0.55**	0.54*
	(0.45)	(0.46)	(0.28)	(0.28)
Controls				
$LN\_ASSETS_{i,t}$	0.01	0.01	0.04	0.03
	(0.05)	(0.05)	(0.04)	(0.04)
$BIOTECH_{i,t}$	1.27***	1.25***	1.04***	0.98***
	(0.26)	(0.26)	(0.20)	(0.21)
$COMP\_HARDWARE_{i,t}$	0.96	0.89	0.95***	0.92**
	(0.59)	(0.61)	(0.37)	(0.37)
$ELECTRONICS_{i,t}$	0.20	0.15	0.20	0.18
	(0.37)	(0.38)	(0.28)	(0.28)
$RETAIL_{i,t}$	0.44	0.37	0.27	0.23
	(0.34)	(0.34)	(0.25)	(0.25)
$COMP\_SOFTWARE_{i,t}$	0.46	0.39	0.54**	0.53*
	(0.36)	(0.37)	(0.27)	(0.27)

Table A.31, Continued

INTERCEPT	-3.43***	-3.01***	-3.87***	-3.83***
	(0.44)	(0.52)	(0.31)	(0.40)
Year FE	No	Yes	No	Yes
N	2,291	2,291	4,747	4,747
Pseudo R^2	0.109	0.124	0.073	0.082

TABLE A.32
Determinants of Plaintiff-Lawyer Views—Expanding Voluntary Disclosure

Table A.32 presents results based on estimating equation (A) in the manuscript using OLS, similar to those presented in Table 9 examining determinants of scrutiny by top (remaining) plaintiffs' lawyers in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B after expanding the voluntary disclosure into individual components of forms. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 TOP_PLF_LN_VIEWS$_{i,t}$	2 REM_PLF_LN_VIEWS$_{i,t}$
Accounting Events		
AUDITOR_CHANGE_ANNCT$_{i,t}$	0.01	0.02
	(0.03)	(0.04)
MAJOR_RESTATE_ANNCT$_{i,t}$	0.44***	0.47***
	(0.08)	(0.09)
NON-TIMELY_FILING_ANNCT$_{i,t}$	0.12***	0.21***
	(0.03)	(0.05)
ICW_ANNCT$_{i,t}$	0.04	0.09**
	(0.03)	(0.04)
Personnel Events		
CEO_TURNOVER$_{i,t}$	0.10***	0.15***
	(0.02)	(0.03)
CFO_TURNOVER$_{i,t}$	0.06***	0.09***
	(0.02)	(0.03)
LN_8-Ks_ITEM2.02$_{i,t}$	0.05***	0.12***
	(0.01)	(0.02)
LN_8-Ks_ITEM7.01$_{i,t}$	-0.00	0.02
	(0.01)	(0.01)
LN_8-Ks_ITEM8.01$_{i,t}$	0.06***	0.17***
	(0.01)	(0.01)
EARN_WARN_ANNCT$_{i,t}$	0.05**	0.07**
	(0.02)	(0.03)
Earnings Characteristics		
POSITIVE_NON-GAAP_ADJ$_{i,t}$	0.02	0.04*
	(0.02)	(0.02)
POSITIVE_DISC_ACCR$_{i,t-1}$	-0.02	-0.04**
	(0.01)	(0.02)
Visibility		
LN_ASSETS$_{i,t-1}$	0.08***	0.17***
	(0.01)	(0.01)
LN_AGE$_{i,t-1}$	-0.05***	0.01
	(0.01)	(0.01)
Complexity		
MULTI-SEGMENTS$_{i,t-1}$	-0.03	-0.04**
	(0.02)	(0.02)
MULTINATIONAL$_{i,t-1}$	0.05***	0.12***
	(0.02)	(0.02)
LOSS$_{i,t-1}$	0.02	0.07***
	(0.02)	(0.02)
NO_DIVIDEND_PAID$_{i,t-1}$	0.07***	0.09***
	(0.02)	(0.02)
External Monitors		
BIG4$_{i,t-1}$	-0.02	-0.13***
	(0.02)	(0.03)
INSTIT_OWN$_{i,t-1}$	0.03	-0.00
	(0.03)	(0.04)
Market Turmoil		
CAR$_{i,t-1}$	-0.06***	-0.05**
	(0.02)	(0.02)

Table A.32, Continued

$RETURN\_VOL_{i,t-1}$	0.64***	0.86***
	(0.13)	(0.19)
$RETURN\_SKEW_{i,t-1}$	-0.02***	-0.03***
	(0.01)	(0.01)
$SHARE\_TURN_{i,t-1}$	0.03***	0.04***
	(0.00)	(0.01)
High Risk Industries		
$BIOTECH_{i,t}$	0.14***	0.44***
	(0.04)	(0.07)
$COMP\_HARDWARE_{i,t}$	0.12	0.23**
	(0.08)	(0.09)
$ELECTRONICS_{i,t}$	-0.03	0.00
	(0.03)	(0.04)
$RETAIL_{i,t}$	0.14***	0.13***
	(0.04)	(0.05)
$COMP\_SOFTWARE_{i,t}$	0.11***	0.15***
	(0.04)	(0.05)
INTERCEPT	-0.64***	-1.12***
	(0.06)	(0.07)
Year FE	Yes	Yes
N	17,179	17,179
R^2	0.094	0.178

TABLE A.33
Determinants of Litigation Filings among Highly Scrutinized Firms—Expanding Voluntary Disclosure

Table A.33 presents results similar to those presented in Table A.30 based on estimating equation (VIII) using logistic estimation expanding the voluntary disclosure into individual components of forms. Highly scrutinized firms are defined as firms in the top 25% (50%) of non-zero yearly views from either the most or the less fearful plaintiffs' lawyers in Columns 1–2 (3–4). To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 $SUED_{i,t}$	2 $SUED_{i,t}$	3 $SUED_{i,t}$	4 $SUED_{i,t}$
Defense Resources				
$TOP\_SECURITIES\_LAW\_FIRM_{i,t}$	0.17	0.15	0.12	0.10
	(0.22)	(0.23)	(0.17)	(0.17)
$HIGH\_COMP\_GC_{i,t}$	-0.19	-0.16	-0.18	-0.17
	(0.21)	(0.21)	(0.16)	(0.16)
Litigation Proceeds				
$CAR_{i,t}$	-1.17***	-1.16***	-1.09***	-1.10***
	(0.25)	(0.24)	(0.18)	(0.18)
$NY\_INC_{i,t}$	0.44	0.39	0.34	0.32
	(0.59)	(0.59)	(0.46)	(0.46)
Risk Environment				
$HIGH\_COMP\_CRO_{i,t}$	0.39	0.40	0.30	0.28
	(0.48)	(0.49)	(0.43)	(0.43)
$RISK\_MNGT\_COMMITTEE_{i,t}$	0.23	0.19	0.28	0.25
	(0.36)	(0.36)	(0.25)	(0.25)
$SUED_{i,t-1}$	-1.10***	-1.14***	-0.99***	-1.07***
	(0.36)	(0.37)	(0.33)	(0.34)
$VOLUNTARY\_8\text{-}Ks\_ITEM2.02_{i,t}$	12.85	14.27	27.35	29.87
	(38.29)	(38.84)	(31.70)	(31.61)
$VOLUNTARY\_8\text{-}Ks\_ITEM7.01_{i,t}$	2.62	2.87	1.35	1.54
	(8.96)	(9.24)	(7.06)	(7.18)
$VOLUNTARY8\text{-}Ks\_ITEM8.01_{i,t}$	13.55	14.79	13.19	14.30*
	(10.80)	(10.97)	(8.34)	(8.48)
$EARN\_WARN\_ANNCT_{i,t}$	-0.59***	-0.61***	-0.28*	-0.26
	(0.23)	(0.23)	(0.17)	(0.17)
Insider Trading				
$FORMS\_3\text{-}5_{i,t}$	0.34	0.57	0.41	0.61
	(1.34)	(1.37)	(0.99)	(1.01)
Accounting Events				
$MAJOR\_RESTATE\_ANNCT_{i,t}$	2.92***	3.09***	2.15***	2.28***
	(0.65)	(0.67)	(0.48)	(0.49)
$ICW\_ANNCT_{i,t}$	-1.27**	-1.47**	-1.13**	-1.23**
	(0.63)	(0.67)	(0.48)	(0.49)
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t}$	-0.34	-0.29	0.68**	0.66*
	(0.59)	(0.59)	(0.34)	(0.34)
Controls				
$LN\_ASSETS_{i,t}$	-0.03	-0.04	-0.01	-0.02
	(0.06)	(0.06)	(0.05)	(0.04)
$BIOTECH_{i,t}$	1.09***	1.01***	1.00***	0.91***
	(0.33)	(0.34)	(0.27)	(0.27)
$COMP\_HARDWARE_{i,t}$	1.21**	1.21**	1.13***	1.09***
	(0.51)	(0.51)	(0.38)	(0.38)
$ELECTRONICS_{i,t}$	0.48	0.53	0.21	0.18
	(0.44)	(0.44)	(0.33)	(0.33)
$RETAIL_{i,t}$	0.43	0.35	0.29	0.22
	(0.36)	(0.36)	(0.28)	(0.28)

Table A.33, Continued

COMP_SOFTWARE$_{i,t}$	0.80**	0.78**	0.71**	0.70**
	(0.37)	(0.37)	(0.30)	(0.30)
INTERCEPT	-2.84***	-3.01***	-3.48***	-3.54***
	(0.62)	(0.72)	(0.47)	(0.54)
Year FE	No	Yes	No	Yes
N	1,698	1,698	3,501	3,501
Pseudo R^2	0.105	0.116	0.074	0.086

TABLE A.34

Predicting SUED_INV$_{i,t}$ using a 7% Threshold

Table A.34 presents results similar to Table 6 Panel A when predicting SUED_INV$_{i,t}$, but instead of classifying cases as being predicted positive if the predicted probability is in the top 3% of all probabilities, we use a 7% cut-off value. Column 1 presents results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Column 2 presents results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Column 3 presents results based on combining the two groups of variables other than FPS$_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 SUED_INV$_{i,t}$	2 SUED_INV$_{i,t}$	3 SUED_INV$_{i,t}$
TOP_PLF_LN_VIEWS$_{i,t-1}$		0.28***	0.24***
		(0.03)	(0.04)
REM_PLF_LN_VIEWS$_{i,t-1}$		0.19***	0.17***
		(0.03)	(0.03)
FPS$_{i,t}$	0.26***		
	(0.08)		
LN_ASSETS$_{i,t-1}$	0.14***	0.03**	0.09***
	(0.02)	(0.01)	(0.02)
SALES_GR$_{i,t-1}$	0.62***		0.66***
	(0.13)		(0.13)
CAR$_{i,t-1}$	-0.03		0.06
	(0.07)		(0.07)
RETURN_SKEW$_{i,t-1}$	-0.11***		-0.07
	(0.04)		(0.04)
RETURN_VOL$_{i,t-1}$	4.40***		3.23***
	(0.55)		(0.56)
SHARE_TURN$_{i,t-1}$	0.06***		0.05***
	(0.01)		(0.01)
BIOTECH$_{i,t}$		0.75***	0.66***
		(0.12)	(0.12)
COMP_HARDWARE$_{i,t}$		0.62***	0.59***
		(0.18)	(0.18)
ELECTRONICS$_{i,t}$		-0.13	-0.17
		(0.13)	(0.13)
RETAIL$_{i,t}$		-0.01	-0.03
		(0.16)	(0.16)
COMP_SOFTWARE$_{i,t}$		0.09	0.05
		(0.14)	(0.14)
INTERCEPT	-4.39***	-3.17***	-4.04***
	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179
Pseudo R^2	0.029	0.035	0.050

Table A.34, Continued

In-sample:			
Pred. Posit.	1203	1203	1203
Correct Pred. Posit.	171	216	216
Precision	0.142	0.180	0.180
Sensitivity	0.147	0.186	0.186
False Negative Rate	0.853	0.814	0.814
Specificity	0.936	0.938	0.938
AUC	0.638	0.638	0.681
Out-of-sample:			
Precision	0.136	0.177	0.172
Sensitivity	0.142	0.184	0.179
False Negative Rate	0.858	0.816	0.821
Specificity	0.935	0.938	0.937
AUC	0.635	0.634	0.670

TABLE A.35

Predicting SUED_INV$_{i,t}$ using a 10% Threshold

Table A.35 presents results similar to Table 6 Panel A when predicting SUED_INV$_{i,t}$, but instead of classifying cases as being predicted positive if the predicted probability is in the top 3% of all probabilities, we use a 10% cut-off value. Column 1 presents results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7). Column 2 presents results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views. Column 3 presents results based on combining the two groups of variables other than FPS$_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1 SUED_INV$_{i,t}$	2 SUED_INV$_{i,t}$	3 SUED_INV$_{i,t}$
TOP_PLF_LN_VIEWS$_{i,t-1}$		0.28***	0.24***
		(0.03)	(0.04)
REM_PLF_LN_VIEWS$_{i,t-1}$		0.19***	0.17***
		(0.03)	(0.03)
FPS$_{i,t}$	0.26***		
	(0.08)		
LN_ASSETS$_{i,t-1}$	0.14***	0.03**	0.09***
	(0.02)	(0.01)	(0.02)
SALES_GR$_{i,t-1}$	0.62***		0.66***
	(0.13)		(0.13)
CAR$_{i,t-1}$	-0.03		0.06
	(0.07)		(0.07)
RETURN_SKEW$_{i,t-1}$	-0.11***		-0.07
	(0.04)		(0.04)
RETURN_VOL$_{i,t-1}$	4.40***		3.23***
	(0.55)		(0.56)
SHARE_TURN$_{i,t-1}$	0.06***		0.05***
	(0.01)		(0.01)
BIOTECH$_{i,t}$		0.75***	0.66***
		(0.12)	(0.12)
COMP_HARDWARE$_{i,t}$		0.62***	0.59***
		(0.18)	(0.18)
ELECTRONICS$_{i,t}$		-0.13	-0.17
		(0.13)	(0.13)
RETAIL$_{i,t}$		-0.01	-0.03
		(0.16)	(0.16)
COMP_SOFTWARE$_{i,t}$		0.09	0.05
		(0.14)	(0.14)
INTERCEPT	-4.39***	-3.17***	-4.04***
	(0.17)	(0.11)	(0.16)
N	17,179	17,179	17,179
Pseudo R^2	0.029	0.035	0.050

Table A.35, Continued

In-sample:			
Pred. Posit.	1718	1718	1718
Correct Pred. Posit.	231	280	289
Precision	0.134	0.163	0.168
Sensitivity	0.199	0.241	0.248
False Negative Rate	0.801	0.759	0.752
Specificity	0.907	0.910	0.911
AUC	0.638	0.638	0.681
Out-of-sample:			
Precision	0.130	0.163	0.163
Sensitivity	0.192	0.241	0.242
False Negative Rate	0.808	0.759	0.758
Specificity	0.907	0.910	0.910
AUC	0.635	0.634	0.670

TABLE A.36
Predicting Realized Litigation Risk Proxied with Plaintiff-Lawyer Investigations

Table A.36 presents results similar in nature to Table 6 Panel A examining proxies for ex ante litigation risk using only plaintiff-lawyer investigation announcements in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. Column 1 presents results based on estimating equation (3) in the manuscript using variables recommended by Kim and Skinner (2012) (Model 3 of their Table 7) but replacing the dependent variable to be INVESTIG_ANNCT.$_{i,t}$. Column 2 presents results based on estimating equation (3) in the manuscript using plaintiff-lawyer EDGAR views but replacing the dependent variable to be INVESTIG_ANNCT.$_{i,t}$. Columns 3 presents results based on combining the two groups of variables other than FPS$_{i,t}$ and estimating the equation. All equations use logistic estimation. Both in-sample and out-of-sample model performance metrics are reported. The out-of-sample statistics are calculated using the "K-fold" cross validation procedure similar to Dimmock and Gerken (2012), Kim and Skinner (2012), and Larcker and Zakolyukina (2012) with a 10-fold cross validation procedure. Precision is calculated as the percent of predicted positive cases that are true positives. Sensitivity is calculated as the percent of true positive cases correctly identified. Specificity is calculated as the percent of true negative cases correctly identified. We classify cases as being predicted positive if the predicted probability is in the top 3% of all probabilities. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2	3
	INVESTIG_ANNCT.$_{i,t}$	INVESTIG_ANNCT.$_{i,t}$	INVESTIG_ANNCT.$_{i,t}$
TOP_PLF_LN_VIEWS$_{i,t-1}$		0.27***	0.24***
		(0.04)	(0.04)
REM_PLF_LN_VIEWS$_{i,t-1}$		0.21***	0.18***
		(0.03)	(0.03)
FPS$_{i,t}$	0.25***		
	(0.08)		
LN_ASSETS$_{i,t-1}$	0.12***	0.01	0.06***
	(0.02)	(0.02)	(0.02)
SALES_GR$_{i,t-1}$	0.61***		0.65***
	(0.14)		(0.14)
CAR$_{i,t-1}$	0.06		0.15**
	(0.07)		(0.07)
RETURN_SKEW$_{i,t-1}$	-0.11**		-0.06
	(0.04)		(0.04)
RETURN_VOL$_{i,t-1}$	4.07***		2.85***
	(0.57)		(0.58)
SHARE_TURN$_{i,t-1}$	0.06***		0.05***
	(0.01)		(0.01)
BIOTECH$_{i,t}$		0.73***	0.64***
		(0.13)	(0.13)
COMP_HARDWARE$_{i,t}$		0.57***	0.55***
		(0.20)	(0.20)
ELECTRONICS$_{i,t}$		-0.14	-0.17
		(0.14)	(0.14)
RETAIL$_{i,t}$		-0.01	-0.04
		(0.17)	(0.17)
COMP_SOFTWARE$_{i,t}$		0.01	-0.03
		(0.15)	(0.15)
INTERCEPT	-4.25***	-3.10***	-3.88***
	(0.17)	(0.11)	(0.17)
N	17,179	17,179	17,179
Pseudo R^2	0.025	0.033	0.047

Table A.36, Continued

In-sample:			
Pred. Posit.	516	516	516
Correct Pred. Posit.	72	107	107
Precision	0.140	0.207	0.207
Sensitivity	0.067	0.100	0.100
False Negative Rate	0.933	0.900	0.900
Specificity	0.972	0.975	0.975
AUC	0.631	0.633	0.677
Out-of-sample:			
Precision	0.133	0.202	0.194
Sensitivity	0.065	0.098	0.095
False Negative Rate	0.935	0.902	0.905
Specificity	0.972	0.974	0.974
AUC	0.627	0.628	0.665

TABLE A.37

Association of Prior and Current Quarter's Returns with Current Quarter's Plaintiff-Lawyer Views

Table A.37 presents results using a specification similar to Table 8 examining whether prior and current quarter's abnormal returns are associated with current quarter plaintiff-lawyer views by estimating a modified B in the manuscript in the firm-quarter sample from 2012–2016 as shown in Table A.2 Panel C. We use a Fama-French 4 factor model to calculate expected returns. Panel A examines all firm-quarters, while Panel B (C) excludes quarters in which litigation is filed (class period end occurs for subsequent litigation). To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Refer to Table A.41 for variable definitions.

Panel A. All Quarters

	1 $LN\_VIEWS_{i,t}$	2 $REM\_PLF\_LN\_VIEWS_{i,t}$	3 $TOP\_PLF\_LN\_VIEWS_{i,t}$
$BHAR_{i,t}$	-0.13***	-0.06***	-0.10***
	(0.01)	(0.01)	(0.01)
$BHAR_{i,t-1}$	-0.19***	-0.12***	-0.11***
	(0.02)	(0.01)	(0.01)
$LN\_MVE_{i,t-1}$	0.07***	0.06***	0.02***
	(0.00)	(0.00)	(0.00)
$LN\_BOOK\text{-}TO\text{-}MARKET_{i,t-1}$	-0.06***	-0.05***	-0.01**
	(0.01)	(0.01)	(0.01)
$LN\_TURNOVER_{i,t-1}$	0.08***	0.06***	0.03***
	(0.01)	(0.00)	(0.00)
$ALPHA_{i,t-1}$	-42.52***	-25.78***	-22.97***
	(3.37)	(2.66)	(1.89)
$INSTIT\_OWN_{i,t-1}$	-0.02	-0.02	-0.01
	(0.02)	(0.02)	(0.01)
$NASDAQ_{i,t-1}$	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)
INTERCEPT	-0.20**	-0.13	-0.11***
	(0.09)	(0.08)	(0.02)
FF48 FE	Yes	Yes	Yes
Fyear x Qtr FE	Yes	Yes	Yes
N	87,338	87,338	87,338
R^2	0.081	0.070	0.042

Panel B. Excluding Quarters with a Filing in the Prior Quarter

	1 $LN\_VIEWS_{i,t}$	2 $REM\_PLF\_LN\_VIEWS_{i,t}$	3 $TOP\_PLF\_LN\_VIEWS_{i,t}$
$BHAR_{i,t}$	-0.13***	-0.06***	-0.10***
	(0.01)	(0.01)	(0.01)
$BHAR_{i,t-1}$	-0.14***	-0.08***	-0.09***
	(0.02)	(0.01)	(0.01)
Controls Included	Yes	Yes	Yes
FF48 FE	Yes	Yes	Yes
Fyear x Qtr FE	Yes	Yes	Yes
N	86,715	86,715	86,715
R^2	0.078	0.068	0.039

Panel C. Excluding Quarters with a Class Period End Occurring in the Prior Quarter

	1 $LN\_VIEWS_{i,t}$	2 $REM\_PLF\_LN\_VIEWS_{i,t}$	3 $TOP\_PLF\_LN\_VIEWS_{i,t}$
$BHAR_{i,t}$	-0.13***	-0.06***	-0.10***
	(0.01)	(0.01)	(0.01)
$BHAR_{i,t-1}$	-0.17***	-0.10***	-0.10***
	(0.02)	(0.01)	(0.01)
Controls Included	Yes	Yes	Yes
FF48 FE	Yes	Yes	Yes
Fyear x Qtr FE	Yes	Yes	Yes
N	87,227	87,227	87,227
R^2	0.080	0.069	0.041

TABLE A.38

Determinants of Plaintiff-Lawyer Views using Contemporaneous Values for All Variables

Table A.38 presents results similar to Table 9, but we modify equation (A) in the manuscript to use current year data for all variables. Column 1 (2) presents results based on estimating the modified equation using OLS examining determinants of scrutiny by top (remaining) plaintiffs' lawyers in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2
	$TOP\_PLF\_LN\_VIEWS_{i,t}$	$REM\_PLF\_LN\_VIEWS_{i,t}$
Accounting Events		
$AUDITOR\_CHANGE\_ANNCT_{i,t}$	0.00	0.02
	(0.03)	(0.04)
$MAJOR\_RESTATE\_ANNCT_{i,t}$	0.42***	0.45***
	(0.08)	(0.09)
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t}$	0.10***	0.19***
	(0.03)	(0.05)
$ICW\_ANNCT_{i,t}$	0.04	0.11***
	(0.03)	(0.04)
Personnel Events		
$CEO\_TURNOVER_{i,t}$	0.08***	0.14***
	(0.02)	(0.03)
$CFO\_TURNOVER_{i,t}$	0.05**	0.07***
	(0.02)	(0.03)
Disclosure		
$LN\_VOLUNT\_8\text{-}Ks_{i,t}$	0.06***	0.20***
	(0.01)	(0.01)
$EARN\_WARN\_ANNCT_{i,t}$	0.04*	0.06**
	(0.02)	(0.03)
Earnings Characteristics		
$POSITIVE\_NON\text{-}GAAP\_ADJ_{i,t}$	0.03	0.04*
	(0.02)	(0.02)
$POSITIVE\_DISC\_ACCR_{i,t}$	-0.04***	-0.04**
	(0.01)	(0.02)
Visibility		
$LN\_ASSETS_{i,t}$	0.08***	0.18***
	(0.01)	(0.01)
$LN\_AGE_{i,t}$	-0.03**	0.03**
	(0.01)	(0.02)
Complexity		
$MULTI\text{-}SEGMENTS_{i,t}$	-0.03*	-0.05**
	(0.02)	(0.02)
$MULTINATIONAL_{i,t}$	0.06***	0.12***
	(0.02)	(0.02)
$LOSS_{i,t}$	0.01	0.07***
	(0.02)	(0.02)
$NO\_DIVIDEND\_PAID_{i,t}$	0.06***	0.07***
	(0.02)	(0.02)
External Monitors		
$BIG4_{i,t}$	-0.01	-0.11***
	(0.02)	(0.03)
$INSTIT\_OWN_{i,t}$	0.04	-0.04
	(0.03)	(0.04)

Table A.38, Continued

Market Turmoil		
$\text{CAR}_{i,t}$	-0.16***	-0.15***
	(0.02)	(0.02)
$\text{RETURN\_VOL}_{i,t}$	1.47***	1.20***
	(0.14)	(0.18)
$\text{RETURN\_SKEW}_{i,t}$	-0.06***	-0.01
	(0.01)	(0.01)
$\text{SHARE\_TURN}_{i,t}$	0.03***	0.04***
	(0.00)	(0.01)
High Risk Industries		
$\text{BIOTECH}_{i,t}$	0.12***	0.45***
	(0.04)	(0.07)
$\text{COMP\_HARDWARE}_{i,t}$	0.12	0.27***
	(0.08)	(0.10)
$\text{ELECTRONICS}_{i,t}$	-0.03	0.03
	(0.03)	(0.04)
$\text{RETAIL}_{i,t}$	0.16***	0.15***
	(0.04)	(0.05)
$\text{COMP\_SOFTWARE}_{i,t}$	0.11***	0.17***
	(0.04)	(0.05)
INTERCEPT	-0.79***	-1.31***
	(0.06)	(0.08)
Year FE	Yes	Yes
N	16,963	16,963
R^2	0.114	0.182

TABLE A.39

Determinants of Plaintiff-Lawyer Views using Lagged Values for All Variables

Table A.39 presents results similar to Table 9, but we modify equation (A) in the manuscript to use prior year data for all independent variables. Column 1 (2) presents results based on estimating the modified equation using OLS examining determinants of scrutiny by top (remaining) plaintiffs' lawyers in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2
	$TOP\_PLF\_LN\_VIEWS_{i,t}$	$REM\_PLF\_LN\_VIEWS_{i,t}$
Accounting Events		
$AUDITOR\_CHANGE\_ANNCT_{i,t-1}$	-0.05**	-0.07*
	(0.03)	(0.04)
$MAJOR\_RESTATE\_ANNCT_{i,t-1}$	0.20***	0.19**
	(0.06)	(0.08)
$NON\text{-}TIMELY\_FILING\_ANNCT_{i,t-1}$	0.07**	0.09**
	(0.03)	(0.05)
$ICW\_ANNCT_{i,t-1}$	0.07**	0.07*
	(0.03)	(0.04)
Personnel Events		
$CEO\_TURNOVER_{i,t-1}$	0.04*	0.03
	(0.02)	(0.03)
$CFO\_TURNOVER_{i,t-1}$	0.01	0.08***
	(0.02)	(0.03)
Disclosure		
$LN\_VOLUNT\_8\text{-}Ks_{i,t-1}$	0.05***	0.18***
	(0.01)	(0.01)
$EARN\_WARN\_ANNCT_{i,t-1}$	0.02	0.07***
	(0.02)	(0.03)
Earnings Characteristics		
$POSITIVE\_NON\text{-}GAAP\_ADJ_{i,t-1}$	0.05***	0.06***
	(0.02)	(0.02)
$POSITIVE\_DISC\_ACCR_{i,t-1}$	-0.02	-0.04**
	(0.01)	(0.02)
Visibility		
$LN\_ASSETS_{i,t-1}$	0.08***	0.17***
	(0.01)	(0.01)
$LN\_AGE_{i,t-1}$	-0.05***	0.00
	(0.01)	(0.01)
Complexity		
$MULTI\text{-}SEGMENTS_{i,t-1}$	-0.03	-0.04*
	(0.02)	(0.02)
$MULTINATIONAL_{i,t-1}$	0.05***	0.13***
	(0.02)	(0.02)
$LOSS_{i,t-1}$	0.03*	0.10***
	(0.02)	(0.02)
$NO\_DIVIDEND\_PAID_{i,t-1}$	0.07***	0.09***
	(0.02)	(0.02)
External Monitors		
$BIG4_{i,t-1}$	-0.02	-0.14***
	(0.02)	(0.03)

Table A.39, Continued

INSTIT_OWN$_{i,t-1}$	0.04	-0.02
	(0.03)	(0.04)
Market Turmoil		
CAR$_{i,t-1}$	-0.06***	-0.04*
	(0.02)	(0.02)
RETURN_VOL$_{i,t-1}$	0.64***	0.91***
	(0.13)	(0.19)
RETURN_SKEW$_{i,t-1}$	-0.02***	-0.03***
	(0.01)	(0.01)
SHARE_TURN$_{i,t-1}$	0.03***	0.04***
	(0.00)	(0.01)
High Risk Industries		
BIOTECH$_{i,t}$	0.15***	0.47***
	(0.04)	(0.07)
COMP_HARDWARE$_{i,t}$	0.12	0.24**
	(0.08)	(0.10)
ELECTRONICS$_{i,t}$	-0.04	-0.02
	(0.03)	(0.04)
RETAIL$_{i,t}$	0.15***	0.15***
	(0.04)	(0.05)
COMP_SOFTWARE$_{i,t}$	0.10**	0.14***
	(0.04)	(0.05)
INTERCEPT	-0.60***	-1.08***
	(0.05)	(0.07)
Year FE	Yes	Yes
N	17,179	17,179
R^2	0.082	0.160

TABLE A.40
Determinants of Plaintiff-Lawyer Views After including an Indicator Variable for Firm-Years Missing IBES Coverage

Table A.40 presents results similar in nature to Table 9 after modifying equation (A) in the manuscript to include an indicator variable for firms missing IBES coverage. Column 1 (2) presents results based on estimating the modified equation examining determinants of scrutiny by top (remaining) plaintiffs' lawyers in the firm-year sample from 2012–2016 as shown in Table A.2 Panel B. To minimize the influence of outliers, all continuous variables are winsorized at the 1% and 99% levels. Standard errors appear in parentheses and are adjusted for clustering at the company level. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Refer to Table A.41 for variable definitions.

	1	2
	TOP_PLF_LN_VIEWS$_{i,t}$	REM_PLF_LN_VIEWS$_{i,t}$
Accounting Events		
AUDITOR_CHANGE_ANNCT$_{i,t}$	0.01	0.03
	(0.03)	(0.04)
MAJOR_RESTATE_ANNCT$_{i,t}$	0.44***	0.46***
	(0.08)	(0.09)
NON-TIMELY_FILING_ANNCT$_{i,t}$	0.12***	0.21***
	(0.03)	(0.05)
ICW_ANNCT$_{i,t}$	0.04	0.10**
	(0.03)	(0.04)
Personnel Events		
CEO_TURNOVER$_{i,t}$	0.10***	0.15***
	(0.02)	(0.03)
CFO_TURNOVER$_{i,t}$	0.06***	0.08***
	(0.02)	(0.03)
Disclosure		
LN_VOLUNT_8-Ks$_{i,t}$	0.07***	0.21***
	(0.01)	(0.01)
EARN_WARN_ANNCT$_{i,t}$	0.05**	0.05*
	(0.02)	(0.03)
MISSING_IBES_COVERAGE$_{i,t}$	0.01	-0.03
	(0.02)	(0.03)
Earnings Characteristics		
POSITIVE_NON-GAAP_ADJ$_{i,t}$	0.03	0.03
	(0.02)	(0.02)
POSITIVE_DISC_ACCR$_{i,t-1}$	-0.02	-0.04**
	(0.01)	(0.02)
Visibility		
LN_ASSETS$_{i,t-1}$	0.08***	0.18***
	(0.01)	(0.01)
LN_AGE$_{i,t-1}$	-0.05***	0.00
	(0.01)	(0.02)
Complexity		
MULTI-SEGMENTS$_{i,t-1}$	-0.03	-0.04**
	(0.02)	(0.02)
MULTINATIONAL$_{i,t-1}$	0.05***	0.12***
	(0.02)	(0.02)
LOSS$_{i,t-1}$	0.03	0.08***
	(0.02)	(0.02)
NO_DIVIDEND_PAID$_{i,t-1}$	0.07***	0.08***
	(0.02)	(0.02)
External Monitors		
BIG4$_{i,t-1}$	-0.01	-0.13***
	(0.02)	(0.03)
INSTIT_OWN$_{i,t-1}$	0.03	-0.04
	(0.03)	(0.04)

Table A.40, Continued

Market Turmoil		
$\text{CAR}_{i,t-1}$	-0.05***	-0.04**
	(0.02)	(0.02)
$\text{RETURN\_VOL}_{i,t-1}$	0.64***	0.88***
	(0.13)	(0.18)
$\text{RETURN\_SKEW}_{i,t-1}$	-0.02***	-0.03***
	(0.01)	(0.01)
$\text{SHARE\_TURN}_{i,t-1}$	0.03***	0.04***
	(0.00)	(0.01)
High Risk Industries		
$\text{BIOTECH}_{i,t}$	0.15***	0.46***
	(0.04)	(0.07)
$\text{COMP\_HARDWARE}_{i,t}$	0.12	0.25***
	(0.08)	(0.10)
$\text{ELECTRONICS}_{i,t}$	-0.03	0.00
	(0.03)	(0.04)
$\text{RETAIL}_{i,t}$	0.15***	0.14***
	(0.04)	(0.05)
$\text{COMP\_SOFTWARE}_{i,t}$	0.11***	0.15***
	(0.04)	(0.05)
INTERCEPT	-0.66***	-1.15***
	(0.06)	(0.08)
Year FE	Yes	Yes
N	17,179	17,179
R^2	0.092	0.175

TABLE A.41

Online Appendix Variable Definitions

Variable	Definition
$\text{ALPHA}_{i,t}$	Alpha, based on a Fama-French 4 factor model. We estimate the model over days [−252, −1] relative to fiscal quarter's t end and require at least 60 observations of daily returns to estimate the model.
$\text{ARCSINH\_VIEWS}_{i,[Class\ End,\ Filing-1]}$	Arcsinh (inverse hyperbolic sine) of the total number of EDGAR views by plaintiffs' law firms from the class period end up to, but not including, the litigation filing day.
$\text{ARCTAN\_VIEWS}_{i,[Class\ End,\ Filing-1]}$	Arctan (inverse of the tangent) of the total number of EDGAR views by plaintiffs' law firms from the class period end up to, but not including, the litigation filing day.
$\text{AUDITOR\_CHANGE\_ANNCT}_{i,t}$	An indicator variable set to one if there is an auditor change announcement during the fiscal year; zero otherwise.
$\text{BETA}_{i,t}$	The slope coefficient from a regression of daily returns on the CRSP equal-weighted index. The regression is estimated for each firm for each year. We require at least 20 observations.
$\text{BHR}_{i,t}$	Cumulative buy-and-hold return during the fiscal year based on daily returns.
$\text{BHAR}_{i,t+1}$	Buy-and-hold abnormal returns over the period [1, 60] days relative to fiscal quarter's t end. Expected returns are calculated using a Fama-French 4 factor model. We estimate the model over days [−252, −1] relative to fiscal quarter's t end and require at least 60 observations of daily returns to estimate the model.
$\text{BIG4}_{i,t}$	An indicator variable set to one if the firm is audited by one of the top four audit firms (i.e., the Big 4); zero otherwise.
$\text{BIOTECH}_{i,t}$	An indicator variable set to one if the firm's SIC code is between 2833 and 2836; zero otherwise.
$\text{CAR}_{i,t}$	Cumulative abnormal return during the fiscal year based on monthly returns. For sued firms in Tables A.30, A.31, and A.33, the period is limited from the beginning of the year to the day prior to the filing date using daily returns.
$\text{CEO\_TURNOVER}_{i,t}$	An indicator variable set to one if the firm announces CEO turnover during the fiscal year; zero otherwise. We use the Audit Analytics' Director and Officer Changes data set to maximize coverage. We exclude cases in which ACTION per Audit Analytics is set to "Appointed", "Retracted Resignation", "Re-elected", "Change Misreported", "Nominated", "Returned to Position", or "Engaged".

Table A.41, Continued

Variable	Definition
CFO_TURNOVER$_{i,t}$	An indicator variable set to one if the firm announces CFO turnover during the fiscal year; zero otherwise. We use the Audit Analytics' Director and Officer Changes data set to maximize coverage. We exclude cases in which ACTION per Audit Analytics is set to "Appointed", "Retracted Resignation", "Re-elected", "Change Misreported", "Nominated", "Returned to Position", or "Engaged".
COMP_HARDWARE$_{i,t}$	An indicator variable set to one if the firm's SIC code is between 3570 and 3577; zero otherwise.
COMP_SOFTWARE$_{i,t}$	An indicator variable set to one if the firm's SIC code is between 7371 and 7379; zero otherwise.
EARN_WARN_ANNCT$_{i,t}$	An indicator variable set to one if the firm provides earning warnings during the fiscal year per IBES guidance; zero otherwise. For sued firms in Tables A.30, A.31, and A.33, the period is limited from the beginning of the year to the day prior to the filing date. We rely on IBES guidance codes to identify earning warnings (i.e., cases in which the guidance code is equal to one). The variable is set to zero for firms missing IBES coverage.
ELECTRONICS$_{i,t}$	An indicator variable set to one if the firm's SIC code is between 3600 and 3674; zero otherwise.
EX_ANTE_LIT_RISK$_{i,t}$	Ex ante litigation risk per Kim and Skinner (2012) (i.e., Model 3 of their Table 7). We use the log odds value (i.e., we do not convert the predicted value to probability).
FORMS_3-5$_{i,t}$	Number of Forms 3, 4, and 5 the firm submitted on EDGAR during the fiscal year. For sued firms, the number of filings are limited to the day prior to filing date. We scale by the number of days in the relevant period. We also include amended filings.
FPS$_{i,t}$	Indicator variable set to one if the firm is in a high litigation risk industry as defined in Francis, Philbrick, and Schipper (1994); zero otherwise. Specifically, it is set to one if the firm in in any of the following industries: biotechnology (SIC codes 2833–2836 and 8731–8734), computers (SIC codes 3570–3577 and 7370–7374), electronics (SIC codes 3600–3674), or retailing (SIC codes 5200–5961).
HIGH_COMP_CRO$_{i,t}$	An indicator variable set to one if the "Chief Risk Officer" or "Chief Compliance Officer" is listed as a top paid executive in the financial statements or proxy statement submitted on EDGAR during the fiscal year; zero otherwise.
HIGH_COMP_GC$_{i,t}$	An indicator variable set to one if a general counsel or Chief Legal Officer is listed as a top paid executive in the financial statements or proxy statement during year t; zero otherwise.

Table A.41, Continued

Variable	Definition
ICW_ANNCT$_{i,t}$	An indicator variable set to one if the management (auditor) announces ineffective internal controls under Sarbanes-Oxley Act (SOX) Section 302 (404) during the fiscal year; zero otherwise.
INSTIT_OWN$_{i,t}$	Proportion of institutional ownership as of fiscal year end. For Tables A.5 and A.37, the variable is calculated as of fiscal quarter end. Missing values are set to zero.
INVESTIG_ANNCT.$_{i,t}$	An indicator variable set to one if an investigation by a plaintiffs' law firm is announced against the firm during the fiscal year; zero otherwise. To identify investigation announcements, we search for press release newswires on Factiva that contain: 1) "announces investigation" or 2) "investigating" and "on behalf" and then match the targeted firms to our data set.
LIBERAL_COURT$_{i,t-1}$	Federal judge ideology in firm's i circuit as of the prior fiscal year per Huang et al. (2019).
LN_8-Ks_ITEM2.02$_{i,t}$	Defined similar to LN_VOLUNT_8-Ks$_{i,t}$, but only includes 8-Ks with Item Code 2.02.
LN_8-Ks_ITEM7.01$_{i,t}$	Defined similar to LN_VOLUNT_8-Ks$_{i,t}$, but only includes 8-Ks with Item Code 7.01.
LN_8-Ks_ITEM8.01$_{i,t}$	Defined similar to LN_VOLUNT_8-Ks$_{i,t}$, but only includes 8-Ks with Item Code 8.01.
LN_AGE$_{i,t}$	Natural logarithm of one plus the number of years since the firm first appeared on Compustat.
LN_ASSETS$_{i,t}$	Natural logarithm of total assets, in millions, at the fiscal year end.
LN_BOOK-TO-MARKET$_{i,t}$	Natural logarithm of the book-to-market ratio, calculated as the sum of quarter t total liabilities and market value of equity, scaled by total assets.
LN_DAMAGES$_{i,t}$	Natural logarithm of the difference between the maximum market capitalization during the class period less the market capitalization the day following class period end. Market capitalization is calculated in actual dollar value.
LN_MVE$_{i,t}$	Natural logarithm of the market value of equity, in millions, at the end of the fiscal year. For Tables A.5 and A.37, the variable is calculated as of fiscal quarter end.
LN_MVE$_{v1\ i,t}$	Average daily market value of equity, in thousands, during the fiscal year.
LN_SETTLEM$_{i,t}$	Natural logarithm of the securities class action settlement amount.
LN_TURNOVER$_{i,t}$	Natural logarithm of share turnover. Share turnover is defined as split-adjusted trading volume, scaled by shares outstanding during the first day of the fiscal quarter.

Table A.41, Continued

Variable	Definition
LN_VIEWS$_{i,t}$	Natural logarithm of one plus the total number of EDGAR views by plaintiffs' law firms during the fiscal year. For Tables A.5 and A.37, the variable is calculated as the total number of EDGAR views during the fiscal quarter. We exclude index and web crawler views.
LN_VOLUNT_8-Ks$_{i,t}$	Natural logarithm of one plus the number of Form 8-Ks with Item Codes 2.02, 7.01, or 8.01 that the firm submitted on EDGAR during the fiscal year (see Bourveau et al. (2018), He and Plumlee (2020)). We also include amended filings.
LOSS$_{i,t}$	An indicator variable set to one if the firm's net income for the fiscal year is negative; zero otherwise.
MAJOR_RESTATE_ANNCT$_{i,t}$	An indicator variable set to one if a major restatement (i.e., disclosed via Item 4.02 in an 8-K) is announced during the fiscal year; zero otherwise. We use the filing date of the original restatement announcement, rather than the date of the 8-K filing with Item 4.02, which may occur on a subsequent date.
MISSING_IBES_COVERAGE$_{i,t}$	An indicator variable set to one if the firm is not covered by IBES guidance of if the range of the earliest and latest guidance in IBES does not include the current year; zero otherwise.
MULTI-SEGMENTS$_{i,t}$	An indicator variable set to one if the firm has more than one business segment in the fiscal year; zero otherwise. Missing values for the number of business segments are set to one.
MULTINATIONAL$_{i,t}$	An indicator variable set to one if the firm has pre-tax foreign income in the fiscal year; zero otherwise.
NASDAQ$_{i,t}$	An indicator variable set to one if the firm is trading on NASDAQ as of fiscal quarter end; zero otherwise.
NO_DIVIDEND_PAID$_{i,t}$	An indicator variable set to one if the firm did not pay dividends during the fiscal year; zero otherwise.
NON-TIMELY_FILING_ANNCT$_{i,t}$	An indicator variable set to one if the firm submits a non-timely filing during the fiscal year; zero otherwise.
NY_INC$_{i,t}$	An indicator variable set to one if the firm is incorporated in New York in the fiscal year; zero otherwise.

Table A.41, Continued

Variable	Definition
POSITIVE_DISC_ACCR$_{i,t-1}$	An indicator variable set to one if the prior fiscal year's modified Jones discretionary accruals are positive as per Dechow, Sloan, and Sweeney (1995); zero otherwise. We require at least 15 observations in a given year-SIC2 industry to calculate discretionary accruals. Accruals are calculated as net income minus cash flows from operations.
POSITIVE_NON-GAAP_ADJ$_{i,t}$	An indicator variable set to one if quarterly GAAP-reported EPS (epsfiq per Compustat) announced during the fiscal year is less than management-provided non-GAAP EPS as per Bentley et al. (2018); zero otherwise. If the manager does not provide non-GAAP EPS, the variable is set to zero.
PRED_LITIG_PROB-DETERM_MODEL$_{Logit\ i,t}$	Predicted probability for litigation filing after re-estimating the model from Appendix C in the manuscript using SUED$_{i,t}$ as the dependent variable. The model is estimated using a logistic regression.
PRED_LITIG_PROB-DETERM_MODEL$_{OLS\ i,t}$	Predicted probability for litigation filing after re-estimating the model from Appendix C in the manuscript using SUED$_{i,t}$ as the dependent variable. The model is estimated using OLS regression.
PRED_LN_VIEWS$_{i,t}$	Predicted number of EDGAR views by all plaintiffs' lawyers using the model from Appendix C.
PRED_WITHOUT_BAD_NEWS_LN_VIEWS$_{i,t-1}$	Predicted number of EDGAR views by all plaintiffs' lawyers based on a model similar to Appendix C estimated after excluding the following categories: Accounting Events, Personnel Events, and Disclosure.
REM_PLF_ARCSINH_VIEWS$_{i,t}$	Arcsinh (inverse hyperbolic sine) of the total number of EDGAR views by plaintiffs' law firms not ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) during the fiscal year.
REM_PLF_ARCSINH_VIEWS$_{i,[Class\ End,\ Filing-1]}$	Arcsinh (inverse hyperbolic sine) of the total number of EDGAR views by plaintiffs' law firms not ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) from the class period end up to, but not including, the litigation filing day.
REM_PLF_ARCTAN_VIEWS$_{i,t}$	Arctan (inverse the tangent) of the total number of EDGAR views by plaintiffs' law firms not ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) during the fiscal year.
REM_PLF_ARCTAN_VIEWS$_{i,[Class\ End,\ Filing-1]}$	Arctan (inverse tangent) of the total number of EDGAR views by plaintiffs' law firms not ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) from the class period end up to, but not including, the litigation filing day.

Table A.41, Continued

Variable	Definition
REM_PLF_LN_VIEWS$_{i,t}$	Natural logarithm of one plus the total number of EDGAR views by plaintiffs' law firms not ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) during the fiscal year. For Tables A.5 and A.37, the variable is calculated as the total number of EDGAR views during the fiscal quarter.
REM_PLF_LN_VIEWS-ORTHOGONAL$_{i,t-1}$	Residuals from a model regressing REM_PLF_LN_VIEWS$_{i,t-1}$ on indicator variables set to one for litigation filings; earnings warnings; and announcement of major restatements, internal control weaknesses, CEO turnover, CFO turnover, non-timely filings, or auditor changes; and a count of large daily negative market-adjusted returns (i.e., $<$ –10%). All variables are measured with respect to year t–1 to be consistent with plaintiff-lawyer views.
RETAIL$_{i,t}$	An indicator variable set to one if the firm's SIC code is between 5200 and 5961; zero otherwise.
RETURN_SKEW$_{i,t}$	Skewness of monthly raw returns during the fiscal year.
RETURN_SKEW$_{v1\ i,t}$	Skewness of daily raw returns during the fiscal year.
RETURN_VOL$_{i,t}$	Standard deviation of monthly returns during the fiscal year. For Tables A.5 and A.37, the variable is calculated using daily returns over the fiscal quarter.
RETURN_VOL$_{v1\ i,t}$	Standard deviation of daily returns during the fiscal year.
RISK_MNGT_COMMITTEE$_{i,t}$	An indicator variable set to one if "Risk Management Committee" is mentioned in the financial statements or proxy statement submitted on EDGAR during the fiscal year; zero otherwise.
ROA$_{i,t}$	Net income scaled by total assets as of the end of the fiscal year.
SALES_GR$_{i,t-1}$	Sales growth measured as the change in net sales from fiscal year t–2 to t–1, divided by total assets as of t–2.
SETTLED$_{i,t}$	An indicator variable set to one if the firm is sued during the fiscal year and the lawsuit is eventually settled; zero otherwise. We require the lawsuit to contain fraud allegations (i.e., alleging violations of Rule 10b-5).
SHARE_TURN$_{i,t}$	Share turnover, defined as split-adjusted trading volume scaled by shares outstanding during the first month of the fiscal year.
SHARE_TURN$_{v1\ i,t}$	Average daily trading volume scaled by the average shares outstanding.
SUED$_{i,t}$	An indicator variable set to one if a securities class action is filed against the firm during the fiscal year; zero otherwise.

Table A.41, Continued

Variable	Definition
SUED_INV$_{i,t}$	An indicator variable set to one if SUED$_{i,t}$ is equal to one or INVESTIG_ANNCT.$_{i,t}$ is equal to one in the fiscal year; zero otherwise.
TOBINSQ$_{i,t}$	The sum of the market value of common stock, preferred stock, and firm debt, scaled by total assets as of the end of the fiscal year. Preferred stock and debt are assumed to have a market value equal to book value. Missing values for debt are set equal to zero.
TOP_PLF_ARCSINH_VIEWS$_{i,t}$	Arcsinh (inverse hyperbolic sine) of the total number of EDGAR views by plaintiffs' law firms ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) during the fiscal year.
TOP_PLF_ARCSINH_VIEWS$_{i,[Class\ End,\ Filing-1]}$	Arcsinh (inverse hyperbolic sine) of the total number of EDGAR views by plaintiffs' law firms ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) from the class period end up to, but not including, the litigation filing day.
TOP_PLF_ARCTAN_VIEWS$_{i,t}$	Arctan (inverse tangent) of the total number of EDGAR views by plaintiffs' law firms ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) during the fiscal year.
TOP_PLF_ARCTAN_VIEWS$_{i,[Class\ End,\ Filing-1]}$	Arctan (inverse tangent) of the total number of EDGAR views by plaintiffs' law firms ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 (see TOP_PLF_LN_VIEWS$_{i,t}$) from the class period end up to, but not including, the litigation filing day.
TOP_PLF_LN_VIEWS$_{i,t}$	Natural logarithm of one plus the total number of EDGAR views by plaintiffs' law firms ranked among the top plaintiffs' law firms by Chambers and Partners as of June 2011 during the fiscal year. For Tables A.5 and A.37, the variable is calculated as the total number of EDGAR views during the fiscal quarter. The list of top plaintiff' law firms includes: 1) Bernstein, Litowitz, Berger & Grossmann LLP, 2) Grant & Eisenhofer, 3) Labaton Sucharow LLP, and 4) Robbins Geller Rudman & Dowd LLP.
TOP_PLF_LN_VIEWS-ORTHOGONAL$_{i,t-1}$	Residuals from a model regressing TOP_PLF_LN_VIEWS$_{i,t-1}$ on indicator variables set to one for litigation filings; earnings warnings; and announcement of major restatements, internal control weaknesses, CEO turnover, CFO turnover, non-timely filings, or auditor changes; and a count of large daily negative market-adjusted returns (i.e., $<$ –10%). All variables are measured with respect to year t–1 to be consistent with plaintiff-lawyer views.

Table A.41, Continued

Variable	Definition
TOP_SECURITIES_LAW_FIRM$_{i,t}$	An indicator variable set to one if the firm CC'd a top securities lawyer when replying to the SEC in a comment letter inquiry during the period t–2 to t; zero otherwise. Top securities lawyers are identified as law firms listed as National Tier 1 under https://bestlawfirms.usnews.com/litigation-securities.
VOLUNTARY_8-Ks$_{i,t}$	Number of Form 8-Ks with Item Codes 2.02, 7.01, or 8.01 that the firm submitted on EDGAR during the fiscal year (see Bourveau et al. (2018), He and Plumlee (2020)). For sued firms, the number of filings are limited to the day prior to filing date. We scale by the number of days in the relevant period. We also include amended filings.
VOLUNTARY_8-Ks_ITEM2.02$_{i,t}$	Defined similar to VOLUNTARY_8-Ks$_{i,t}$, but only includes 8-Ks with Item Code 2.02.
VOLUNTARY_8-Ks_ITEM7.01$_{i,t}$	Defined similar to VOLUNTARY_8-Ks$_{i,t}$, but only includes 8-Ks with Item Code 7.01.
VOLUNTARY8-Ks_ITEM8.01$_{i,t}$	Defined similar to VOLUNTARY_8-Ks$_{i,t}$, but only includes 8-Ks with Item Code 8.01.