The Cost of Fraud Prediction Errors

The Cost of Fraud Prediction Errors

Messod D. Beneish
Indiana University Kelley School of Business
dbeneish@indiana.edu

Patrick Vorst
Maastricht University School of Business and Economics
p.vorst@maastrichtuniversity.nl

The Accounting Review, Forthcoming

Running Head: The Cost of Fraud Prediction Errors

We gratefully acknowledge the helpful comments and suggestions from the editor (Robert Knechel), three anonymous reviewers, Ken Merkley, Joe Schroeder, Lori Shefchik-Bhaskar, Flora Sun, Jaeyoon Yu (discussant), workshop participants at Maastricht University and from the 2019 PCAOB and 2021 EAA conference participants. We thank Colleen Honisberg for helpful comments and discussions on the litigation and settlement prediction models in Honisberg, Rajgopal, and Srinivasan (2020). We also thank Dan Amiram, Zahn Bozanic, and Ethan Rouen for giving us access to the code to compute the FSD measure in Amiram et al. (2015), and we thank Yang Bao and Bin Ke for their help in using the code to reproduce their models in Bao et al. (2020). Any remaining errors are our own.

The Cost of Fraud Prediction Errors

Abstract

We compare seven fraud prediction models with a cost-based measure that nets the benefits of correctly anticipating instances of fraud, against the costs borne by incorrectly flagging non-fraud firms. We find that even the best models trade off false to true positives at rates exceeding 100:1. Indeed, the high number of false positives makes all seven models considered too costly for auditors to implement, even in subsamples where misreporting is more likely. For investors, M-Score and, at higher cut-offs, the F-Score, are the only models providing a net benefit. For regulators, several models are economically viable as false positive costs are limited by the number of investigations regulators can initiate, and by the relatively low market value loss a "falsely accused" firm would bear in denials of requests under the Freedom of Information Act (FOIA). Our results are similar whether we consider fraud or two alternative restatement samples.

Keywords: financial statement fraud, restatements, false positive, false negative, cost of errors, true positive benefits

JEL Classifications: G31; G32; G34; M40

Data Availability: Data are available from the public sources cited in the text

I. INTRODUCTION

Costs have long been central to accounting, auditing, and financial economics research as economic agents trade off costs and benefits in making decisions (e.g., Stigler 1971; Alchian and Demsetz 1972; Jensen and Meckling 1976). Yet, despite their important role, empirical documentation of the magnitude of many costs has been scarce. This scarcity is more acute in the context of models predicting financial statement fraud as costs vary not only by decision-maker, but also by type of error as depicted in Figure 1. Indeed, the costs of incorrectly classifying firms as frauds (e.g., false positives) have not been investigated, and thus not contrasted with the benefit of detecting frauds (e.g., true positives) to assess the net effect of using fraud prediction models.

Our investigation of the costs of fraud prediction errors is important for four reasons. First, comparing models by assessing their net costs or benefits is important because thus far, estimates of the rate at which decision makers trade off the costs of false positives and false negatives (e.g., Type I and II errors) have remained in the domain of assumptions. Second, the most commonly used comparison metrics in the literature—the area under the receiver operating characteristic curve (AUC) and expected costs of misclassification (ECM)—either overestimate model performance in studies of rare events and/or make unrealistic assumptions about the relative costs of prediction errors. Third, anecdotal discussions with auditors reveal that litigation concerns create a reluctance to use fraud prediction models in practice. Hence, our findings that increasing a model's rate of success at detecting fraud comes at the cost of an increasing number of false positives helps explain this tendency as false positives increase expected litigation costs. Fourth, recognizing that evaluating model prediction results requires human investigation, we analyze model performance in extreme subsamples that *a priori* have a greater likelihood of misreporting to provide insights into the conditions under which decision makers would find fraud prediction models to be most cost effective.

As we detail in Figure 2, we consider separately decisions by auditors, investors, and regulators. To auditors, benefits accrue by avoiding the costs of false negatives which we argue consist of litigation costs, reputation costs, and the foregone profits associated with any unusual client losses following the public revelation that accounting fraud went undetected at one of the auditors' clients. False positive costs to auditors consist of any incremental audit investment (if auditors cannot bill the client for it), lost audit fees if auditors resign, and an incremental litigation cost that we label 'discovery' cost.[1] To investors, benefits accrue by avoiding the abnormal value loss that occurs when the fraud is revealed, and false positive costs are the sum of any incremental audit fees paid by the client, and the profit foregone (or loss avoided) by not investing in false positives. To regulators, benefits accrue by avoiding the average value loss for the market as a whole across days when frauds are publicly revealed. In terms of regulators' false positive costs, resource constraints limit the number of false positives by restricting the number of new investigations regulators can initiate in any given time period, and the cost of falsely accusing a firm is also low, as it is more likely to occur indirectly via a FOIA request denial rather than by a public announcement that the firm is under investigation.

Our tests are based on three samples of non-financial firms: (1) firms that were charged by the SEC with accounting violations other than violations of the Foreign Corrupt Practices Act (hereafter AAER or fraud firms), (2) firms that have restated their financial statements, or (3) firms that have a severe restatement, defined as restatements disclosed in press releases or 8-Ks (see e.g.,

[1] The auditor costs we consider are similar to those explicitly recognized but not measured in prior work. For example, in a study comparing the detective performance of alternative analytical procedures using simulated account data Knechel (1986, 385) suggests that "auditors are concerned primarily with expected costs. A Type I error almost always results in overauditing and increased costs. A Type II error, on the other hand, may be cost-free in most cases, even though it may carry a potentially large cost under specific circumstances (e.g., a lawsuit or loss of a client)." Similar to Knechel (1986) who argues that an auditor's analytical procedure choice depends on the auditor's risk preference, we view expected litigation costs, which we label as 'discovery costs,' as the choice of auditing firms' general counsels who are by definition risk averse.

Huang and Scholz 2012). There are 768 fraud firm-years (313 unique fraud firms), 5,408 restatement firm-years (2,391 unique restatements) and 2,869 firm-year observations corresponding to 1,115 unique severe restatements.

We compare model performance both with traditional metrics and with our costs estimates to determine the net benefit or cost to a decision maker implementing the model.[2] In terms of traditional measures, the five models with higher AUC are the models in Alawadhi et al. (2020), Amiram et al. (2015), Chakrabarty et al. (2020), Cecchini et al. (2010) and Dechow et al. (2011). These models have the highest true positive (or sensitivity) rates that range from 57.9 percent to 82.0 percent, but also the highest false positive rates (ranging from 32 percent to 58.8 percent). Indeed, the models with the lowest sensitivity rates (Beneish 1999 and Bao et al. 2020) also have the lowest false positive rates with 17 percent for the Beneish (1999) model and 17.1 percent for the Bao et al. (2020) model. Except for Bao et al. (2020), who recognize the importance of reducing the number of false positives, fraud prediction studies in the last two decades have increased their true positive rates at the cost of higher false positive rates. The resulting *number* of false positives has become even larger given the low proportion of misreporting firms in the population, suggesting these models trade off increasing numbers of false positives for the benefit of a true positive. This in part motivates our comparison of models based on net benefits.

[2] Our comparison focuses on seven models that combine analyses of financial statement numbers and/or text with statistical and/or machine-learning algorithms to achieve increasing true positive rates and which we describe in Appendix A: the Beneish M-Score model (1997, 1999), the financial kernel model from Cecchini, Aytug, Koehler, and Pathak (2010), the F-Score (Dechow, Ge, Larson, and Sloan 2011), the measure of financial statement divergence based on how the distribution of first digits differs from Benford's Law (Amiram, Bozanic, and Rouen 2015), the misrepresentation model in Alawadhi, Karpoff, Koski, and Martin (2020), the machine-learning model in Bao, Ke, Li, Yu, and Zhang (2020), and the modification of Amiram et al. (2015)'s measure proposed by Chakrabarty, Moulton, Pugachev, and Wang (2020). We do not compare models that rely on textual or tonal analyses (the 'bag-of-words' analysis of Purda and Skillicorn (2015), the Brown, Crowley, and Elliott (2020) model that combines financial data with textual information), the model in Hribar, Kravet, and Wilson (2014) due to a limited time period of available data or, absent a specific decision rule, models based on raw, abnormal, or performance-matched abnormal accruals (Kothari, Leone ,and Wasley. 2005). Our comparison also differs from those in Brazel, Jones and Zimbelman (2009) and Price, Sharp, and Wood (2011) in that their analyses do not compare models on the basis on net costs or benefits.

Our investigation of net benefits reveals the following insights. First, for auditors, there are only two economically viable models/decision rules: (i) investigating firms with F-Scores greater than 2.45 provides net benefits across various samples, despite the low sensitivity rate at this cut-off (18 percent), and (ii) investigating restatement firms with M-Scores greater than -1.78. Second, for investors, only the M-Score and, at higher cut-offs, the F-Score provide a net benefit. This result obtains across different measures of abnormal returns, irrespective of whether we use raw or winsorized cost and benefit data. As these models and/or cutoffs also have the lowest false positive rates, we interpret this result as consistent with the notion that a large number of false positives makes model usage costly. Third, we take into account that regulator resources are constrained and consider scenarios under which there is capacity to conduct between 50 and 100 new investigations per year. We find that the models in Beneish (1999), Dechow et al. (2011), Amiram et al. (2015) and Chakrabarty et al. (2020) are economically viable at low costs for frauds and severe restatements, and for a wider range of costs in the samples of restatements.[3] Fourth, we assess the usefulness of prediction models within classes of firms that ex-ante have the potential for higher misreporting risk. For investors, we find that models are generally economically viable for young growth firms, firms with low operating cash flows, with large investing cash outflows, and with extreme accruals. In contrast, for auditors, the models only appear cost-effective among firms that are more likely over-investing (firms in the lowest quintile of investing cash flows).

Some perspective on our results is in order. Our estimation of net costs or benefits assumes that the costs we estimate represent the costs that would be borne by the various decision makers if they chose to implement one of the fraud screening models considered. Although we rely on

[3] We consider costs as low as 1 percent following recent studies that exploit data obtained via FOIA requests and which report that denials of FOIA requests lower future returns (Blackburne, Kepler, Quinn, and Taylor 2021, Coleman Merkley, Miller, and Pacelli 2021). We use 1 percent of the market value of equity as a representation of the likely cost of indirectly informing the public that the SEC is investigating a false positive.

previous research for much of our estimation of costs, consider alternative constructs, and conduct sensitivity tests, we cannot rule out that our costs estimates are incomplete, nor assess the extent to which they measure decision makers' costs with error.

Notwithstanding these cautions, our evidence on the number and costs of false positives has implications for professionals and researchers which we discuss in increasing order of implementation difficulty. For investors, the M-Score and, at higher cut-offs, the F-Score are economically viable, as are generally most models when applied to extreme quintiles of firm characteristics that capture the incentives and/or ability of firms to misreport. For the SEC, several models are economically viable because constrained resources limit the number of possible investigations and a limited market reaction to denials of FOIA requests indicates that the number and cost of false positives are relatively low. For auditors, the number and costs of false positives make fraud screening models too costly to implement in practice, even in extreme subsamples.

Researchers who rely on these models to identify misreporting in large samples should exercise caution because models with false positive rates in the range of 40 to 60 percent–rates many times larger than the actual incidence of misreporting in the population—could lead to falsely rejecting the null hypothesis of no misreporting more frequently than warranted. Moreover, as research continues to exploit statistical and machine-learning algorithms to analyze financial and/or textual data and propose new models on the basis of their potential usefulness to auditors, regulators, and investors, we believe it is important that future research carefully considers the user-specific potential costs and benefits when evaluating model performance. Indeed, assessing the models' economic viability is even more important because traditional metrics overstate the models' predictive performance as the misreporting/no-misreporting binary frequency tends to be highly unbalanced.

The remainder of the paper is structured as follows: Section II presents the samples and compares the seven fraud prediction models using traditional measures. Sections III, IV and V describe cost-benefit analyses for auditors, investors, and regulators, respectively. Section VI describes the economic viability of the models across types of firms and Section VII concludes.

II. TRADITIONAL MEASURES OF MODEL PERFORMANCE

Sample

Our primary evaluation of the performance of extant fraud classifiers is based on a comprehensive sample drawn from Accounting and Auditing Enforcement Actions (AAER) by the Securities and Exchange Commission (SEC) between April 1982 and July 2016 (AAER#1 to #3793). We identify accounting enforcement actions against 574 firms after eliminating multiple and unassigned AAERs (2351), those related to financial institutions (319), auditing actions against independent CPAs (280), enforcement actions for paying bribes under the Foreign Corrupt Practices Act (112), and those related to violations in 10-Qs resolved within the fiscal year (131). We complement our analyses of the sample of AAERs with two broader samples of accounting misconduct by using restatements drawn from Audit Analytics in the period 2000-2018.

Table 1 reports the selection of the final misconduct samples. The main sample consists of 574 fraud cases over the period 1979-2016, of which 494 cases can be matched to the COMPUSTAT-CRSP merged database. Those 494 cases relate to 1,185 firm-years with misstated financial statements. We then drop firms with missing returns around the revelation date of the fraud, for example because the firm delists prior to the date on which the fraud is revealed, leaving a sample of 413 fraud cases (1,041 firm-years). This criterion is imposed because the stock market reaction to the revelation of the fraud is an important measure of the costliness of the fraud from the perspective of investors and regulators. As we are interested in determining the usefulness of

fraud prediction models, we further drop observations for which we do not have data to compute the fraud prediction models that we compare. Overall, these sample restrictions lead to a final sample of 313 unique fraud cases involving 768 misstatement years. In terms or restatements, we exclude those cases that Audit Analytics codes as either resulting from clerical errors or that are non-adverse and subject the remainder of the restatements to the same exclusion criteria as applied to the fraud sample. This leaves us with 5,408 firm-year restatement observations corresponding to 2,391 unique restatements. We further investigate a subset of severe restatements (identified as restatements announced via a press release or 8-K Filing, see also Huang and Scholz 2012), which consists of 2,869 firm-year observations corresponding to 1,115 unique restatements.

In Table 1, Panel B we report the market reaction to the announcements of frauds and restatements. Not surprisingly, the mean abnormal return for fraud cases is the most adverse (-16.2 percent), followed by the response to the announcement of severe restatements (-2.8 percent), and to that of all restatements (-1.3 percent). Correspondingly, the market value losses are also more adverse averaging $446.63 million for frauds, and $47.92 ($18.49) million for severe (all) restatements.[4] In Panel C, we report the incidence of AAERs and restatements by fiscal year.

Traditional Measures of Performance

In Table 2, we report five metrics traditionally used to assess model performance: (1) the area under the Receiver Operating Characteristics curve (*AUC*), (2) model *sensitivity* or hit rate--a measure of success at identifying frauds or restatements– is computed as the ratio of true positives to all positives (e.g., the number of flagged frauds to all frauds in Panel A and B, and the number of flagged restatements to all restatements in Panel C and D), (3) model *precision*—a measure of

[4] Prior research has documented that the revelation of a restatement and/or fraud is typically associated with highly adverse abnormal price reactions. In terms of AAERs, Beneish (1999) and Karpoff et al. (2008) document three-day losses of -20 percent and -25 percent, respectively, while Palmrose and Scholtz (2004) document a market reaction to restatement announcements of approximately -4 percent.

the precision of the model's positive predictions—is the ratio of true positives to the sum of true and false positives (the number of flagged frauds or restatements relative to all flagged observations), (4) the *false positive rate,* which is the ratio of non-fraud (or non-restatement) firms that are flagged relative to all non-fraud (or non-restatement) observations, and (5) the *false to true positive rate*, which measures the average number of false flags for each true flag; assuming equal costs, it represents the rate at which a model trades of the cost of a false positive for the benefit of avoiding the cost of a false negative. Our calculations rely on published cut-offs with two exceptions: (1) for the Bao et al. (2020) model, we use the fraud flags generated by their machine-learning algorithm when applied to our sample firms; and (2) for the Cecchini et al. (2010) model, we estimate the cut-off that minimizes expected misclassification costs on our entire sample assuming a 200:1 cost ratio, which is the cost ratio that produces the best classification percentages in their original article.

The first two panels in Table 2 differ in that Panel A uses all fraud firm-years to compare models while Panel B uses only the first year in which a fraud occurs. Treating firms that commit fraud in consecutive years as a unique instance of fraud is consistent with the costs of fraud being imposed just once after the fraud is publicly revealed. Given that, on average, a firm has fraudulent reports for two consecutive years, focusing on the first year is consistent with the notion that frauds so detected can lead to cost avoidance for all decision makers. That is, while investors would benefit from a fraud detection flag any time before the fraud is publicly revealed, a flag of a firm in its second or in a latter fraud year is unlikely to reduce the costs borne by auditors or regulators.

In Panel A, in terms of AUC, the top four models are Alawadhi et al. (2020) (0.728), Chakrabarty et al. (2020) (0.689), Amiram et al. (2015) (0.679), and Dechow et al. (2011) (0.673).[5]

[5] The number of observations differs across models as a result of data requirements. While our sample criteria require data availability to compute the M-Score and F-Score, models by Alawadhi et al. (2020), Bao et al. (2020) and

Three of these models also have the highest sensitivity or true positive rates. For example, the Chakrabarty et al. (2020) model successfully identifies 82.0 percent of the frauds, and the Amiram et al. (2015) and Dechow et al. (2011) models do so with sensitivity rates of 68.4 percent and 64.7 percent, respectively. The Cecchini et al. (2010) model is the fourth highest with a sensitivity rate of 57.9 percent. The Bao et al. (2020) and the Alawadhi et al. (2020) models have the largest precision at 1.44 percent and 1.14 percent, respectively. Finally, the models with the two lowest false positives rates are the Beneish (1999) model (17.0 percent) and the Bao et al. (2020) model (17.1 percent). In general, we find that the models that have the highest true positive rates also have the highest false positive rates.

The results in Panel B for unique fraud observations are generally similar. For example, the AUC results reveal that the same four models have the highest AUC measures in the analysis of unique fraud cases, however, the order is different: the highest AUC in Panel B is for Dechow et al. (2011) (0.717), closely followed by Chakrabarty et al. (2020) (0.713). In addition, the three models that have the highest true positive rates also have the highest false positive rates, with 82.1 percent [58.8 percent] for Chakrabarty et al. (2020), 71.7 percent [42.8 percent] for Amiram et al. (2015), and. 71.6 percent [39.4 percent] for Dechow et al. (2011), respectively. The false to true positive ratio ranges from a low of 168 for the Bao et al. (2020) model to a high of 324 for the Cecchini et al. (2010) model. As this is the rate at which each model trades off true positives for false positives, this result implies that the benefit of identifying a true positive (measured as avoiding the cost of a false negative or missed detection) needs to be, on average, 168 to 324 times greater than the average cost of a false positive, for the model to be economically useful.

Cecchini et al. (2010) have more data restrictions. For example, to classify a firm as a potential fraud using Alawadhi et al. (2020), requires assessing the level of their model estimated probability in three consecutive years. Similarly, the Bao et al. (2020) model is estimated using a method that tolerates no missing observations in any of the 28 inputs that their model considers.

The relation between false and true positives in the fraud sample is such that false positives increase at an increasing rate as depicted in Figure 3 such that a one percent increase in the true positive rate comes at a greater percentage increase in the rate of false positives. This effect is magnified by the low proportion of fraud firms in the population and suggests that true positives come at the cost of a greater number of false positives.[6]

In Panel C we report the corresponding statistics for 2,391 unique restatements. In line with the fraud results, the three models that have the higher true positive rates also have the higher false positive rates (Chakrabarty et al. 2020: 63.0 percent [52.9 percent]; Cecchini et al. 2010: 54.1 percent [51.8 percent]; and Amiram et al. 2015: 49.2 percent [42.3 percent]). The Beneish (1999) and the Bao et al. (2020) models continue to have the lowest false positive rates at 12.1 percent and 17.0 percent, respectively. Compared to the fraud sample, we find that precision percentages are higher in the sample of restatements, ranging from 4.82 percent for the Alawadhi et al. (2020) model to 5.68 percent for the Beneish (1999) model. However, once we consider that restatements occur in the sample at rate of 4.65 percent, the higher precision does not translate into improved performance in the restatement sample. The true to false positive trade-off rate ranges from 17 to 20 which is significantly lower than that implied by the fraud sample analysis in Panel B and suggests lower cost-benefit thresholds for applying the models to predicting restatements.

The results in Panel D for the 1,115 unique severe restatements are similar in terms of sensitivity (the top three models are the same) and precision, which ranges from 1.94 percent for the Alawadhi et al. (2020) model to 2.81 percent for the Beneish (1999) model. Although the

[6] In an unreported analysis we find that if the goal is to increase the rate at which frauds are identified, combining model pairs based on either the union or the intersection of their flags is not desirable relative to using the models individually. The former because it results in more costly tradeoffs and the latter because lowering the false positive rate comes at the cost of lowering the true positive rate. The analysis does confirm that more true positives come at the cost of more false positives.

precision is higher than in the fraud sample, once we consider that severe restatements occur in the sample at a rate of 2.1 percent, the higher precision also does not translate into improved performance in the severe restatement sample. The true to false positive trade-off rate ranges from 35 to 51, which is lower than the rate for frauds, but higher than that of all restatements.

In sum, while our analysis of traditional performance measures enables us to relate our work to prior studies, it also highlights the sharp contrast between the implicit assumptions of equal costs of prediction errors and the rates at which the models trade off true and false positives. In addition, the lower cost-benefit trade-off when fraud prediction models are applied to samples of restatements, warrants investigation as to the models' economic usefulness in these samples. This underscores the need to consider detailed model-specific cost-benefit analyses.

A Cost-Based Measure of Model Performance

While AUC and ECM are frequently used comparison metrics, both suffer from major limitations, some of which arise because the underlying cost assumptions are not descriptive. First, although the AUC measure is not threshold-dependent and has been widely used (e.g., Cecchini et al. 2010; Larker and Zakolyukina 2012; Perols Bowen, Zimmerman, and Samba 2017; Alawadhi et al. 2020; Bao et al 2020; Chackrabarty et al. 2020), it treats false positives and false negatives as equally costly and focuses on model sensitivity (flagged frauds relative to all frauds), rather than model precision (flagged frauds relative to all firms flagged). This leads to potentially misleading conclusions about the relative performance of models because false positives and false negatives are not equally costly, and because AUC measures can mask poor performance when analyzing imbalanced binary classes (Davis and Goadrich 2006; Sokolova 2006; Maratea, Alfredo, and Manzo, 2014; Ozene, Subtil, and Maucort-Boulch 2015). For example, medical studies

indicate that the AUC tends to overestimate model performance when disease prevalence is low (Davis and Goadrich 2006; Ozene et al. 2015).[7]

Second, although costs vary across firms within a given class, ECM assumes that all classification errors of a given type are equally costly. The market value loss at revelation statistics in Table 1, Panel B indicate that such assumption is not descriptive.[8] For these reasons, we turn to evaluating the costs and benefits of using each model for different decision makers.

III. THE COST OF CLASSIFICATION ERRORS FOR AUDITORS

We compute the net benefit (cost) of using a model as the difference between the benefit of detecting an instance of misreporting (which we estimate as avoiding the costs of a missed detection [e.g., the costs of a false negative]), and the costs of false positives flagged by the model. We posit that auditors benefit from detecting an instance of misreporting by avoiding three types of costs that are associated with missed detections: litigation costs, client losses, and reputation losses.[9] We subtract from these benefits two types of false positive costs; those that result from auditors' own decisions (incremental audit work, resignation) and those that depend on the actions

[7] Although ROC-AUC modifications exist that allow for differential false positive and false negative costs, analyses relying on ROC-AUC primarily treat false positives and false negatives as equally costly (e.g., Adams and Hand 1999, Lobo, Jiménez-Valverde, and Real 2008). To our knowledge, they have only been implemented once in accounting research. Specifically, Alawadhi et al. (2020) conduct a sensitivity test that considers a relative cost ratio of 1.5 to 1, which does not alter their findings. However, this ratio is not sufficiently large to capture relative error costs across error types.

[8] Further, Beneish, Lee, and Nichols (2013) show that even among the top instances of fraud, the loss on the market - a component of investors' costs of missed detection (false negative) –varies considerably as it is 15 to 20 times larger for frauds at Enron and Waste Management compared to frauds at Sunbeam and Vivendi.

[9] We draw from prior work to identify the benefits of successfully detecting misreporting as the absolute value of the costs of missed detections. For litigation, we rely on prior research that has long recognized the importance of the costs of potential litigation for audit pricing and the design of financial reporting (e.g., Simunic 1980, Simunic and Stein 1996, Palmrose 1998, Lys and Watts 1994). For client losses, we follow Lyon and Maher (2005) and measure auditor client losses as the difference between an auditor's rate of attrition in the two years following the discovery of a missed detection and the rate of attrition for the same 'undetecting' auditor in the two years prior to the public revelation of the fraud. For reputation losses, we follow Franz, Crawford, and Johnson (1998), Gleason, Jenkins and Johnson (2008), and Weber, Willenborg, and Zhang (2008) who provide evidence that announcements that reveal either litigation against auditors or the existence of financial statement fraud at a firm are associated with negative spillover effects to the other clients of the auditor.

from the plaintiff's bar. While the former relates to lost audit fees and higher audit investment costs, the latter relates to the expected costs of litigation. We describe these below before presenting an analysis of the net costs or benefits by model.

Auditor False Negative Costs

We estimate the litigation component of auditors' false negative costs using actual damages paid by auditors in lawsuits filed against the auditor. Our evidence suggests that the incidence of lawsuits increases with the severity of the misreporting: We find that the auditors of 87 fraud firms were sued (27.8 percent of the sample) and that the incidence of lawsuits against auditors among firms with severe restatements is 2.6 percent compared to 1.55 percent for all restatements. The aggregate settlement is $3.128 billion for the fraud firms, $58.6 million for all restatements, and $55.8 million for severe restatements. On average, fraud firms that are sued pay $36 million to settle (e.g., $3,128.5/87), whereas restatement firms pay $1.58 million, and severe restatement firms pay $1.92 million. Our findings on aggregate are in line with results in Honigsberg, Rajgopal, and Srinivasan (2020) who identify 540 lawsuits naming auditors over the period 1996-2016 and estimate the aggregate value of auditor settlement payments at $3.53 billion.

We estimate the cost of client losses by multiplying the abnormal number of clients lost by the average audit fee per lost client. Relying on COMPUSTAT auditor data, we find that the average loss due to client losses is equal to $20.1 million for fraud firms, $16.3 million for all restatements and $18.2 million for severe restatements.

We rely on the contagion effect to measure auditor reputational losses.[10] Specifically, we use the average price reaction in the three days surrounding the fraud-revelation announcement of

[10] For example, Weber et al. (2008) find that other clients of KPMG Germany suffered abnormal returns of -3 percent around the three events revealing the accounting scandal at another KPMG auditee in Germany (ComROAD AG). Similarly, Gleason et al. (2008) document a contagion effect and suggest that investors re-assess their reliance on the financial statements of non-restating firms in the same industry.

the other clients of the auditor in the same industry as the fraud firm (6-digit GICS) relative to that of the other firms in the same industry, but which are not clients of the auditor. We find that the average reputation loss is equal to $116.54 million in 2016 dollars for fraud firms, but for restatements, perhaps because they represent less severe cases of misreporting, we do not find evidence of reputational losses.[11]

Auditors' Costs of False Positives

Incremental Audit Work

To the extent that auditors view a flag from a fraud screening model as an indication of fraud risk, they have two possible courses of action. First, auditors can increase the nature and extent of their audit investment to ascertain and report on whether fraud has indeed occurred. Second, they can resign from the client and forego their typical profit margin on the lost audit fee. Thus, for any given model, we can estimate auditor *i*'s expected false positive costs (E_{AUD} [FP COST]) for a given client *j* that is falsely flagged in year *t* as:

$$E_{AUDit}[\text{FP COST}] = p\,(AUDFEE_{ijt} * pm\%) + (1\text{-}p)\,(INCRAUDFEE_{ijt} * (1\text{-}pm\%)) \quad (1)$$

where p is the probability of resignation, and resignation results in the auditor losing its usual profit margin (pm%) on the audit fee billed to client *j* in year *t* ($AUDFEE_{ijt}$). In the absence of resignation, the auditor undertakes additional work costing $INCRAUDFEE_{ijt} * (1\text{-}pm\%)$, which we assume alternatively that (i) they cannot pass on to the client in year *t*, or (ii) that they can bill the client for. Equation (1) does not take into account the potential negative effects of time pressure on the quality of the auditor's output (e.g., Bills, Swanquist, and Whited 2016). Time pressure could increase because of the additional audit investment and auditors' limited ability to increase hiring

[11]The data we report are winsorized at the 1 percent-99 percent level, so these findings are less likely to reflect the effect of an extreme observation. In our analysis of net costs and benefits by model, we assume that reputation losses are zero when we find that, on average, industry-peers with a different auditor have a more severe drop in price than the industry-peer clients of the auditor. This avoids treating reputation losses as benefits.

and training in the short run. We compute incremental audit fee estimates using audit fees drawn from Audit Analytics for the period 2000-2016 and rely on a model based on Hribar et al. (2014)-which we describe in Table A.1 of the Online Appendix--to backfill audit fees for observations in the period 1980-1999. Drawing on both prior audit judgment research and on empirical audit fee research, we estimate that if auditors increase their investment in the audit after observing a flag from a fraud screening model, either they or investors bear a cost that ranges from 20 percent to 30 percent of lagged audit fees.[12]

In Table 3, Panel B we present statistics on the incremental audit fees we estimate as part of the expected false positive costs faced by auditors. We assume that the likelihood of resignation equals 3 percent [range 1-3 percent], that incremental fees amount to 23 percent [range 20-28 percent] of the contemporaneous audit fee, and, relying on recent reports by the BIG4 in the United Kingdom, that audit firms' profit margins equal 23 percent [range 20-30 percent]. In terms of resignation, we follow Landsman, Nelson, and Rountree (2009) who document a resignation rate of 2.93 percent and also consider a 1 percent resignation rate given that Bronson, Masli and Schroeder (2021) document a likelihood of resignation of 1.03 percent. The margin percentage is consistent with evidence provided by Chen, Elemes, Hope, and Yoon (2020) in their examination of audit firm profitability in the context of critical audit matters. In 2016 dollars, the average

[12] For example, Glover, Schultz, and Zimbelman (2003) suggest that auditors budget 21.8 percent more audit hours in high fraud risk engagements after the issuance of SAS 82. More recently, experimental evidence in Boritz, Kochetova-Kozloski, and. Robinson (2015) indicates that auditors increase audit budgets by 20.4 percent in the presence of fraud risks. The upper bound of the range is based on an analysis that assumes that an internal control weakness (ICW) is a fraud risk factor that leads auditors to increase their audit investment. The analysis which we describe in Online Appendix Table A.2, both follows and extends prior empirical work that has studied the effect of ICWs on audit fees (e.g., Raghunandan and Rama 2006; Hogan and Wilkins 2008; Munsif, Raghunandan, Rama, and Singhvi 2011). Consequently, we examine the behavior of audit fees before and after the revelation of an ICW, and also conditional on whether an ICW was subsequently remediated or not.

(median) fee-related false positive cost to auditors is $0.27 ($0.09) million for fraud firms, $0.50 ($0.20) million for all restatements and $0.50 ($0.21) million for severe restatements.[13]

Discovery Costs

Audit research has long recognized that auditors are concerned about expected costs and that litigation plays an important role in auditors' decisions (e.g., Simunic 1980; Knechel 1986). We expect that the presence of a fraud flag in a prior year's audit working papers for a continuing client reduces the plaintiff's bar discovery costs and makes it easier for the plaintiff to argue *scienter* even for an unrelated fraud occurring in a subsequent period. Discussions with auditors reveal that this has long been a concern of auditors' general counsel who have advised against implementing prediction models in audit practice, claiming that plaintiffs' attorneys could then argue on discovery that auditors had advance knowledge of the fraud and were negligent in conducting their audit, even if the auditor undertook incremental audit work to re-assess the likelihood of misreporting. This risk likely seemed significant as Arthur Andersen undertook to shred paper documents and destroy emails related to fraud alerts on Enron, which led to its indictment for obstruction of justice and its subsequent demise.[14]

Evidence that the plaintiff's bar exploits the circumstances of the audit engagement in pursuing litigation against auditors is also provided by Schmidt (2012) who shows that the plaintiff's bar exploits the extent to which auditors provided non-audit services in litigation against

[13] The estimates of fee-related costs for the restatement samples are higher than those for the fraud sample. This is likely the case because the vast majority of our sample restatements take place after the enactment of the Sarbanes Oxley Act (SOX), and prior research has shown a significant increase in average audit fees following SOX (Ettredge, Scholz, and Li 2007; Ghosh and Pawlewicz 2009).

[14] On January 25, 2002, the Wall Street Journal reported that in recovering deleted e-mails at Arthur Andersen, Congress found evidence that the Chicago office of Arthur Andersen had issued two "alerts" to the Houston office in the Spring of 2001 concerning earnings manipulation at Enron. The alerts came from a tailored version of the fraud prediction model that one of the authors had estimated under a consulting relationship with Andersen. ("Andersen Knew of `Fraud' Risk at Enron --- October E-Mail Shows Firm Anticipated Problems Before Company's Fall", 01/25/2002, A3).

auditors in discovered fraud cases. Circumstantial evidence from AAERs and PCAOB inspection reports is also consistent with the issue of discovery costs. Specifically, in several cases the PCAOB inspection reports allege auditors' negligence by not "considering whether a change in auditing procedures is needed to obtain more reliable evidence because of a higher risk of fraud" or to appropriately "evaluate the risk of fraud in the financial statements" (Pearson 2011). Clearly, an ex-post indication that the auditor did not act upon a flag from a fraud prediction model exposes the auditor to severe litigation risk, which may be low on an individual case basis, but will increase when accumulated across all of the auditors' clients. For example, in PCAOB Rel. No. 105- 2007-002, 7, the PCAOB revokes auditors' registrations on the basis of failure to "exercise due professional care, to exercise professional skepticism, and to obtain sufficient competent evidence, including responding appropriately to indicators of a risk of material misstatement due to fraud."

We label these false positive costs as 'discovery costs' to the extent they facilitate the plaintiff's bar discovery process and reduce the likelihood that judges dismiss the lawsuit for failure to plead *scienter*. We estimate them as the product of the estimated probability of litigation against an auditor and the estimated settlement amount relying on models motivated by Honigsberg et al. (2020). We describe the models in Table 3 where we report that in the sample of frauds, average (median) discovery costs equal \$0.93 (\$0.230) million in 2016 dollars for each false positive. The corresponding statistics are \$0.64 (\$0.01) million for all restatements and \$0.62 (\$0.01) million for severe restatements. As the discovery costs incorporate the probability of being sued, our estimates are plausibly lower than the settlement costs paid by auditors when they are actually sued. We treat these as estimates of the expected cost of facilitating discovery.

Net Benefits (Costs) to Auditors

In Table 4, we report estimates of the net costs or benefits to auditors using each model on fraud firms (Panel A), all restatements (Panel B), and severe restatements (Panel C). We compute net benefits by considering two specifications of false positive costs. In one specification, auditor's false positive costs include discovery costs and the cost of the extra audit investment, and in a second specification, we exclude the extra audit investment from false positive costs as auditors may be able to bill all or part of the incremental audit work to the client. The results are similar in both calculations of net benefits and suggest that there are only two economically viable models/decision rules for auditors. First, auditors investigating firms with F-Scores greater than 2.45 would gain net benefits across all three samples. Second, investigating restatement firms with M-Scores greater than -1.78 also provides net benefits to auditors. All other models and/or decision rules provide results that indicate that these models are too costly for auditors to use. Note that both models that provide benefits to auditors (M-Score > -1.78 and F-Score > 2.45) have relatively low true positive rates (36 percent for the M-Score and 18 percent for F-Score > 2.45, respectively), yet these models also present the lowest false positive rates (17 percent for the M-Score and 4 percent for F-Score > 2.45). We interpret this finding as consistent with the notion that a large number of false positives makes model usage costly to auditors.[15]

A caution is in order because our analysis relies on publicly available data whereas auditors have access to private information about the firms' strategy, control systems and management incentives and also to more up-to-date information (e.g., monthly data), which have been shown

[15] In the online Appendix Table A.3, we also consider an alternative computation of true positive benefits with similar results. Specifically, we exclude reputation losses from the true positive benefits as reputation and client losses are related as the likely outcome of a loss of auditor reputation is that the auditor loses clients. The benefits and costs reported in Table 4 and in Table A.3 are winsorized at the 1 percent-99 percent level and based on market-adjusted abnormal returns. We obtain similar results when the benefits and costs are not winsorized and when calculations are based on size-adjusted abnormal returns (available on request).

to improve the effectiveness of the audit and to more accurately assess the risk of material misstatement (Knechel 1988, 2007). As a result, we are not able to assess whether the fraud prediction models are economically viable in conjunction with non-public information that auditors can also access to assess misreporting risk.

IV. THE COST OF CLASSIFICATION ERRORS FOR INVESTORS

We posit that investors' benefit from detecting an instance of misreporting by avoiding the investment loss associated with the discovery of the misreporting (i.e., the stock price drop associated with the misreporting revelation). Investors' false positive costs are the sum of two components: (1) the incremental audit fee which we assume is billed by the incumbent auditor or required by the newly hired auditor, and (2) the profit foregone (or minus the loss avoided) by not investing over the next year in firms flagged by the prediction model as misreporting firms. In Table 5, we report these false positive costs in aggregate, and do not distinguish these two components as the majority of the net costs or benefits arise from the investor profit foregone/loss avoided. The inferences we discuss below are unchanged if we exclude the incremental audit fee from investors' false positive costs.

In Table 5, we report the net benefits or costs to investors using each model on fraud firms (Panel A), all restatements (Panel B), and severe restatements (Panel C). We compute two estimates of costs and benefits, one using actual data and the other using winsorized data. Although uncommon when dealing with portfolios, we report results after having winsorized extreme returns to allay concerns that our results are driven by either extreme observations or by mismeasured abnormal returns as one-year-ahead abnormal returns are a key component of investors' false positive costs. For example, actual and winsorized market-adjusted (size-adjusted) abnormal returns average 4.17 percent and 1.90 percent (2.04 percent and -.04 percent) for all non-fraud

firms.[16] Table 5 estimates of net benefits are based on market-adjusted returns, and as we discuss and report in the online Appendix (Table A.4), we find similar results using size-adjusted returns.

Panel A reports that there is a net benefit to investors using either M-Score or F-Score (at cut-offs of 1.85 and 2.45), suggesting that these models are economical in detecting fraud firms in the AAER sample. All the other models generate net costs to investors ranging from $138 billion for F-Score (at a cut-off of 1.00) to $2,001 billion for the Chakrabarty et al. model. Interestingly, it is not the benefits from detecting frauds (i.e., the true positives) that are driving these results. To explain, the M-Score has a lower hit rate (sensitivity) than the other models and although it identifies firms that are among the highest in percentage market value loss at revelation (mean of 21.85 percent), the mean dollar amount of benefit is among the lowest at $140.87.[17] Indeed, the principal source of benefits is that *by not investing* in the false positives generated from the M-Score or, at cut-offs of 1.85 and 2.45 for the F-Score, investors avoid, over the year that follows, percentage and dollar losses ranging from -4.0 percent (-$36.1 million) to -11.1 percent (-$62.3 million). Overall, the M-Score has the highest net benefit for investors which amounts to $1,588 billion or $847 billion winsorized.

Similarly, analyzing the sample of all restatements, we find that investors only benefit when using either the M-Score or the F-Score at higher cut-offs. However, unlike the AAER sample, the

[16] Although not always reported, it is not unusual for sample abnormal returns over one-year periods to be non-zero. For example, Blackburne et al. (2020, 7) examine over 81,000 firm-year observations over the period 2000-2017 and report market-adjusted one-year abnormal returns averaging 6.0 percent for the 73,197 firm-year observations without SEC investigations and 1 percent in 8,616 firms with SEC investigations. Similarly, Brown et al. (2020, 267), report market-adjusted one-year abnormal returns averaging 10.6 percent for 37,301 non-AAER firm-year observations in the period 1994-2010, and 14.8 percent for 37,571 non-irregularity firm-year observations in the period 2000-2012. The averages for the sample of non-restatement firms are as follows: actual and winsorized market-adjusted (size-adjusted) abnormal returns average 7.46 percent and 3.26 percent (5.30 percent and 1.40 percent); for non-severe-restatement firms averages are as follows: actual and winsorized market-adjusted (size-adjusted) abnormal returns average 7.38 percent and 5.24 percent (3.21 percent and 1.38 percent).

[17] This contrasts, for example, with the Alawadhi et al. (2020) model which has the lowest percentage loss (mean of 13.28 percent), but the highest mean benefit of $977 million ($901 million winsorized). The findings are consistent with the M-Score identifying more frequently smaller growth firms (Beneish 1999) and with the Alawadhi et al. (2020) model identifying larger firms, as their model contains market capitalization which loads positively.

F-Score at the 2.45 cut-off has the highest net benefit which amounts to $224 billion ($71 billion winsorized). Finally, analyzing the sample of severe restatements, we find that investors benefit when using the same models: the F-Score at the 2.45 cut-off has the highest net benefit ($232 billion) and the M-Score has the highest net benefit ($80 billion) in the winsorized sample calculation.

In sum, only the M-Score and, at higher cut-offs, the F-Score, provide a net benefit to investors, whereas the other models generate net costs, consistent with the latter models not systematically exploiting fundamental signals. To explain, the main component of the investors' false positive costs is the profit foregone by not investing in a falsely flagged firm. For both M-Score and F-Score (at high cut-offs) this is turned into a benefit because the firms that are falsely flagged by these models typically face significant headwind one-year-ahead in terms of both earnings and returns. This is not surprising since these models rely on signals about future firm prospects that are documented in the academic and professional literatures to predict future earnings and future returns (e.g., O'Glove 1987, Kellogg and Kellogg 1991, Siegel 1991, Lev and Thiagarajan 1993, Sloan 1996, Abarbanell and Bushee 1997, 1998). Indeed, in an extensive out-of-sample analysis, Beneish et al. (2013) show that the M-Score identifies firms that experience one-year-ahead negative abnormal returns. This is likely because the typical firm identified by the M-Score is a growth firm with "deteriorating fundamentals (as evidenced by a decline in asset quality, eroding profit margins, and increasing leverage)" and with "aggressive accounting practices (e.g., receivables growing much faster than sales, large income inflating accruals, and decreasing depreciation expense)." (Beneish et al. 2013, 57)

V. THE COST OF CLASSIFICATION ERRORS FOR REGULATORS

Regulators' Costs

An important objective of regulators (e.g., SEC, PCAOB) is to minimize wealth losses related to asymmetric and/or misleading information (e.g., incomplete or inaccurate reports and disclosures). We argue that regulators' costs of false negatives (e.g., the benefit when these costs are avoided as misreporting is detected) is the cost of losing investors' trust in the functioning of capital markets which we measure as the abnormal value drop in the market as a whole on days that reveal a failure in regulatory oversight. To the extent that investors update their beliefs about the quality of regulatory oversight and the presence of asymmetric information in the stock market, we expect that market-wide returns are lower on days on which frauds are revealed. We estimate this cost as the market-wide dollar abnormal returns in three periods centered on the fraud-revealing announcement (day 0, days -1 to +1, and days -2, +2). Specifically, we estimate a regression of CRSP daily value-weighted market returns on an indicator variable that is equal to one on the fraud revelation date and zero on days on which there is no fraud revealed. The results are reported in Table 6, Panel A. Whereas we find significant results only in the regressions with the one-day event window, the negative and significant coefficient on *EVENT* indicates that market-wide returns are on average slightly lower on fraud revelation dates (-0.119 percent). This result is consistent with investors being more pessimistic about stocks in general, perhaps as a result of a loss of trust in the operation of capital markets. The effect is lower if we consider three-day and five-day event windows (-0.033 percent and -0.032 percent). In terms of economic significance, these percentages yield estimates of regulators' false negative costs (and consequently benefits from detecting frauds) that range from $5.4 billion to $20.1 billion in 2016 dollars per instance of fraud.

Regulator false positives costs arise if it becomes known that they are investigating a firm incorrectly flagged by the model: in that case, the costs are the loss in market value of the firm for which regulators mistakenly investigated misreporting. Thus, estimating regulators' false positive costs requires two ingredients: an estimate of the number of regulators' misreporting investigations and an estimate of the market value loss for falsely flagged firms.

With respect to the number of investigations, even though the use of a prediction model can be common knowledge among market participants, investors understand that regulators have constrained resources. As such, there is a limit to the number of investigations regulators can undertake and the number of false positives for regulators is thus capped. Until recently, it was not possible to estimate the number of investigations undertaken by the SEC, but FOIA requests, a number of studies estimate that the number of misreporting investigations ranges from 50 to 100 per year (Blackburne et al., 2020; Bonsall, Holzman, and Miller 2021; Holzman, Marshall, and Schmidt 2021).[18] For this reason, when estimating regulators false positive costs, we limit the number of flagged firms investigated by the SEC to range from the top 50 to the top 100 ranks for each model in a given year.

We consider false positive costs as the loss in market value if regulators make public an investigation on a falsely flagged firm. Although we understand regulator aversion to publicly making false accusations (we are not aware of any), investors and financial intermediaries carefully monitor regulators' actions. We consider estimates of market value losses ranging from 1 percent to 15 percent. We treat 1 percent as a lower bound on account of evidence in Coleman

[18] These studies rely on closed formal SEC investigations data initially obtained by Blackburne et al. (2020) under FOIA for the period January 1, 2000 to August 2, 2017. Bonsall et al. (2021) document that the SEC opened investigations for approximately 13.5 percent of the 4,698 restatements in Audit Analytics in the period from January 1, 2000 to December 31, 2013, which averages to approximately 46 investigations per year. Holzman et al. (2021) estimate that over this period the SEC undertook 1,249 misreporting investigations out of a total of 1,685 investigations for publicly traded firms. These totals yield an average of 70 to 95 investigations per year.

et al. (2020) that SEC denials of FOIA requests due to ongoing enforcement proceedings are predictive of SEC investigations and are associated with negative future return performance. In this case, the cost of falsely accusing a firm is low, because it is more likely to occur indirectly via a FOIA request denial rather than by a public announcement that the firm is under investigation. However, if regulators make the investigations public, we assume that falsely identified firms experience a market value loss in the range of 3 percent to 15 percent, which is the typical market reaction to comment letters and announcements of SEC investigations and charges.

Net Benefits (Costs)

Table 6, Panel B reports the net benefits or costs for regulators under the seven models assuming 100 investigations per year and market value losses of 1, 3, and 5 percent for false positives.[19] Three features are noteworthy. First, three models are too costly to use regardless of the number of investigations or the magnitude of the loss. These are the Cecchini et al. (2010), the Bao et al. (2020), and the Alawadhi et al. (2020) models. This is likely because these models identify larger firms, thereby increasing the dollar loss associated with false positives. For example, the Alawadhi et al. (2020) model includes market capitalization with a positive weight as shown in Appendix A equation (6).[20] Second, if the cost of falsely accusing a firm is 1 percent or less of its market value, four models – Beneish (1999); Dechow et al. (2011); Amiram et al. (2015); Chakrabarty et al. (2020) - are systematically beneficial. Third, if the cost of falsely accusing a firm is 3 percent or greater, none of the models are beneficial.

[19] We also consider 50, 60, 70, 80 and 90 investigations per year. The results which we report in Table A.5 of the Online Appendix, are generally similar. We report 5 percent as an upper bound on losses in both Table 6 and Table A.5 because the models cease to be beneficial at that level independently of the number of annual investigations.
[20] We obtain similar results when evaluating the data after winsorizing at the 1 percent and 99 percent values of market capitalization (these results are available on request).

Table 6, Panel C and D report the corresponding results for restatements and the subset of severe restatements, respectively. In line with the results in Panel A, the Cecchini et al. (2010) and the Alawadhi et al. (2020) models are too costly to use regardless of the magnitude of the loss. On the other hand, the Bao et al. (2020) model is economically viable if the cost of falsely accusing a firm is 1 percent or less of its market value. In contrast to the results on frauds in Panel A, we find that the M-Score, F-Score, and the models in Amiram et al. (2015) and Chakarbarty et al. (2020) are economically viable for the identification and detection of restatements, for all loss percentages considered.

VII. ASSESSING ECONOMIC VIABILITY ACROSS TYPES OF FIRMS

In this section, we assess the usefulness of prediction models within classes of firms that have the potential for higher misreporting risk and/or misreporting costs either because they present characteristics that provide ex-ante greater incentives to misreport, or because weaker monitoring mechanisms are less likely to prevent misreporting. We do so for investors and auditors and consider quintile partitions of accruals, operating cash flows (CFO), investing cash flows (CFI), prior stock returns, sales growth, size, return volatility, firm life cycle, as well as dedicated and transient institutional ownership. While the analyses reported in this section are exploratory and many other variables could be considered as viable alternatives, this set of variables captures various broad constructs that are likely associated with misreporting, such as growth, performance, and oversight.

Table 7 (Table 8) reports, for each characteristic quintile, the average net benefit to investors (auditors) of using the fraud prediction models. We report four computations of net benefits for investors based either on market or size-adjusted returns and using either actual data or data

winsorized at the 1 percent and 99 percent levels. We average net benefits over the seven models to provide insights on the performance of fraud prediction models across various characteristics.[21] For investors we find that, on average, fraud prediction models are most useful for young growth firms that make large investments. These are firms in the introductory and growth stages of their life cycle, and they experience high sales growth (highest quintile) and more negative investing cash outflows (lowest quintile). Similarly, we find that net benefits are greater for firms that are riskier (firms with higher volatility and extreme accrual firms), and firms that have potentially inflated stock prices (higher past abnormal returns), poor current performance (low cash flows), and either weaker monitoring or higher short-term pressure (low dedicated institutional ownership or higher transient institutional ownership).

While the results in Table 7 indicate that the net benefits of fraud prediction models to investors vary considerably across these characteristics, the results in Table 8 suggest that using fraud prediction models represents a net cost to auditors across most of the values of these characteristics. If we assume that true positive benefits include avoidance of both client losses and reputation costs, results indicate that the models are useful to auditors in some instances. In particular, this is the case for firms that are small, have high sales growth (top quintile), large investments (lowest CFI quintile) and either low dedicated (or high transient) institutional ownership. However, when we exclude reputation cost avoidance from the calculation of net benefits, there is only a net benefit to auditors for firms in the lowest CFI quintile.

Finally, Table 9 reports fraud occurrence and traditional model performance measures across quintiles of the characteristics we consider. We find that the number of fraud cases is generally

[21] In line with our main results, cross-sectional results can vary considerably across models. However, for brevity, we only report the average performance of all models across the quintiles of the characteristics we investigate (results available on request).

higher in quintiles of high accruals, high growth, high investment, high momentum, low dedicated ownership, or high transient ownership. This is consistent with the notion that extreme values of these measures likely reflect higher misreporting risk. In addition, we find that it is in these extreme quintiles that the models on average exhibit higher precision—a summary measure of performance that combines model success and model error (number of true positives to flagged firms). These findings suggest that results across quintiles of characteristics are driven by differences in both fraud occurrence and in model performance.

Overall, when evaluating the models' usefulness to investors, we find considerable variation in the net benefits of fraud prediction models across firm characteristics. However, using these models generally appears costly for auditors, even when considering samples in which misreporting is ex-ante more likely and/or costly.

VIII. CONCLUSION

This paper estimates the costs of fraud prediction errors from the perspective of auditors, investors, and regulators, and proposes a cost-based measure for model comparison that nets the benefits of correctly anticipating instances of fraud (true positives), against the costs borne by incorrectly flagging non-fraud firms (false positives). We supplement traditional comparison measures with our cost-based measure for two reasons. First, traditional measures typically overestimate model performance in studies of rare events and/or assume cost equality, in contrast to evidence that costs vary both within and across classes of firms. Second, unlike traditional measures, cost-based measures can be estimated specifically for each of the decision makers we consider, thereby allowing the costs and benefits to vary depending on the type of decision at hand.

We compare seven fraud prediction models that have been proposed in prior research. We find that the higher true positive rates in recent models come at the cost of higher false positive

rates and that even the best models trade off false to true positives at rates exceeding 100:1. Indeed, the high number of false positives makes all seven models considered too costly for auditors to implement, even when we consider extreme subsamples where a priori firms' management has higher incentives and/or ability to misreport. We believe this could explain audit practitioners' apparent reluctance to use these models, despite the fact that models have nearly doubled their success at identifying fraud when compared to the initial models in Beneish (1997, 1999).

For investors, M-Score and the F-Score when used at higher cut-offs are the only models providing a net benefit when applied to the sample as a whole. We conjecture this occurs because the M-Score and the F-Score exploit fundamental signals that have been shown to predict future earnings and returns, and the main component of investors' false positive costs is the profit foregone (or the loss avoided) by not investing in a falsely flagged firm. In addition, we find that most models are economically viable if applied to top or bottom quintiles of characteristics of firms in which managers *a priori* have greater incentives and/or ability to misreport.

For the SEC, several models are economically viable because constrained resources limit the number investigations and the limited market reaction to denials of FOIA indicates that the costs of false positives are relatively low. For researchers, the use of these models to identify misreporting in large sample studies without the benefit of further examination is concerning, because the typical 40 to 60 percent false positive rate is much larger than the actual incidence of misreporting in the population, which would lead to falsely rejecting the null hypothesis of no misreporting more frequently than warranted.

Overall, as the number of false positives increases, the use of fraud prediction models becomes more costly. Hence, our evidence suggests that researchers focus on lowering the false positive rates of their models rather than pursuing higher true positive rates. Alternatively, they

could investigate ways to make the resolution of false positive investigations more efficient, either by isolating the main causes of the flag (e.g., growth, profitability in the M-Score) or by examining compensation, governance, investing and trading patterns as a way of corroborating or contradicting the model's predictions.

REFERENCES

Abarbanell, J., and B. Bushee. 1997. Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research* 35 (1): 1–24.

Abarbanell, J., and B. Bushee. 1998. Abnormal returns to a fundamental analysis strategy. *The Accounting Review* 73 (1):19-45.

Adams, N. M., and D. J. Hand. 1999. Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognition* 32 (7): 1139-1147.

Alawadhi, A., J. Karpoff, J. Koski, and J. Martin. 2020. The prevalence and costs of financial misrepresentation. Working paper. University of Washington.

Alchian, A. A., and H. Demsetz. 1972. Production, information costs, and economic organization. *The American Economic Review* 62 (5): 777-795.

Amiram, D., Z. Bozanic, and E. Rouen. 2015. Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies* 20 (4): 1540-1593.

Bao, Y., B. Ke, B. Li, Y. J. Yu, and J. Zhang. 2020. Detecting accounting fraud in publicly traded U.S. firms: A new perspective and a new method. *Journal of Accounting Research* 58 (1): 199-235.

Beneish, M. D. 1997. Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16 (3): 271-309.

Beneish, M. D. 1999. The detection of earnings manipulation. *Financial Analysts Journal* 55 (5): 24-36.

Beneish, M. D., C. M. Lee, and D. C. Nichols. 2013. Earnings manipulation and expected returns. *Financial Analysts Journal* 69 (2): 57-82.

Bills, K. L., Q. T. Swanquist, and R. L. Whited. 2016. Growing pains: Audit quality and office growth. *Contemporary Accounting Research* 33 (1): 288-313.

Blackburne T, J. Kepler, P. Quinn, and D. Taylor. 20210. Undisclosed SEC investigations. Management Science 67 (6), Article in Advance: 34031-341816.

Bonsall, S., E. Holzman, and B. P. Miller. 2021. Wearing out the watchdog: SEC case backlog and investigation likelihood. Working paper, Pennsylvania State University, State College.

Boritz, J. E., N. Kochetova-Kozloski, and L. Robinson. 2015. Are fraud specialists relatively more effective than auditors at modifying audit programs in the presence of fraud risk? *The Accounting Review* 90 (3): 881-915.

Brazel, J. F., K. L. Jones, and M. F. Zimbelman. 2009. Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research* 47 (5): 1135–1166.

Bronson, S. N., A. Masli, and J. H. Schroeder. 2021. Releasing Earnings when the audit is less cComplete: Implications for audit quality and the auditor/client relationship. *Accounting Horizons* 35 (2): 27-55.

Brown, N. C., R.M. Crowley, and W. B. Elliott. 2020. What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research* 58 (1): 237-291.

Cecchini, M., H. Aytug, G.J. Koehler, and P. Pathak. 2010. Detecting management fraud in public companies. *Management Science* 56 (7): 1146-1160.

Chakrabarty, B., P. C. Moulton, L. Pugachev, and F. Wang. 2020. Catch me if you can: Improving the scope and accuracy of fraud prediction. Working paper, St Louis University.

Chen, J.Z., A. Elemes, O. K. Hope, and A. Yoon. 2020. Determinants and consequences of audit-firm profitability: Evidence from key audit matters. Working paper University of Toronto.
Ciesielski, J. 1998. What's happening to the quality of assets? *Analyst's Accounting Observer,* 7 (3): 19.
Coleman, B., K. Merkley, B. Miller, and J. Pacelli. 2021. Does FOIA foil the SEC's intent to keep investigations confidential? *Management Science* 67 (6): 3419-3428.
Davis, J., and M. Goadrich. 2006. The relationship between precision recall and ROC Curves. Proceedings of the 23rd international conference on Machine learning June 2006: 233–240.
Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28 (1): 17-82.
Dickinson, V. 2011. Cash flow patterns as a proxy for firm life cycle. *The Accounting Review* 86 (6): 1969-1994.
Ettredge, M. L., S. Scholz, and C. Li. 2007. Audit fees and auditor dismissals in the Sarbanes-Oxley era. *Accounting Horizons* 21 (4): 371-386.
Franz, D. R., D. Crawford, and E. N. Johnson. 1998. The impact of litigation against an audit firm on the market value of nonlitigating clients. *Journal of Accounting, Auditing & Finance* 13 (2): 117-134.
Ghosh, A., and R. Pawlewicz. 2009. The impact of regulation on auditor fees: Evidence from the Sarbanes-Oxley Act. *Auditing: A Journal of Practice & Theory* 28 (2): 171-197.
Gladwell, M. 2009. *What the Dog Saw and Other Adventures*. New York: John Wiley & Sons.
Gleason, C. A., N. T. Jenkins, and W. B. Johnson. 2008. The contagion effects of accounting restatements. *The Accounting Review* 83 (1): 83-110.
Glover, D. P., J. Schultz, and M. Zimbelman. 2003. A comparison of audit planning decisions in response to increased fraud risk: Before and after SAS No. 82. *Auditing: A Journal of Practice & Theory* 22 (3): 237-251.
Hammersley, J. S., K. M. Johnstone, and K. Kadous. 2011. How do audit seniors respond to heightened fraud risk? *Auditing: A Journal of Practice & Theory* 30 (3): 81-101.
Harrington, C. 2005. Analysis of ratios for detecting financial statement fraud. *Fraud Magazine (March-April)*: 24-27.
Hogan, C. E., and M. S. Wilkins. 2008. Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies? *Contemporary Accounting Research* 25 (1): 219-242.
Holzman, E., N. Marshall, and B. Schmidt. 2021. Who's on the hot seat for an SEC investigation? Working Paper, The Ohio State University.
Honigsberg, C., S. Rajgopal, and S. Srinivasan. 2020. The changing landscape of auditor liability. *Journal of Law and Economics* 63 (2): 367-409.
Hribar, P., T. Kravet, and R. Wilson. 2014. A new measure of accounting quality. *Review of Accounting Studies* 19 (1): 506-538.
Huang, Y., S. Scholz. 2012. Evidence on the association between financial restatements and auditor resignations. *Accounting Horizons* 26 (3): 439-464.
Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305-360.
Karpoff, J. M., D. S. Lee, and G. S. Martin. 2008. The cost to firms of cooking the books. *Journal of Financial and Quantitative Analysis* 43 (3): 581-611.

Kellogg, I., and L.B. Kellogg. 1991. Fraud, Window Dressing, and Negligence in Financial Statements. New York: McGraw-Hill.
Knechel, W. R. 1986. A simulation study of the relative effectiveness of alternative analytical review procedures. *Decision Sciences* 17 (3): 376–394.
Knechel, W. R. 1988. The effectiveness of statistical analytical review as a substantive auditing Procedure: A simulation analysis. *The Accounting Review* 63 (1): 74-95.
Knechel, W. R. 2007. The business risk audit: Origins, obstacles and opportunities. *Accounting Organizations and Society* 32 (4-5): 383-408.
Kothari, S. P., A. J. Leone, and C. E. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39 (1): 163-197.
Landsman, W. R., K. K. Nelson, and B. R. Rountree. 2009. Auditor switches in the pre- and post-Enron eras: Risk or realignment? *The Accounting Review* 84 (2): 531-558.
Larcker, D. F., and A. A. Zakolyukina. 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50 (2): 495-540.
Lev, B., and S. R. Thiagarajan. 1993. Fundamental information analysis. *Journal of Accounting Research* 31 (2): 190-215.
Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17 (2): 145-151.
Lyon, J. D., and M. W. Maher. 2005. The importance of business risk in setting audit fees: Evidence from cases of client misconduct. *Journal of Accounting Research* 43 (1): 133-151.
Lys, T., and R. L. Watts. 1994. Lawsuits against auditors. *Journal of Accounting Research* 32 (Supplement): 65-93.
Maratea, A., P. Alfredo, and M. Manzo, 2014. Adjusted F-measure and kernel scaling for imbalanced data learning. *Information Sciences* 257: 331-341.
Merrill Lynch. (2000). Financial Reporting Shocks (March 31).
Munsif, V., K. Raghunandan, D. V. Rama, and M. Singhvi. 2011. Audit fees after remediation of internal control weaknesses. *Accounting Horizons* 25 (1): 87-105.
O'Glove, T.L. 1987. Quality of Earnings. New York: The Free Press.
Ozenne, B., F. Subtil, and D. Maucort-Boulch. 2015. The precision recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of Clinical Epidemiology* 68 (8): 855-859.
Palmrose, Z.-V. 1988. 1987 Competitive Manuscript Co-Winner: An analysis of auditor litigation and audit service quality. *The Accounting Review* 63 (1): 55-73.
Palmrose, Z.-V., and S. Scholz. 2004. Auditor independence, non-audit services, and restatements: Was the U.S. government right? Journal of Accounting Research 42 (3): 561-588.
Pearson, T. C. 2011. Potential litigation against auditors for negligence. *Brooklyn Journal of Corporate, Financial & Commercial Law* 5 (2): 4.
Perols, J. L., R. M. Bowen, C. Zimmermann, and B. Samba. 2017. Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review* 92 (2): 221-245.
Price III, R.A., N. Y. Sharp, and D. A. Wood. 2011. Detecting and predicting accounting irregularities: A comparison of commercial and academic risk measures. *Accounting Horizons* 25 (4): 755-780.

Purda, L. and D. Skillicorn, 2015. Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection. *Contemporary Accounting Research* 32 (3): 1193-1223.

Raghunandan, K., and D. V. Rama. 2006. SOX Section 404 material weakness disclosures and audit fees. *Auditing: A Journal of Practice & Theory* 25 (1): 99-114.

Schmidt, Jaime J. 2012. Perceived auditor independence and audit litigation: The role of nonaudit services fees. The Accounting Review 87(3): 1033-1065.

Siegel, J.G. 1991. How to Analyze Businesses, Financial Statements, and the Quality of Earnings. 2nd ed. Englewood Cliffs, NJ:Prentice-Hall.

Simunic, D. A. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research* 18 (1): 161-190.

Simunic, D. A., and M. T. Stein. 1996. Impact of litigation risk on audit pricing: A review of the economics and the evidence. *Auditing* 15: 119.

Sloan, R. G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289-315.

Sokolova, M. N. Japkowicz, and S. Szpakowicz, 2006. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *Australasian Joint Conference on Artificial Intelligence*, Springer Berlin Heidelberg.

Stigler, G. J. 1971. The theory of economic regulation. *The Bell Journal of Economics and Management Science 2* (1): 3-21.

Weber, J., M. Willenborg, and J. Zhang. 2008. Does auditor reputation matter? The case of KPMG Germany and ComROAD AG. *Journal of Accounting Research* 46 (4): 941-972.

Wells, J.T. 2001. Irrational ratios. *Journal of Accountancy* 192 (2): 80-83.

Appendix A - Fraud Prediction Models

We evaluate the ability of a wide range of fraud detection models to identify firms that are subsequently subject to SEC accounting and enforcement actions. Specifically, in chronological order, we use the M-Score (Beneish 1999), the Cecchini et al. (2010) model based on support vector machines (SVM), the F-Score (Dechow et al. 2011), an extended F-Score model that incorporates a measure of financial statement divergence based on how the distribution of first digits differs from Benford's Law (Amiram et al. 2015), the Adjusted Benford Score from Chakrabarty et al. (2020), the misrepresentation model from Alawadhi et al. (2020), and, finally, the Bao et al. (2020) fraud prediction model developed using a machine learning approach.

The Beneish M-Score

Beneish (1997; 1999) profiles firms that manipulate earnings (firms either charged by the SEC or that admit to manipulation in the public press) and develops a statistical model to discriminate manipulators from non-manipulators. In this paper, we use the unweighted probit model presented in Beneish (1999) which relies exclusively on financial statement data and whose usefulness in assessing fraud potential *out-of-sample* has been shown by academics and professionals.[22] The M-Score below classifies a firm as a potential manipulator if the M-Score exceeds -1.78, assuming that false negatives are 20 times more costly than false positives:

$$M\text{-}SCORE = -4.840 + 0.920DSRI_{it} + 0.528GMI_{it} + 0.404AQI_{it} + 0.892SGI_{it} + 0.115DEPI_{it} - 0.172SGAI_{it} + 4.679TATA_{it} - 0.327LGVI_{it} \quad (1)$$

Where:

DSRI = day's sales receivable index
= $(AR_t/REV_t)/(AR_{t-1}/REV_{t-1})$;

GMI = gross margin index
= $[(REV_{t-1} - Cost\ of\ Goods\ Sold_{t-1})/REV_{t-1}]/[(REV_t - Cost\ of\ Goods\ Sold_t)/REV_t]$;

AQI = asset quality index
= $(1 - [Current\ Assets_t + PPE_t]/AT_t)/(1 - [Current\ Assets_{t-1} + PPE_{t-1}]/AT_{t-1})$;

SGI = sales growth index
= REV_t/REV_{t-1};

DEPI = depreciation index
= $(Depreciation_{t-1}/[Depreciation_{t-1} + PPE_{t-1}])/(Depreciation_t/[Depreciation + PPE])$;

SGAI = sales, general, and administrative expenses index
= $(SGA\ Expense_t/REV_t)/(SGA_{t-1}/REV_{t-1})$;

TATA = total accruals to total assets
= $(IBC_t\text{-}CFO_t)/AT_t$; and

LGVI = leverage index
= $([Long\text{-}Term\ Debt_t + Cur.\ Liab_t]/AT_t)/([Long\text{-}Term\ Debt_{t-1} + Cur.Liab_{t-1}]/AT_{t-1})$.

[22] Since the publication of the original study (which used data from 1982 to February 1993), the model has been featured in financial statement analysis textbooks and in articles directed at auditors, certified fraud examiners, and investment professionals (e.g., Ciesielski 1998, Merrill Lynch 2000, Wells 2001, DKW 2003, Harrington 2005). The model gained notoriety when a group of MBA students at Cornell University posted the earliest warning about Enron's accounting manipulation score using the Beneish (1999) model a full year before the first professional analyst reports (Morris 2009). This episode in American financial history is preserved in the Enron exhibit at Museum of American Finance, New York (www.moaf.org) and is also recounted in Gladwell (2009).

The Cecchini et al. (2010) Financial Kernel Model

Cecchini et al. (2010) use support vector machines, a machine learning approach, to develop a fraud prediction model using basic financial data. Specifically, they develop a "financial kernel" that allows for a non-linear mapping of (combinations of) raw accounting measures into predictions of whether firms have engaged in fraudulent reporting. We follow Alawadhi et al. (2020) and calculate the Cecchini-Score based on the coefficients from a regression of an indicator variable for whether firms' financial statements are fraudulent (*AAER*) on Cecchini et al.'s (2010) most predictive measures:

$$AAER_{it} = \beta_0 + \beta_1 Sales_{it-1}/PSTK_{it-1} + \beta_2 SG\&A_{it}/IVAO_{it} + \beta_3 Assets_{it}/IVAO_{it-1} + \beta_4 Sales_{it}/IVAO_{it-1} + \beta_5 Assets_{it}/STI_{it} \quad (2)$$

Where *Sales* = Total sales, *PSTK* = Preferred Stock, *SG&A* = Selling, General, and Administrative Expenditures, *IVAO* = Investments, Advances, and Other, *Assets* = Total assets, and *STI* = Short-Term Investments.

The Dechow et al. (2011) F-Score

Dechow et al. (2011) follow a methodology similar to Beneish (1997; 1999) in developing a score to detect accounting fraud and estimate three alternative models that differ in whether the models include returns and/or other non-financial statement data (e.g., number of employees, security issuance). The results are similar across models and we use the following version of the F-Score:

$$F\text{-}SCORE = -7.893 + 0.790RSST_{it} + 2.518\Delta REC_{it} + 1.191\Delta INV_{it} + 1.979\%SOFTAST_{it} + 0.171\Delta CSales_{it} - 0.932\Delta Earnings_{it} + 1.029Issuance_{it} \quad (3)$$

Where:

RSST = *(*Δ*WC* + Δ*NCO* + Δ*FIN*) / Average Total Assets
WC = [Current Assets - Cash and Short-Term Investments] - [Current Liabilities - Debt in Current Liabilities];
NCO = [Total Assets - Current Assets - Investments and Advances - [Total Liabilities - Current Liabilities - Long-Term Debt];
FIN = [Short-Term Investments + Long-Term Investments] - [Long-Term Debt + Debt in Current Liabilities + Preferred Stock]; all following Richardson et al. (2005);

Δ*REC* = ΔAccounts Receivables/Average Total Assets;

Δ*INV* = ΔInventory/Average Total Assets;

%SOFTAST = (Total Assets – Cash - PPE)/Total Assets

Δ*CSALES* = percentage change in cash sales [ΔSales - ΔAccounts Receivables];

Δ*EARNINGS* = [$Earnings_t$/Average Total $Assets_t$] - [$Earnings_{t-1}$/Average Total $Assets_{t-1}$]; and

Issuance = indicator variable coded 1 if the firm issued debt/equity securities during year t.

Dechow et al. (2011) suggest three potential cut-offs for classifying firms as frauds, depending on whether the F-score exceeds either 1.0, 1.85, and 2.45, which correspond to assumed costs ratios of false negatives to false positives of 143:1, 86:1, and 82:1, respectively. In our main analysis, we rely on the 1.0 cutoff, as most studies that employ F-score make use of this cutoff.

The Amiram et al. (2015) FSD Score

Amiram et al. (2015) construct an *FSD Score* by comparing the distribution of first digits in over 100 financial statement items relative to Benford's law, as research in different disciplines has used deviations from Benford's distribution to detect errors or manipulations in data. The FSD Score is based on the mean absolute deviation statistic (MAD) for the financial items considered and is calculated as MAD = (∑|AD-ED|)/K, where AD (ED) is the actual (expected) proportion of leading digits and K is the number of leading digits being analyzed. Amiram et al. (2015) suggest that the FSD score is a useful tool for detecting financial statement errors, as they document that lagged FSD scores positively correlate with material misstatements while contemporaneous FSD scores are negatively correlated with material misstatements.

To evaluate the performance of FSD scores we follow Amiram et al. (2015) and add the FSD score to the F-Score model to investigate the incremental impact of FSD scores on the accuracy of fraud prediction models. Specifically, we re-estimate the F-Score regression on our sample of fraud and non-fraud firms in which we add the FSD score to the model. When it comes to the replication of the F-Score, we find that, with the exception of the change in inventory, coefficients are of the same sign and generally similar in magnitude as the coefficients in Dechow et al. (2011). In line with Amiram et al. (2015), we find a negative and significant relation between the contemporaneous FSD score and the likelihood of misstatements. We calculate the fraud probability based on the MAD-Score as follows.

$$MAD\text{-}SCORE = -6.8188 + 0.8949RSST_{it} + 1.8492\Delta REC_{it} + 0.3407\Delta INV_{it} + 0.0931\Delta CSales_{it} -1.0684\Delta Earnings_{it} + 1.0829Issuance_{it} +1.8808SOFTASSETS_{it} – 0.1684FSD\_Score \quad (4)$$

Where *SOFTASSETS* is defined as total assets less net property, plant, and equipment and cash and cash equivalents, divided by total assets, and all other variables are as defined before.

The Chakrabarty et al. (2020) Adjusted Benford Score

Chakrabarty et al. (2020) develop a fraud prediction model that is in spirit close to the MAD-Score calculated in the previous section. However, rather than directly including the raw FSD-Score in the model, they evaluate the performance of a range of alternative models that further include several standardized variations of the raw score, which adjust the measure for systematic variation across firms, financial statement lengths, industries, and over time. Although they develop a model that relies on (variations of) Benford's Law only, they further test the performance of an expanded model in which the FSD-Score and its relevant standardized counterparts are added to the F-Score specification of Dechow et al. (2011). They conclude that although the model that makes use of Benford's Law data only ensures the broadest applicability, a model that incorporates both the F-Score variables and the variables based on Benford's Law has the highest accuracy, Hence, in our analyses, we evaluate the performance of the adjusted Benford Score that utilizes both the F-Score and Benford's Law variables. Specifically, we calculate the Adjusted Benford Score following the coefficients in Chakrabarty et al. (2020):

$$ABF\text{-}SCORE = -13.75 + 0.74RSST_{it} + 2.27\Delta REC_{it} + 0.72\Delta INV_{it} + 0.11\Delta CSales_{it} – 0.58\Delta Earnings_{it} + 1.08Issuance_{it} +1.81SOFTASSETS_{it} + 1.71B\_Raw_{it} + 0.66B\_Input_{it} – 3.70B\_Year \quad (5)$$

Where *B_Raw* is the raw FSD Score, *B_Input* is the raw score that is adjusted for the number of inputs used in the computation of raw score, and *B_Industry* is the raw score that is adjusted for the average raw FSD Score in an industry and year.[23]

The Misrepresentation Model from Alawadhi et al. (2020)

Alawadhi et al. (2020) develop a misrepresentation model that combines variables previously used in other well-known prediction models (Beneish 1997; 1999; Dechow et al. 2011) as well as a set of new variables. Retaining those variables that, after pooling them into an expanded model, retain their significance, we calculate the score to capture the likelihood of misrepresentation as follows:

$$AKKM\text{-}SCORE = -8.850 + 0.089GMI_{it} + 0.186SGI_{it} + 0.380SGAI_{it} + 0.810\Delta REC_{it} + 2.079\Delta INV_{it} + 1.715SOFTASSETS_{it} + 0.342OLEASE_{it} + 0.497ACT\_Issuance_{it} + 0.3923MCAP_{it} - 0.455ALS\_Accruals_{it} + 0.4153Loss_{it} - 0.13SEGHHI_{it} + 0.1253GEOSEG_{it} + 0.8046OPIN_{it} - 0.6007BigN_{it} + 0.3239BusinessEquipment_{it} - 0.7688Telecom_{it} \quad (6)$$

Where *OLEASE* is an indicator variable that is equal to one in the presence of operating lease arrangements, zero otherwise, *MCAP* is the natural logarithm of the market value of equity, *ALS_Accruals* are discretionary accruals calculated following Allen, Larson, and Sloan (2013), *Loss* is an indicator variable that is equal to one if net income is negative, zero otherwise, *SEGHHI* is the Herfindahl-Hirschmann index of (business) segment sales, *GEOSEG* is the natural logarithm of one plus the number of geographic segments, *OPIN* is an indicator variable that is equal to one for audit opinions other than unqualified, zero otherwise, *BigN* is an indicator variable that is equal to one if the firm is audited by a BigN auditor, zero otherwise, and *BusinessEquipment* and *Telecom* are indicator variables for whether the firm is active in one of those two industries. Other variables that derive from the M-Score and the F-Score are as defined before.

The Machine Learning-Based Prediction from Bao et al. (2020)

Bao et al. (2020) use a machine learning approach to develop a fraud prediction model that, in line with Cecchini et al. (2010), makes use of raw accounting data, rather than predetermined accounting ratios. They then use ensemble learning and an AdaBoost algorithm on a training sample to develop a strong predictor of fraud. Mechanically, the training occurs in iterations in which each subsequent iteration is aimed at improving the predictions of cases that were misclassified in the previous iteration. The final prediction is then a weighted average of the outcomes of the previous iterations in which iterations with a higher accuracy receive a higher weight.

We follow Bao et al. (2020) and use their list of 28 raw financial statement items in the model.[24] For every year in our sample, we train the sample on all previous years of data, requiring a

[23] To calculate *B_Input*, every year, we sort observations into 20 groups based on the number of inputs used in the calculation of *B_Raw*. We then calculate *B_Input* by subtracting from *B_Raw*, the average value of *B_Raw* of the respective bins and dividing by the standard deviation of *B_Raw* in each of the bins. *B_Industry* is calculated by subtracting from *B_Raw*, the industry-year average value of *B_Raw* and dividing by the standard deviation of *B_Raw* in the industry-year.

[24] The 28 items are common shares outstanding, total current assets, sale of common and preferred stock, net PPE, trade payables, cash and short-term investments, stock price at fiscal year-end, retained earnings, inventory, common equity, debt in current liabilities, depreciation and amortization, accounts receivable, cost of goods sold, total assets, long-term debt issuance, income before extraordinary items, long-term debt, interest and related expense, income tax

minimum of 5 years of data and a one-year lag between the training sample and our predictions. Hence, we start the sample in 1985 in which the fraud predictions are based on a sample and algorithm that is trained on the years 1980-1984. Similarly, the predictions for 1991 follow from an algorithm that is trained over the period 1980-1990, whereas predictions for 2000 follow from the 1980-1999 training period.[25]

expense, total current liabilities, sales, income tax payable, investment and advances, total liabilities, short-term investments, net income, and preferred stock.

[25] We are thankful to Yang Bao, Bin Ke, Bin Li, Julia Yu, and Jie Zhang for providing access to the code to estimate their model and to run the RUSBoost algorithm.

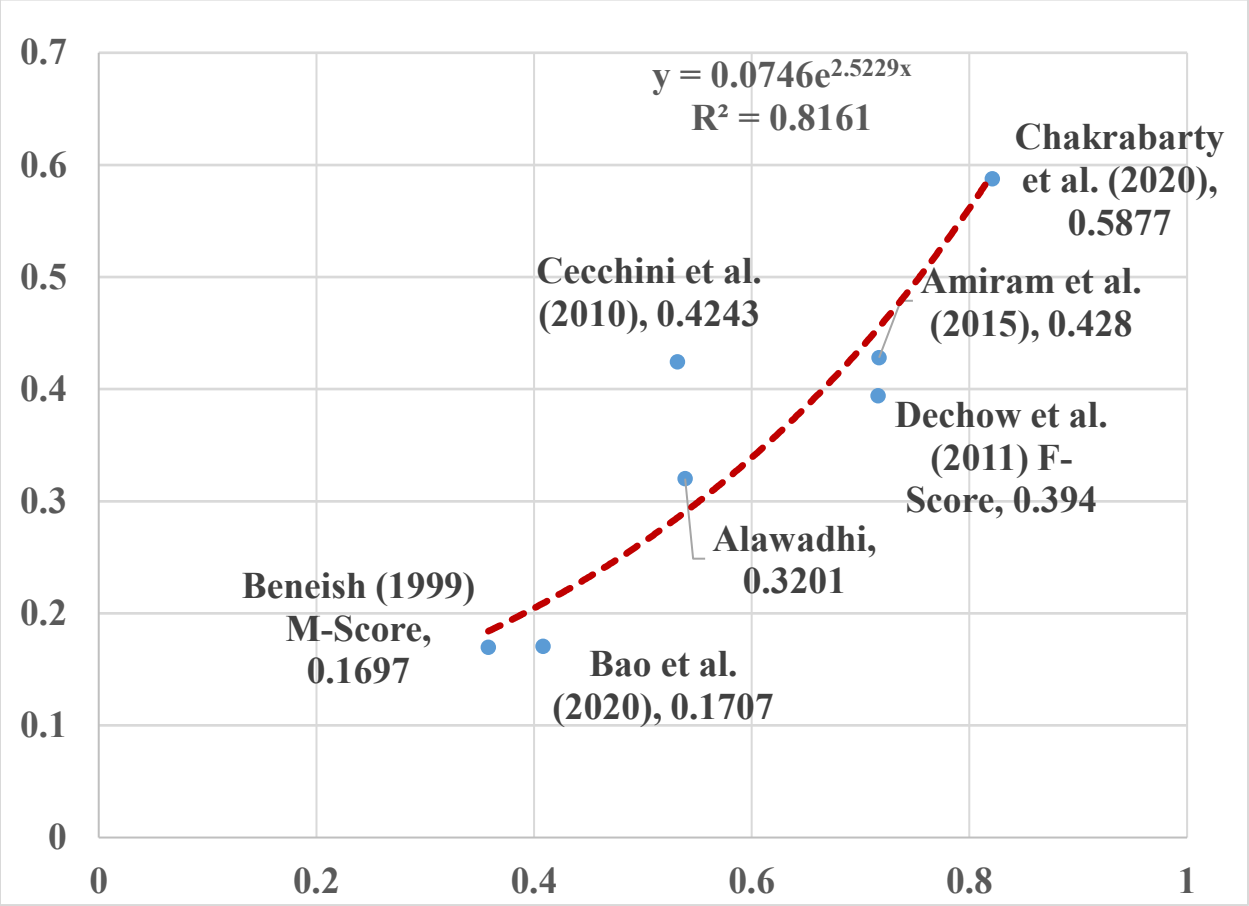
Figure 1: Overview of the Costs & Benefits of Fraud Prediction Models

		Actual	
		Fraud	Non-Fraud
Predicted	Fraud	True Positive (Benefit is the cost of false negative that is avoided)	False Positive (Cost of false positive is incurred)
	Non-Fraud	False Negative (False negative cost is incurred whether model is used or not)	True Negative (No cost incurred whether model is used or not)
Precision		True positive /Predicted Fraud	
Sensitivity or recall		True positive / Total Fraud	
The net benefit or net cost of using a fraud prediction model is based on the *upper row* of the matrix where the model either predicts fraud correctly (true positive) or incorrectly (false positive). If fraud is not predicted, there is no difference in costs relative to not using a model.			

Figure 2: Overview of the Components of the Costs of Fraud Model Classification Errors

	Auditors	Investors	Regulators
False Positive Costs (Incorrectly Flagged Non-Fraud Firms)	**Incremental audit work assuming it cannot be billed to audit client--AT COST.** This follows experimental evidence that auditors increase the audit investment when they perceive a higher risk of misstatement (e.g., Glover et al. 2003; Hammersley et al. 2011; for a Boritz et al. 2015), and empirical evidence that a risk factor such as a control weakness or the payment of bribes results in higher fees (Hogan and Wilkins 2008; Lyon and Maher 2008; Munsif et al. (2011). Alternatively, one year of lost audit fees from resigning from the audit client (or profit or such fees).	**Incremental audit work assuming it is billed to audit client--AT MARKET.** This follows experimental evidence that auditors increase the audit investment when they perceive a higher risk of misstatement (e.g., Glover et al. 2003; Hammersley et al. 2011; Boritz et al. 2015), and empirical evidence that a risk factor such as a control weakness or the payment of bribes results in higher fees (Hogan and Wilkins 2008; Lyon and Maher 2008; Munsif et al. (2011).	If investigations into false positives are made public: Costs estimates range from **1 to 15% of the market value of equity** at the end of month three subsequent to firms fiscal year ends. Percentages encompass potential market value losses associated with denials of Freedom of Information Act requests, and three-day average losses associated with initial announcements of Wells notices, restatements involving irregularities, and SEC investigations.
	Discovery Costs. We expect that presence of a fraud flag in a prior year's audit working papers to make it easier for the plaintiff's bar to argue scienter for a fraud occurring in a subsequent period. We estimate 'discovery costs' as the product of the estimated probability of litigation against an auditor and the estimated settlement amount relying on models motivated by Honigsberg et al. (2020).	**Loss avoided (profit foregone)** by not investing in false positive firms over a 12-month period beginning three months after fiscal year end.	
False Negatives Costs (Missed Fraud Detection)	**Litigation Costs:** Auditor litigation costs are based on actual litigation outcomes in litigation cases against auditors. Data are obtained from extensive searches on Audit Analytics, Westlaw, Lexis Nexis, and Factiva. Our searches begin with the firm name at the time the fraud is discovered, but also consider name changes as applicable, specify a search period beginning two years before and ending two years after the misreporting period identified in AAERs	**Loss on the market:** Abnormal Returns in period varying from days -1 to +1 to days -1 to +252 relative to the first public revelation of the fraud times the stock's market value on day -2.	The difference in value-weighted market returns on revelation days over the period 1982-2016 times the value of the stock market in aggregate two days prior to the fraud-revealing announcements.
	Reputation Losses: Market value loss at fraud revelation due to contagion effect on other clients of the auditor in the same industry (e.g., Gleason et al. 2008; Weber et al. 2008)		
	Client Loss: Profits foregone on Abnormal Client Loss in the two years following the revelation of the fraud (e.g., Lyon and Maher 2004)		

Figure 3: False Positive Rates versus True Positive Rates



False positive rates are depicted on the vertical axis, true positive rates on the horizontal axis.

TABLE 1

Sample Selection & AAER/Restatement Market Reactions

Panel A: Sample Selection

AAER (1982-2016)	**# Firms**	**# Firm-Years**
Total Sample of Fraud Cases	574	
Less: Fraud Cases Not Matched to Compustat-CRSP	-80	
Fraud Cases Matched to Compustat-CRSP	494	1,185
Less: Fraud Cases with Missing Announcement Returns	-79	-144
Fraud Cases with Announcement Returns	413	1,041
Less: Fraud Cases with Missing M-Score or F-Score	-100	-273
Final Sample of Fraud Cases	313	768

Restatements (2000-2018)	**# Firms**	**# Firm-Years**
Audit Analytics Restatements	18,768	
Less Clerical Errors and non-Adverse Restatements	-3,281	
Restatements with adverse impact	15,487	
Less: Restatements Not Matched to Compustat-CRSP	-10,490	
Restatements Matched to Compustat-CRSP	4,997	11,604
Less: Restatements with Missing Announcement Returns	-211	-350
Fraud Cases with Announcement Returns	4,786	11,254
Less: Restatements with Missing M-Score or F-Score	-2,395	-5,846
Final Sample of all Restatement Cases	2,391	5,408
Final Sample of Severe Restatements (Cases disclosed via Press Releases or 8-K)	1,115	2,869

Panel B: Three-day Market reaction

Market-Adjusted Abnormal Returns	**N**	**Mean (%)**	**Mean ($)**	**Median %**	**Median $**	**Sum ($)**
Fraud Cases	313	-16.15%	-$446.63	-10.61%	-$20.61	-$139,794.72
All Restatements	2,391	-1.30%	-$18.49	-0.66%	-$0.95	-$44,209.59
Severe Restatements	1,115	-2.79%	-$47.92	-1.50%	-$2.47	-$53,430.80

Panel C: Overview of the Number of AAERs, Restatements, and Severe Restatements by Fiscal Year.

Fiscal Year	*AAER (all)*	*AAER (first)*	*Res (all)*	*Res (first)*	*Severe Res (all)*	*Severe Res (first)*	*Fiscal Year*	*AAER (all)*	*AAER (first)*	*Res (all)*	*Res (first)*	*Severe Res (all)*	*Severe Res (first)*
1979	2	2					1999	60	31				
1980	6	4					2000	75	37				
1981	6	1					2001	73	26	219	215	138	135
1982	8	5					2002	60	12	416	238	285	163
1983	7	3					2003	51	11	530	228	380	150
1984	10	6					2004	36	7	556	174	435	129
1985	9	6					2005	25	5	398	121	301	70
1986	13	6					2006	16	4	289	111	196	62
1987	12	7					2007	15	5	246	97	142	47
1988	11	7					2008	12	4	224	98	110	41
1989	11	7					2009	15	6	237	116	103	43
1990	10	4					2010	15	3	297	144	119	58
1991	15	8					2011	11	2	342	161	130	49
1992	20	14					2012	8	2	369	172	116	42
1993	16	7					2013	6	1	330	130	103	37
1994	17	13					2014	3		275	123	93	35
1995	17	8					2015			262	107	91	27
1996	24	15					2016			201	77	66	13
1997	35	19					2017			139	53	40	11
1998	38	15					2018			78	26	21	3

This table reports the sample selection (Panel A) and reports descriptive statistics of the number of fraud and restatement firms by year (Panel C). Panel B reports descriptive statistics of the market reaction to AAER and restatement announcements using the three-day market-adjusted CAR around fraud and restatement announcement. Dollar returns are calculated by multiplying the three-day announcement CAR by the market value of equity at day -2 relative to announcement. **Data are in millions of 2016 dollars.**

TABLE 2:

Assessing Model Performance with Traditional Measures

Panel A: AAERs - All Observations

	Fraud Observations	Non-Fraud Observations	AUC	Sensitivity	Precision	False Positive Rate	False to True Positive Rate
Beneish (1999) M-Score	768	136,144	0.577	23.18%	0.76%	16.97%	130
Cecchini et al. (2010)	615	102,273	0.578	57.89%	0.81%	42.43%	122
Dechow et al. (2011) F-Score	768	136,144	0.673	64.71%	0.92%	39.40%	108
Amiram et al. (2015) FSD Score	766	131,978	0.679	68.41%	0.92%	42.83%	108
Alawadhi et al. (2020)	626	106,281	0.728	62.62%	1.14%	32.01%	87
Bao et al. (2020)	722	115,948	0.563	40.17%	1.44%	17.07%	68
Chakrabarty et al. (2020) ABF Score	768	134,797	0.689	82.03%	0.79%	58.77%	126

Panel B: AAERs - Unique Observations

	Fraud Observations	Non-Fraud Observations	AUC	Sensitivity	Precision	False Positive Rate	False to True Positive Rate
Beneish (1999) M-Score	313	136,144	0.668	35.78%	0.48%	16.97%	206
Cecchini et al. (2010)	252	102,273	0.562	53.17%	0.31%	42.43%	324
Dechow et al. (2011) F-Score	313	136,144	0.717	71.57%	0.42%	39.40%	239
Amiram et al. (2015) FSD Score	311	131,978	0.709	71.70%	0.39%	42.83%	254
Alawadhi et al. (2020)	219	106,281	0.691	53.88%	0.35%	32.01%	288
Bao et al. (2020)	289	115,948	0.537	40.83%	0.59%	17.07%	168
Chakrabarty et al. (2020) ABF Score	313	134,797	0.713	82.11%	0.32%	58.77%	308

Panel C: Restatements - Unique Observations

	Number of Restatements	Number Non-Restatements	AUC	Sensitivity	Precision	False Positive Rate	False to True Positive Rate
Beneish (1999) M-Score	2,391	49,075	0.518	14.89%	5.68%	12.06%	17
Cecchini et al. (2010)	1,895	38,679	0.514	54.14%	4.87%	51.82%	20
Dechow et al. (2011) F-Score	2,391	49,075	0.556	44.33%	5.49%	37.18%	17
Amiram et al. (2015) FSD Score	2,217	45,736	0.555	49.21%	5.34%	42.32%	18
Alawadhi et al. (2020)	2,074	40,281	0.489	48.75%	4.82%	49.54%	20
Bao et al. (2020)	2,189	45,461	0.515	19.96%	5.35%	17.00%	18
Chakrabarty et al. (2020) ABF Score	2,371	48,506	0.576	63.01%	5.51%	52.86%	17

Panel D: Restatements - Unique Observations - 8K/Press/Expanded

	Number of Restatements	Number Non-Restatements	AUC	Sensitivity	Precision	False Positive Rate	False to True Positive Rate
Beneish (1999) M-Score	1,115	51,780	0.510	16.05%	2.81%	11.98%	35
Cecchini et al. (2010)	940	40,818	0.486	53.09%	2.29%	52.05%	43
Dechow et al. (2011) F-Score	1,115	51,780	0.570	46.91%	2.64%	37.29%	37
Amiram et al. (2015) FSD Score	1,087	48,271	0.565	51.61%	2.66%	42.49%	37
Alawadhi et al. (2020)	982	42,644	0.522	42.67%	1.94%	49.73%	51
Bao et al. (2020)	1,082	47,939	0.587	20.52%	2.64%	17.06%	37
Chakrabarty et al. (2020) ABF Score	1,114	51,182	0.581	65.26%	2.60%	53.19%	37

This table reports traditional fraud prediction model performance measures. Panel A and Panel B report results on a sample of AAERs. They differ in that the former uses all fraud firm-years and the latter only the first year in which a fraud occurs to compare models. Treating firms that commit fraud in consecutive years as a unique instance of fraud is consistent with the costs of fraud being imposed just once after the fraud becomes publicly revealed. Panel C and Panel D report results on a sample of all or severe restatements, respectively. The number of observations differs across models as a result of data requirements. Area under the Receiver Operating Characteristics curve (AUC) is the area under the curve that plots the true positive and false positive rates using every observation as a possible cut-off. Sensitivity is the true positive rate computed as the ratio of true positives to all positives (flagged frauds to all frauds). Precision is the ratio of true positives to the sum of true and false positives (flagged frauds to all flagged observations). The False Positive Rate is the ratio of non-fraud flagged firms to all non-fraud observations. Sensitivity to precision measures the relative size of the sample of flagged firms to that of the sample of actual misreporting firms. The false to true positive rate measures the average number of false flags for each true flag: assuming equal costs, it represents the rate at which a model trades of the costs of false positives for the benefit of avoiding the cost of a false negative. Precision to prevalence is the precision deflated by the unconditional probability of fraud/restatement occurrence in the sample, which is useful in comparing model performance across fraud and restatement samples that are characterized by different levels of occurrence. Our calculations rely on published cut-offs with two exceptions: (1) for the Bao et al. (2020) measure, we use the flags generated by their model when applied to our sample firms; (2) for the Cecchini et al. (2010) measure, we estimate the cut-off that minimizes expected misclassification costs on our whole sample assuming a 200:1 cost ratio, which is the cost ratio that produces the best classification results in their original article.

TABLE 3

Costs and Benefits of Predictions Errors for Auditors

Panel A: False Negative Costs (True Positive Benefits)

		Litigation Against Auditors		Settlement in 2016 $	
Litigation Costs	**N**	**Number Sued**	**% Sued**	**Mean**	**Total**
Fraud Firms	313	87	27.80%	$10.12	$3,128.50
All Restatements	2,391	37	1.55%	$0.02	$58.58
Severe Restatements	1,115	29	2.60%	$0.05	$55.80

Client Losses	**N**	**Mean**	**Std Dev**	**Median**	**Total**
Fraud Firms	313	$20.11	$48.72	$6.50	$6,296.10
All Restatements	2,391	$16.34	$36.34	$6.54	$39,064.89
Severe Restatements	1,115	$18.21	$37.69	$10.35	$20,303.34

Reputation Losses	**N**	**Mean**	**Std Dev**	**Median**	**Total**
Fraud Firms	313	$116.54	$2.085.85	$0.00	$36,477.25
All Restatements	2,391	-$21.02	$1.847.05	$0.00	-$50,263.50
Severe Restatements	1,115	-$95.27	$1.835.29	$0.00	-$106,231.72

Panel B: False Positive Costs

Extra Audit Investment	**N**	**Mean**	**Std Dev**	**Median**	**Total**
Fraud Firms	136,144	$0.27	$0.60	$0.09	$36,585.13
All Restatements	49,075	$0.50	$0.90	$0.20	$24,440.63
Severe Restatements	51,780	$0.50	$0.90	$0.21	$26,061.36

Discovery Cost	**N**	**Mean**	**Std Dev**	**Median**	**Total**
Fraud Firms	136,144	$0.93	$1.74	$0.23	$126,953.11
All Restatements	49,075	$0.64	$1.46	$0.01	$31,328.31
Severe Restatements	51,780	$0.62	$1.44	$0.01	$32,342.76

This table reports the true positive benefits and false positive costs for auditors. Panel A reports auditors' true positive benefits/false negative costs. Auditor litigation costs are based on actual litigation outcomes in litigation cases against auditors. For each of the 313 fraud firms in the sample, we identify lawsuits filed against its auditors, as well as the lawsuit's resolution. Data are obtained from extensive searches on Audit Analytics, Westlaw, Lexis Nexis, and Factiva. Our searches begin with the firm name at the time the fraud is discovered, but also consider name changes as applicable, specify a search period beginning two years before and ending two years after the fraud period identified in AAERs, and uses search terms such as "securities", "class action", litigation", "lawsuit" and "settlement." Data on litigation in restatement cases is obtained from Audit Analytics. Auditors' reputation loss is the average price reaction in the three days surrounding the fraud-revelation announcement of the other clients of the auditor in the same industry as the fraud/restatement firm (6-digit GICS) relative to that of the other firms in the same industry, but which are not clients of the auditor. Client loss is the abnormal fee loss following the fraud/restatement announcement, measured as the audit fee lost in the two years after the AAER/restatement announcement less the "normal" fee loss which is estimated as the auditor's fee loss in the two years prior to the AAER/restatement announcement. Panel B reports auditors' false positive costs. Extra audit investment is calculated as EAUDi[FP COST] = p (AUDFEEjt) + (1-p) (INCRAUDFEEjt * (1-pm%)), where p is the probability of resignation, AUDFEEjt is the fee lost by the resigning auditor in year t, INCRAUDFEEjt is the value of the additional work undertaken by the auditor as a result of the warning flag (which we assume that they cannot pass on to the client in year t), and where pm% represents the auditor's usual profit margin. Discovery Cost is estimated as the product of the estimated probability of litigation against an auditor times the estimated settlement relying on models motivated by Honigsberg et al. (2020) as described in equations (7) and (8) below. We estimate the probability of litigation (PROB_LIT) to be equal to 1/(1+exp(-Litigation)) where Litigation is estimated using logistic regression on the sample of 313 instances of frauds as:

Litigation = -1.8688 + (FY_ABNRET*-0.0905) + (GROWTH*0.393) + (BTM*-0.4851) +

(HI-LIT_IND*-0.0428) + (LN_MVAL*0.155) (7)

Where Litigation is an indicator variable with 1 if the auditor is sued, FY_ABNRET is the value-weighted market-adjusted return over the fiscal year, GROWTH represents the most recent sales growth, BTM is the book to market ratio at the end of the current fiscal year, HI-LIT_IND represents high litigation industries and LN_MVAL is the natural logarithm of the market value of equity at the end of the fiscal year. GROWTH and LN_MVAL are the only variables that attain significance with p-values of 0.01 and 0.02, respectively and results are qualitatively similar when the model only contains those variables.

We estimate settlement amounts as (-1+ exp(log(1+settle)) using OLS on the subsample of 87 fraud cases in which the auditor is sued, and log(1+settle) is predicted from the following model:

log(1+settle) = -0.74409 + (GROWTH*-.06655) + (LN_MVAL*0.38181) + (ROA*0.59099) + HI_LIT*-.04124) + (BIGN*-0.34835) + (NINE_ELEVEN*0.49598) + (TWO_THREE*0.48605) + (POST_TELLABS*-1.97252) (8)

Where ROA is the ratio of income before extraordinary items to total assets, BIGN is an indicator variable that is equal to one if the firm is audited by a BigN auditor, NINE_ELEVEN, TWO_THREE, and POST_TELLABS are variables identified by Honigsberg et al. (2020) as affecting the likelihood of successful lawsuits against auditors. Specifically, Honigsberg et al. (2020) suggest that federal regional court districts nine and eleven (two and three) held plaintiffs to a lower (higher) standard to plead scienter in lawsuits against auditors. Further, they provide evidence on how two Supreme Court rulings (Tellabs in 2007 and Janus in 2011) altered the likelihood of a successful lawsuit outcome as a function of the circuit court in which the litigation proceeds. Although we do not have enough data to analyze a number of these combinations, we do find that the amounts of settlement amounts decline after 2007. LN_MVAL is the only variable attaining significance at the 5% level, although POST_TELLLABS is marginally significant at the 7% level. We find similar results when the model is restricted to only those variables.

Data are in millions of 2016 dollars.

TABLE 4

Summary of Net Benefits to Auditors Across Models and Misstatement Types

Panel A: AAERs

	True Positives			*False Positives*				*Net Benefits (Costs)*	
Models/Cutoffs	N	Mean Benefit (Avoiding litigation, client, & reputation costs)	Total Benefit (Avoiding litigation, client loss, & reputation costs)	N	Mean Cost (Discovery + Extra Audit Inv.)	Mean Cost (Discovery)	Total Cost (Discovery+ Extra Audit Inv.)	Total Cost (Discovery)	All benefits less all costs	All benefits less discovery costs
M-Score	112	$127.21	$14,247	23,101	$0.88	$0.77	$20,363	$17,842	-$6,116	-$3,595
F-Score > 1	224	$30.95	$6,933	53,643	$1.22	$0.96	$65,362	$51,488	-$58,429	-$44,555
F-Score > 1.85	93	$88.77	$8,256	12,590	$0.98	$0.84	$12,375	$10,532	-$4,119	-$2,276
F-Score > 2.45	56	$311.66	$17,453	5,431	$0.95	$0.83	$5,143	$4,526	**$12,310**	**$12,927**
Cecchini et al.	134	$131.33	$17,598	43,391	$1.44	$1.15	$62,274	$50,064	-$44,676	-$32,466
Amiram et al.	223	$59.31	$13,225	56,532	$1.29	$1.00	$72,907	$56,685	-$59,682	-$43,460
Alawadhi et al.	118	$487.76	$57,556	34,023	$2.61	$2.01	$88,939	$68,373	-$31,383	-$10,817
Bao et al.	118	$33.37	$3,937	19,793	$1.59	$1.23	$31,481	$24,338	-$27,544	-$20,401
Chakrabarty et al.	257	$137.20	$35,260	79,217	$1.42	$1.14	$112,785	$90,158	-$77,525	-$54,898

Panel B: Restatements

	True Positives			*False Positives*				*Net Benefits (Costs)*	
Models/Cutoffs	N	Mean Benefit (Avoiding litigation, client, & reputation costs)	Total Benefit (Avoiding litigation, client loss, & reputation costs)	N	Mean Cost (Discovery + Extra Audit Inv.)	Mean Cost (Discovery)	Total Cost (Discovery+ Extra Audit Inv.)	Total Cost (Discovery)	All benefits less all costs	All benefits less discovery costs
M-Score	356	$71.06	$25,298	5,916	$0.71	$0.50	$4,228	$2,981	$21,071	$22,318
F-Score > 1	1060	$14.97	$15,866	18,248	$1.29	$0.75	$23,560	$13,726	-$7,693	$2,141
F-Score > 1.85	245	$10.64	$2,608	3,114	$1.01	$0.67	$3,161	$2,078	-$553	$530
F-Score > 2.45	95	$112.00	$10,640	974	$0.94	$0.64	$911	$624	**$9,729**	**$10,016**
Cecchini et al.	1026	$17.31	$17,762	20,043	$1.15	$0.74	$23,049	$14,843	-$5,287	$2,919
Amiram et al.	1091	$17.25	$18,819	19,356	$1.37	$0.81	$26,572	$15,640	-$7,753	$3,179
Alawadhi et al.	1011	$15.42	$15,587	19,957	$1.88	$1.09	$37,605	$21,766	-$22,018	-$6,178
Bao et al.	437	$20.63	$9,015	7,728	$1.96	$1.27	$15,111	$9,802	-$6,096	-$787
Chakrabarty et al.	1494	$16.76	$25,037	25,638	$1.42	$0.83	$36,365	$21,338	-$11,328	$3,699

Panel C: Severe Restatements

	True Positives			*False Positives*				*Net Benefits (Costs)*	
Models/Cutoffs	N	Mean Benefit (Avoiding litigation, client, & reputation costs)	Total Benefit (Avoiding litigation, client loss, & reputation costs)	N	Mean Cost (Discovery + Extra Audit Inv.)	Mean Cost (Discovery)	Total Cost (Discovery+ Extra Audit Inv.)	Total Cost (Discovery)	All benefits less all costs	All benefits less discovery costs
M-Score	179	$13.64	$2,441	6,201	$0.72	$0.50	$4,439	$3,109	-$1,998	-$668
F-Score > 1	523	$15.97	$8,350	19,309	$1.28	$0.73	$24,643	$14,106	-$16,294	-$5,757
F-Score > 1.85	122	$9.04	$1,102	3,303	$1.01	$0.66	$3,349	$2,179	-$2,247	-$1,077
F-Score > 2.45	51	$66.27	$3,380	1,041	$0.93	$0.63	$972	$655	**$2,408**	**$2,725**
Cecchini et al.	499	$19.44	$9,700	21,245	$1.13	$0.72	$24,051	$15,270	-$14,351	-$5,570
Amiram et al.	561	$17.44	$9,782	20,512	$1.35	$0.78	$27,789	$16,070	-$18,007	-$6,288
Alawadhi et al.	419	$18.25	$7,645	21,208	$1.85	$1.05	$39,278	$22,360	-$31,633	-$14,714
Bao et al.	222	$20.70	$4,596	8,178	$1.92	$1.23	$15,730	$10,063	-$11,134	-$5,467
Chakrabarty et al.	727	$18.38	$13,362	27,225	$1.40	$0.81	$38,085	$22,000	-$24,723	-$8,638

This table reports the results of analyzing the cost of fraud prediction errors to auditors for AAERs (Panel A), Restatements (Panel B), and Severe Restatements (Panel C). Auditors' false positive costs consist of two components: (i) the costs of the incremental audit work, and (ii) discovery costs. Extra audit investment is calculated as EAUDi[FP COST] = p (AUDFEEjt) + (1-p) (INCRAUDFEEjt * (1-pm%)), where p is the probability of resignation, AUDFEEjt is the fee lost by the resigning auditor in year t, INCRAUDFEEjt is the value of the additional work undertaken by the auditor as a result of the warning flag (which we assume that they cannot pass on to the client in year t), and where pm% represents the auditor's usual profit margin. Discovery Cost is estimated as the product of the estimated probability of litigation against an auditor times the estimated settlement relying on models motivated by Honigsberg et al. (2020) as described in equations (8) and (9). Auditors' true positive benefits consist of three components: (i) litigation costs, (ii) client losses, and (iii) reputation losses. Auditor litigation costs are based on actual litigation outcomes in litigation cases against auditors. For each of the 313 fraud firms in the sample, we identify lawsuits filed against its auditors, as well as the lawsuit's resolution. Data are obtained from extensive searches on Audit Analytics, Westlaw, Lexis Nexis, and Factiva. Our searches begin with the firm name at the time the fraud is discovered, but also consider name changes as applicable, specify a search period beginning two years before and ending two years after the fraud period identified in AAERs, and uses search terms such as "securities", "class action", litigation", "lawsuit" and "settlement." Data on litigation in restatement cases is obtained from Audit Analytics. Auditors' reputation loss is the average price reaction in the three days surrounding the fraud-revelation announcement of the other clients of the auditor in the same industry as the fraud/restatement firm (6-digit GICS) relative to that of the other firms in the same industry, but which are not clients of the auditor. Client loss is the abnormal fee loss following the fraud/restatement announcement, measured as the audit fee lost in the two years after the AAER/restatement announcement less the "normal" fee loss which is estimated as the auditor's fee loss in the two years prior to the AAER/restatement announcement. Mean benefit/cost measures the average true positive and false positive benefits/costs per instance. Net benefits are calculated by summing for each model the true positive benefits and subtracting the sum of false positive costs (mean benefit/cost * N). The table reports several cost aggregated that differ depending on whether auditors' false positive costs include the extra audit investment or whether the auditors' true positive benefits include reputation losses. **Data are in millions of 2016 dollars.**

TABLE 5

Summary of Net Benefits to Investors Across Models and Misstatement Types

Pane A: AAER

	True Positives				*False Positives*				*Net: TPB - FPC*	
Models/Cutoffs	**N**	**Mean Benefit (%)**	**Mean Benefit ($)**	**Mean Benefit $ (Winsorized)**	**N**	**Mean Cost (%)**	**Mean Cost ($)**	**Mean Cost $ (Winsorized)**	**Net Benefit (Cost)**	**Net Benefit (Cost) Winsorized**
M-Score	112	21.85%	$ 140.87	$ 140.87	23,101	-4.00%	$ -68.06	$ -36.05	$1,588,013	$848,673
F-Score > 1	224	16.50%	$ 406.22	$ 364.12	53,643	1.26%	$ 1.66	$ 4.09	$2,039	-$137,972
F-Score > 1.85	93	22.08%	$ 723.06	$ 582.39	12,590	-6.05%	$ -67.03	$ -37.18	$911,092	$522,275
F-Score > 2.45	56	23.64%	$ 213.89	$ 213.89	5,431	-11.15%	$ -158.04	$ -62.30	$870,309	$350,347
Cecchini et al.	134	14.19%	$ 595.30	$ 524.92	43,391	5.42%	$ 29.48	$ 10.37	-$1,199,374	-$379,759
Amiram et al.	223	16.76%	$ 427.00	$ 384.71	56,532	1.57%	$ 10.83	$ 8.03	-$517,164	-$368,048
Alawadhi et al.	118	13.28%	$ 977.33	$ 900.77	34,023	2.19%	$ 25.09	$ 10.74	-$738,397	-$259,017
Bao et al.	118	16.87%	$ 340.01	$ 331.06	19,793	-1.31%	$ 24.72	$ 16.64	-$449,204	-$290,286
Chakrabarty et al.	257	16.65%	$ 397.35	$ 362.20	79,217	3.25%	$ 26.55	$ 14.90	-$2,000,916	-$1,087,271

Panel B: Restatements

	True Positives				*False Positives*				*Net: TPB - FPC*	
Models/Cutoffs	**N**	**Mean Benefit (%)**	**Mean Benefit ($)**	**Mean Benefit $ (Winsorized)**	**N**	**Mean Cost (%)**	**Mean Cost ($)**	**Mean Cost $ (Winsorized)**	**Net Benefit (Cost)**	**Net Benefit (Cost) Winsorized**
M-Score	356	2.19%	$ 20.31	$ 14.41	5,916	3.65%	$ -12.86	$ -4.44	$83,281	$31,410
F-Score > 1	1,060	1.62%	$ 16.92	$ 16.53	18,248	5.27%	$ 56.85	$ 49.73	-$1,019,462	-$889,961
F-Score > 1.85	245	3.70%	$ 19.22	$ 26.27	3,114	-0.53%	$ -39.80	$ -12.63	$128,648	$45,758
F-Score > 2.45	95	3.35%	$ 10.96	$ 32.09	974	-5.22%	$ -228.89	$ -69.92	$223,980	$71,147
Cecchini et al.	1,026	1.06%	$ 43.11	$ 26.75	20,043	8.28%	$ 53.44	$ 38.72	-$1,026,939	-$748,713
Amiram et al.	1,091	1.80%	$ 48.07	$ 20.31	19,356	6.15%	$ 55.14	$ 61.12	-$1,014,767	-$1,160,877
Alawadhi et al.	1,011	1.36%	$ 22.17	$ 20.56	19,957	3.14%	$ 30.11	$ 33.78	-$578,399	-$653,383
Bao et al.	437	1.55%	$ 32.88	$ 24.97	7,728	5.58%	$ 85.97	$ 60.55	-$649,989	-$457,055
Chakrabarty et al.	1,494	1.50%	$ 29.66	$ 20.16	25,638	6.67%	$ 57.39	$ 55.95	-$1,426,919	-$1,404,323

Panel C: Severe Restatements

	True Positives				*False Positives*				*Net: TPB - FPC*	
Models/Cutoffs	**N**	**Mean Benefit (%)**	**Mean Benefit ($)**	**Mean Benefit $ (Winsorized)**	**N**	**Mean Cost (%)**	**Mean Cost ($)**	**Mean Cost $ (Winsorized)**	**Net Benefit (Cost)**	**Net Benefit (Cost) Winsorized**
M-Score	179	3.95%	$ 41.14	$ 25.61	6,201	2.96%	$ -20.14	$ -12.12	$132,268	$79,728
F-Score > 1	523	3.62%	$ 49.68	$ 37.25	19,309	5.18%	$ 51.38	$ 45.14	-$966,022	-$852,130
F-Score > 1.85	122	7.09%	$ 61.32	$ 48.25	3,303	-0.72%	$ -46.75	$ -21.27	$161,888	$76,131
F-Score > 2.45	51	5.83%	$ 47.04	$ 42.58	1,041	-4.96%	$ -220.90	$ -70.86	$232,355	$75,939
Cecchini et al.	499	2.87%	$ 38.26	$ 22.90	21,245	8.18%	$ 48.85	$ 34.65	-$1,018,756	-$724,657
Amiram et al.	561	3.71%	$ 45.48	$ 48.32	20,512	6.20%	$ 52.62	$ 57.95	-$1,053,879	-$1,161,647
Alawadhi et al.	419	3.10%	$ 11.94	$ 54.92	21,208	3.03%	$ 20.96	$ 23.02	-$439,556	-$465,164
Bao et al.	222	3.93%	$ 101.01	$ 73.40	8,178	5.68%	$ 86.74	$ 57.29	-$686,902	-$452,238
Chakrabarty et al.	727	3.18%	$ 42.30	$ 39.09	27,225	6.49%	$ 49.33	$ 47.67	-$1,312,300	-$1,269,413

This table reports the results of analyzing the cost of fraud prediction errors to investors for AAERs, Restatements, and Severe Restatements, using market-adjusted abnormal returns. Investors' false positive costs have two components: (i) as residual claimants, investors bear the costs of additional audit fees imposed by auditors as a result of classifying the firm as a potential fraud (INCRAUDFEEjt), and (ii) the profit foregone or the loss avoided by not investing in the firms that are flagged by a model. In case of losses avoided models could thus generate a "false positive benefit" to investors. The latter variable is measured as the abnormal return over the 12-month period starting in the fourth month after the end of the fiscal year in which the firm is flagged multiplied by the market value of equity at the start of the return accumulation period. Investors' true positive benefits are the abnormal dollar value loss (market value of equity on day -2 multiplied by the abnormal return over days -1 to 1 relative the fraud/restatement revelation announcement. Mean benefit/cost measures the average true positive and false positive benefits/costs per instance. Net benefits are calculated by summing for each model the true positive benefits and subtracting the sum of false positive costs (mean benefit/cost * N). In case of winsorized returns, variables are winsorized at the 1% and 99% levels, respectively, **Data are in millions of 2016 dollars.**

TABLE 6

Regulators' Costs and Benefits

Panel A: False Negative Costs/True Positive Benefits: Announcements and Market-Wide Returns

Return on market-index on Indicator for Revelation date				
Value-weighted	Coeff.	SE	T-stat	p-value
Intercept	0.00054145	0.00011071	4.89	<.0001
Event	-0.00119	0.00050534	-2.36	0.0185
Equal-weighted				
Intercept	0.00084437	0.00009201	9.18	<.0001
Event	-0.00096486	0.00041997	-2.30	0.0216
Return on market-index on indicator for the three days surrounding the revelation date				
Value-weighted				
Intercept	0.00052891	0.00011608	4.56	<.0001
Event	-0.00033379	0.00031751	-1.05	0.2932
Equal-weighted				
Intercept	0.00083902	0.00009647	8.70	<.0001
Event	-0.00030651	0.00026387	-1.16	0.2454
Return on market-index on indicator for the five days surrounding the revelation date				
Value-weighted				
Intercept	0.00055173	0.00012147	4.54	<.0001
Event	-0.00032297	0.00026583	-1.21	0.2244
Equal-weighted				
Intercept	0.00083851	0.00010095	8.31	<.0001
Event	-0.00019378	0.00022093	-0.88	0.3804

N= 9584 daily market-index returns in the period 1980-2017

Panel B: Estimation of Net Benefits (Costs) Depending on 100 Investigations by Year - AAERs

	Number of Frauds Identified	Total Benefit	Number of False Positives	False Positives Cost (1%)	False Positives Cost (3%)	False Positive Cost (5%)	Net Benefit (Cost) at 1%	Net Benefit (Cost) at 3%	Net Benefit (Cost) at 5%
M-Score	14	70,765	3709	23,932	71,796	119,659	46,833	(1,031)	(48,894)
F-Score >1	27	113,465	3773	55,108	165,323	275,538	58,357	(51,858)	(162,073)
F-Score >1.85	21	70,537	1386	50,291	150,872	251,454	20,246	(80,335)	(180,917)
F-Score >2.45	21	70,537	1386	50,291	150,872	251,454	20,246	(80,335)	(180,917)
Chakrabarty et al.	33	139,113	3767	120,269	360,807	601,346	18,844	(221,695)	(462,233)
Cecchini et al.	21	146,159	3649	558,262	1,674,785	2,791,309	(412,103)	(1,528,627)	(2,645,150)
Amiram et al.	37	162,113	3663	67,897	203,691	339,485	94,216	(41,578)	(177,372)
Alawadhi et al.	28	188,995	3770	1,362,436	4,087,309	6,812,182	(1,173,441)	(3,898,314)	(6,623,187)
Bao et al.	23	117,639	3126	210,981	632,944	1,054,907	(93,342)	(515,305)	(937,268)

Panel C: Estimation of Net Benefits (Costs) Depending on 100 Investigations by Year - All Restatements

	Number of Restatements Identified	Total Benefit	Number of False Positives	False Positives Cost (1%)	False Positives Cost (3%)	False Positive Cost (5%)	Net Benefit (Cost) at 1%	Net Benefit (Cost) at 3%	Net Benefit (Cost) at 5%
M-Score	16	121,056	1413	14,391	43,172	71,954	106,665	77,884	49,102
F-Score >1	34	268,553	1348	41,338	124,013	206,689	227,215	144,540	61,864
F-Score >1.85	32	244,106	1281	35,326	105,977	176,629	208,780	138,128	67,476
F-Score >2.45	24	166,506	915	21,908	65,725	109,542	144,597	100,781	56,964
Chakrabarty et al.	56	467,950	1346	83,220	249,659	416,099	384,730	218,290	51,851
Cecchini et al.	28	216,265	1405	322,737	968,211	1,613,684	(106,472)	(751,946)	(1,397,420)
Amiram et al.	28	223,466	1294	43,855	131,566	219,276	179,611	91,901	4,190
Alawadhi et al.	26	229,108	1511	848,049	2,544,148	4,240,247	(618,941)	(2,315,040)	(4,011,139)
Bao et al.	36	271,044	1274	149,844	449,532	749,219	121,200	(178,488)	(478,175)

Panel D: Estimation of Net Benefits (Costs) Depending on 100 Investigations by Year - Severe Restatements

	Number of Restatements Identified	Total Benefit	Number of False Positives	False Positives Cost (1%)	False Positives Cost (3%)	False Positive Cost (5%)	Net Benefit (Cost) at 1%	Net Benefit (Cost) at 3%	Net Benefit (Cost) at 5%
M-Score	16	124,363	1462	14,708	44,125	73,542	109,655	80,238	50,821
F-Score >1	30	246,102	1416	39,660	118,980	198,300	206,442	127,122	47,802
F-Score >1.85	28	221,655	1369	36,360	109,080	181,799	185,295	112,575	39,855
F-Score >2.45	18	131,064	923	21,966	65,898	109,829	109,098	65,166	21,235
Chakrabarty et al.	52	437,728	1424	90,322	270,966	451,609	347,406	166,762	(13,882)
Cecchini et al.	22	171,230	1454	328,173	984,520	1,640,867	(156,943)	(813,290)	(1,469,637)
Amiram et al.	24	202,234	1368	46,187	138,562	230,936	156,046	63,672	(28,703)
Alawadhi et al.	24	213,922	1564	864,200	2,592,600	4,321,000	(650,278)	(2,378,678)	(4,107,078)
Bao et al.	28	216,843	1382	160,082	480,247	800,412	56,761	(263,404)	(583,569)

This table reports the results of analyzing the cost of fraud prediction errors to regulators for AAERs, Restatements, and Severe Restatements. Regulators' true positive benefits are estimated using a regression of market-wide stock returns on an indicator variable for whether there is an AAER announcement, using one-, three-, and five-day event windows. Regulators' true positive benefits are then the regression estimates of the slope times the value of the market on days -2 relative to the day of revelation. Regulators' false positive costs consist of two components: an estimate of the number of regulators' misreporting investigations and an estimate of the market value loss for false positive firms. Using Freedom of Information Act (FOIA) requests, a number of studies estimate that the number of misreporting investigations ranges from 50 to 100 per year (Blackburne et al., 2020; Bonsall et al. 2021; Holzman et al. 2021). For this reason, when estimating regulators false positive costs, we limit the number of flagged firms investigated by the SEC to the top 100 ranks for each model in a given year. We consider false positive costs as the loss in market value for the falsely flagged firms and we use estimates that range from 1% to 15%. We consider 1% as a lower bound on account of evidence in Coleman and al. (2020) that SEC denials of FOIA requests due to ongoing enforcement proceedings are predictive of SEC investigations and are associated with negative future return performance. If regulators make the investigations public, we assume that falsely identified firms experience a market value loss in the range of 3% to 15%, which is the typical market reaction to comment letters and announcements of SEC investigations and charges. As all models become value-reducing when estimated market value losses exceed 5%, the table is restricted to instances in which the estimated market value loss is 5% or less. **Data are in millions of 2016 dollars.**

TABLE 7

Cross-Sectional Analysis of Net Benefits to Investors

	Group	Net Benefit VW	Net Benefit SZ	Net Benefit VW-Win	Net Benefit SZ-Win		Group	Net Benefit VW	Net Benefit SZ	Net Benefit VW-Win	Net Benefit SZ-Win
Size	Quin. 1	-$92	$159	-$92	$159	**CFO**	Quin, 1	$234,451	$250,819	$204,862	$221,649
	Quin. 2	-$3,606	-$3,728	-$3,606	-$3,728		Quin, 2	$284,934	$301,492	$61,870	$87,983
	Quin. 3	-$30,700	-$23,558	-$30,589	-$23,451		Quin, 3	$152,187	$204,304	-$47,047	$5,524
	Quin. 4	-$111,972	-$17,604	-$96,284	-$2,396		Quin, 4	-$557,979	-$483,281	-$290,693	-$213,428
	Quin. 5	-$24,286	-$40,421	$40,695	$143,703		Quin, 5	-$278,044	-$354,706	-$34,692	-$7,620
Life Cycle	Intro	$447,023	$458,243	$284,179	$295,474	**CFI**	Quin, 1	$576,114	$549,576	$419,868	$453,197
	Growth	$647,669	$641,653	$258,336	$307,106		Quin, 2	-$74,737	-$91,142	-$23,874	$16,590
	Mature	-$1,331,338	-$1,268,661	-$614,189	-$499,557		Quin, 3	-$222,964	-$157,122	-$244,628	-$187,427
	Shakeout	$111,839	$119,623	-$10,965	$4,840		Quin, 4	-$420,201	-$381,688	-$202,246	-$160,888
	Decline	-$40,052	-$32,491	-$23,470	-$14,015		Quin, 5	-$22,051	-$408	-$54,208	-$26,775
Lreturn	Quin. 1	-$21,523	-$32,030	-$48,970	-$36,731	**Idiorisk**	Quin, 1	-$764,482	-$725,522	-$545,706	-$462,473
	Quin. 2	-$457,331	-$440,665	-$107,794	-$75,699		Quin, 2	$435,404	$426,439	$39,137	$90,191
	Quin. 3	-$40,496	-$20,697	-$164,754	-$122,545		Quin, 3	-$246,543	-$222,486	$122,751	$153,141
	Quin. 4	-$449,097	-$387,156	-$113,023	-$37,789		Quin, 4	$214,497	$229,312	$187,335	$206,339
	Quin. 5	$702,030	$701,738	$231,350	$275,139		Quin, 5	$135,093	$154,907	$95,006	$115,610
Accruals	Quin. 1	$227,058	$233,833	$137,001	$171,012	**Dedown**	Quin, 1	$927,066	$978,127	$355,137	$467,513
	Quin. 2	-$495,552	-$466,849	-$182,460	-$137,922		Quin, 2	-$172,209	-$160,315	-$25,026	$4,136
	Quin. 3	-$278,751	-$254,340	-$196,548	-$131,086		Quin, 3	-$43,022	-$38,429	-$48,688	-$25,111
	Quin. 4	-$120,682	-$111,898	-$168,001	-$126,464		Quin, 4	-$381,445	-$357,282	-$115,231	-$101,677
	Quin. 5	$700,919	$715,778	$385,004	$405,959		Quin, 5	-$504,688	-$499,289	-$266,799	-$237,348
Sales Growth	Quin. 1	-$20,493	$7,132	-$2,386	$24,711	**Traown**	Quin, 1	$320,447	$287,123	$32,381	$80,265
	Quin. 2	-$365,261	-$288,340	-$205,832	-$138,845		Quin, 2	-$508,254	-$503,397	-$176,065	-$165,034
	Quin. 3	-$653,530	-$643,648	-$477,957	-$420,809		Quin, 3	$5,404	$23,352	-$119,742	-$95,843
	Quin. 4	$192,518	$167,115	$353	$25,111		Quin, 4	-$35,747	$9,191	$2,814	$57,039
	Quin. 5	$676,367	$672,847	$596,201	$624,377		Quin, 5	$43,853	$106,543	$160,005	$231,085

This table reports the net benefits of fraud prediction models to investors across various firm characteristics. Net benefits are calculated as the equal-weighted average of the net benefits of the seven fraud prediction models and three F-Score cutoffs for each firm characteristic. Size is calculated as the market value of equity as the end of the fiscal year. Life cycle is measured following the life cycle classification in Dickinson (2011). Lreturn is the market-adjusted abnormal return in the previous fiscal year. Accruals are calculated as net income less cash flow from operations, scaled by lagged total assets. Sales Growth is the percentage sales growth over the current fiscal year. CFO is cash flow from operations scaled by lagged total assets. CFI is cash flow from investing activities scaled by lagged total assets. Idiorisk is the standard deviation of daily returns measured over the 252 daily return observations in the fiscal year. Dedown is the percentage of shares owned by dedicated institutional owners at the end of the fiscal year. Traown is the percentage of shares owned by transient institutions at the end of the fiscal year. Except for Life cycle firms are grouped in quintiles that are calculated separately for every fiscal year. Net benefits are calculated as described in Table 5 using both market- and size-adjusted returns to measure investors' false positive costs. In case of winsorized returns, variables are winsorized at the 1% and 99% levels, respectively, **Data are in millions of 2016 dollars.**

TABLE 8

Cross-Sectional Analysis of Net Benefits to Auditors

	Group	All benefits less all costs	Benefits excluding reputation less all costs		Group	All benefits less all costs	Benefits excluding reputation less all costs
Size	Quin. 1	$1,027.49	-$444.38	**CFO**	Quin. 1	-$413.28	-$13,208.25
	Quin. 2	-$18,560.38	-$18,668.00		Quin. 2	-$10,522.35	-$10,522.35
	Quin. 3	-$6,180.95	-$6,950.12		Quin. 3	$15,351.97	-$9,266.71
	Quin. 4	-$8,999.14	-$9,853.56		Quin. 4	-$4,667.86	-$7,766.82
	Quin. 5	-$27,869.49	-$35,264.81		Quin. 5	-$5,856.58	-$21,830.46
Life Cycle	Intro	-$1,440.78	-$2,210.07	**CFI**	Quin. 1	$4,537.11	$1,259.58
	Growth	-$11,345.73	-$16,391.81		Quin. 2	-$16,004.83	-$17,823.00
	Mature	-$13,520.74	-$23,018.82		Quin. 3	-$4,995.41	-$6,781.74
	Shakeout	$2,109.60	-$3,090.17		Quin. 4	-$9,524.91	-$10,555.41
	Decline	-$16,047.31	-$17,865.40		Quin. 5	-$33,253.73	-$37,349.18
Lreturn	Quin. 1	-$3,072.85	-$3,726.29	**Idiorisk**	Quin. 1	-$8,539.33	-$22,345.68
	Quin. 2	$3,274.49	-$9,061.62		Quin. 2	-$7,069.09	-$13,927.35
	Quin. 3	-$3,315.71	-$12,257.62		Quin. 3	-$2,969.25	-$8,465.66
	Quin. 4	-$10,574.67	-$13,502.46		Quin. 4	-$3,741.03	-$5,210.42
	Quin. 5	-$17,099.55	-$31,440.46		Quin. 5	-$19,533.22	-$21,158.36
Accruals	Quin. 1	-$5,491.12	-$7,728.08	**Dedown**	Quin. 1	-$8,224.40	-$19,795.36
	Quin. 2	-$459.24	-$10,064.45		Quin. 2	$2,782.84	-$9,298.19
	Quin. 3	-$8,494.27	-$12,648.77		Quin. 3	-$21,093.54	-$21,105.45
	Quin. 4	-$3,259.69	-$13,308.55		Quin. 4	-$5,457.69	-$6,279.20
	Quin. 5	-$20,971.07	-$26,732.88		Quin. 5	-$12,993.69	-$12,993.69
Sales Growth	Quin. 1	-$4,434.08	-$4,434.08	**Traown**	Quin. 1	-$19,739.35	-$21,392.01
	Quin. 2	-$7,085.88	-$10,449.21		Quin. 2	-$9,575.76	-$9,575.76
	Quin. 3	-$27,720.18	-$30,256.61		Quin. 3	-$18,901.30	-$20,803.78
	Quin. 4	-$10,718.12	-$12,278.54		Quin. 4	$12,926.08	-$5,435.50
	Quin. 5	$25,745.05	-$13,740.20		Quin. 5	$7,808.82	-$12,264.83

This table reports the net benefits of fraud prediction models to auditors across various firm characteristics. Net benefits are calculated as the equal-weighted average of the net benefits of the seven fraud prediction models and three F-Score cutoffs for each firm characteristic. Size is calculated as the market value of equity as the end of the fiscal year. Life cycle is measured following the life cycle classification in Dickinson (2011). Lreturn is the market-adjusted abnormal return in the previous fiscal year. Accruals are calculated as net income less cash flow from operations, scaled by lagged total assets. Sales Growth is the percentage sales growth over the current fiscal year. CFO is cash flow from operations scaled by lagged total assets. CFI is cash flow from investing activities scaled by lagged total assets. Idiorisk is the standard deviation of daily returns measured over the 252 daily return observations in the fiscal year. Dedown is the percentage of shares owned by dedicated institutional owners at the end of the fiscal year. Traown is the percentage of shares owned by transient institutions at the end of the fiscal year. Except for Life cycle firms are grouped in quintiles that are calculated separately for every fiscal year. Net benefits are calculated as described in Table 4. **Data are in millions of 2016 dollars.**

TABLE 9

Cross-Sectional Analysis of Traditional Model Performance Measures

	Group	% Frauds	Sensitivity (Mean)	Precision (Mean)	FP Rate (Mean)		Group	% Frauds	Sensitivity (Mean)	Precision (Mean)	FP Rate (Mean)
Size	Quin. 1	8.63%	38.35%	0.24%	19.20%	**CFO**	Quin. 1	24.50%	54.51%	0.65%	29.19%
	Quin. 2	14.60%	42.83%	0.33%	25.64%		Quin. 2	20.58%	52.28%	0.48%	31.76%
	Quin. 3	20.23%	50.40%	0.50%	30.58%		Quin. 3	24.54%	51.61%	0.68%	31.09%
	Quin. 4	29.96%	54.73%	0.80%	34.31%		Quin. 4	16.85%	43.59%	0.30%	30.71%
	Quin. 5	26.58%	53.97%	0.68%	36.24%		Quin. 5	13.54%	43.18%	0.49%	27.18%
Life Cycle	Introduction	7.22%	38.58%	0.60%	21.22%	**CFI**	Quin. 1	34.42%	55.03%	0.82%	34.30%
	Growth	42.84%	50.89%	0.67%	33.13%		Quin. 2	17.79%	52.52%	0.49%	30.44%
	Mature	21.40%	61.99%	0.75%	35.43%		Quin. 3	19.17%	47.72%	0.47%	30.72%
	Shakeout	22.95%	43.64%	0.23%	28.94%		Quin. 4	14.35%	47.20%	0.26%	30.90%
	Decline	5.59%	35.68%	0.22%	23.55%		Quin. 5	14.27%	41.21%	0.38%	23.57%
Lreturn	Quin. 1	14.70%	38.01%	0.27%	22.46%	**Idiorisk**	Quin. 1	16.29%	48.59%	0.45%	30.89%
	Quin. 2	14.62%	48.53%	0.33%	27.66%		Quin. 2	22.42%	51.71%	0.57%	32.82%
	Quin. 3	16.63%	51.52%	0.41%	29.70%		Quin. 3	24.29%	54.61%	0.63%	31.44%
	Quin. 4	20.10%	53.14%	0.46%	31.29%		Quin. 4	25.39%	48.99%	0.59%	28.61%
	Quin. 5	33.95%	51.97%	0.66%	33.47%		Quin. 5	11.61%	45.37%	0.32%	22.12%
Accruals	Quin. 1	15.22%	46.34%	0.54%	21.18%	**Dedown**	Quin. 1	29.40%	52.33%	0.78%	34.03%
	Quin. 2	11.92%	41.76%	0.37%	24.04%		Quin. 2	21.77%	52.09%	0.55%	30.76%
	Quin. 3	14.07%	37.68%	0.28%	27.25%		Quin. 3	15.76%	47.32%	0.40%	26.08%
	Quin. 4	20.03%	47.50%	0.38%	31.70%		Quin. 4	11.06%	52.21%	0.36%	24.65%
	Quin. 5	38.75%	61.85%	0.73%	42.05%		Quin. 5	22.00%	48.62%	0.52%	30.69%
Growth	Quin. 1	11.04%	40.48%	0.59%	19.55%	**Traown**	Quin. 1	14.76%	51.67%	0.46%	28.30%
	Quin. 2	15.86%	44.04%	0.29%	26.18%		Quin. 2	15.61%	46.56%	0.40%	26.16%
	Quin. 3	14.57%	50.14%	0.44%	30.00%		Quin. 3	17.61%	44.45%	0.40%	27.32%
	Quin. 4	20.68%	48.09%	0.42%	32.51%		Quin. 4	18.84%	51.71%	0.52%	29.64%
	Quin. 5	37.86%	58.05%	0.72%	37.74%		Quin. 5	33.18%	54.96%	0.78%	34.80%

This table reports the analysis of traditional model performance measures across various firm characteristics. Measures are calculated as the equal-weighted average of the seven fraud prediction models and three F-Score cutoffs for each firm characteristic. Size is calculated as the market value of equity as the end of the fiscal year. Life cycle is measured following the life cycle classification in Dickinson (2011). Lreturn is the market-adjusted abnormal return in the previous fiscal year. Accruals are calculated as net income less cash flow from operations, scaled by lagged total assets. Sales Growth is the percentage sales growth over the current fiscal year. CFO is cash flow from operations scaled by lagged total assets. CFI is cash flow from investing activities scaled by lagged total assets. Idiorisk is the standard deviation of daily returns measured over the 252 daily return observations in the fiscal year. Dedown is the percentage of shares owned by dedicated institutional owners at the end of the fiscal year. Traown is the percentage of shares owned by transient institutions at the end of the fiscal year. Except for Life cycle firms are grouped in quintiles that are calculated separately for every fiscal year. Measures are calculated as described in Table 2.