How Test Power Impacts Research Relevance: The Case of Earnings Management Research

Zhuoan Feng
UTS Business School
University of Technology Sydney

Yaowen Shan
UTS Business School
University of Technology Sydney

Stephen Taylor*
UTS Business School
University of Technology Sydney

August 2017

We appreciate the financial support of the Accounting and Audit Quality Research program funded by the Capital Markets Co-operative Research Centre (CMCRC Ltd), a center funded by the Federal Government of Australia. We appreciate the comments of Jean Bedard, Philip Brown, Pete Clarkson, Greg Clinch, Jeff Coulton, Gerry Gallery, Alister Hunt, Andrew Jackson, Kin-Wai Lee, Terry Walter and Jilnaught Wong, as well as attendees at the 2017 AFAANZ Conference and seminar participants at the following universities: Auckland, New South Wales, Queensland University of Technology and University of Technology Sydney. We especially acknowledge the role of Luke Bayley in helping identify our abnormal sales index measure. Early versions of this paper were circulated as “Identifying earnings overstatements: A practical test.”

* Corresponding author
Stephen Taylor
School of Accounting
University of Technology Sydney
PO Box 123 Broadway
NSW 2007 Australia
Stephen.Taylor@uts.edu.au

Electronic copy available at: https://ssrn.com/abstract=3015043

How Test Power Impacts Research Relevance: The Case of Earnings Management Research

Abstract

We argue that the broader applicability of accounting research is often limited by the way accounting researchers typically place far greater weight on the relative cost of type I versus type II errors. To illustrate the extent of this problem, we examine the performance of simple financial ratio-type analysis for detecting earnings overstatements when the total misclassification costs are minimized subject to the relative cost of type I versus type II errors. We then contrast the likelihood of type I versus type II errors from this approach with those arising from several widely used measures of unexpected accruals. The results demonstrate how commonly-used unexpected accruals measures reduce the type I error rate by sacrificing the type II error rate. Given that accounting information users and auditors typically face much higher costs of type II errors, we explicitly identify why unexpected accruals models are likely far less useful in detecting earnings overstatements than a relatively simple approach using financial statement analysis red flags. Our results highlight the fundamentally contrasting incentives facing accounting researchers relative to those who might otherwise use the research in practice, and serve as a warning when the broader relevance of accounting research is increasingly under question.

Electronic copy available at: https://ssrn.com/abstract=3015043

1. Introduction

The relevance of accounting research beyond the academic community has been frequently questioned. In some countries, broad national assessments of research relevance and/or researchers' engagement with end-users is either already undertaken, or proposed.[1] Hence, we expect that accounting researchers will face increasing pressure to demonstrate how their research has impact beyond the academy. Our purpose is to highlight how a fundamental concern for researchers (avoiding type I errors) potentially works against the production of accounting research with broad external relevance. We use the extensive literature directed towards identifying instances of earnings management to demonstrate this dilemma.

There are many reasons advanced for the limited broader relevance of accounting research. While some critics focus on the choice of research questions (Kaplan 2011; Dyckman and Zeff 2015), others argue that there are significant barriers to the successful transmission of knowledge created by academic research to end users (Hoang et al. 2017). Still, others argue that the relevance of research is limited by researchers' obsession with demonstrating "statistical significance" (Ohlson 2015). While these are all important considerations, we argue that a fundamental restriction on the broader relevance of accounting research arises from the extremely high weighting given by researchers to the costs of type I errors (i.e., falsely rejecting the null hypothesis). Our argument is consistent with Hopwood (2007), who criticizes the "cautious" approach of accounting researchers, and argues that social science journals generally (including accounting) are more concerned with

[1] For example, in the United Kingdom the Research Assessment Exercise includes an explicit requirement to demonstrate impact (typically in the form of case studies). A similar assessment is scheduled for Australian universities in 2018, which will be an extension of the periodic assessment of research quality known as Excellence in Research for Australia (ERA).

Electronic copy available at: https://ssrn.com/abstract=3015043

ensuring the methodological and statistical validity of the findings than the novelty or broader applicability of the research.[2]

We argue that this "cautiousness" reflects researchers' overriding concern with type I errors, namely minimizing the probability of falsely rejecting the null hypothesis. One such example is the development of methods used to detect earnings management. The detection of accounting manipulation is surely a topic of considerable practical interest to regulators, auditors, and investors (Fields et al., 2001). In highlighting the potential impact of accounting research on practice, an American Accounting Association committee specifically cites the extensive literature addressing the identification of earnings management and tests of its causes and consequences (Moehrle et al. 2009). However, we are struck by the near total absence of evidence demonstrating any substantial impact this research has had on practice.[3] Likewise, Ball (2013) argues that the absence of regulatory or prosecutorial action based on such findings is evidence that it has little external relevance. Our study is motivated by this same concern.

Despite early evidence that the method suggested by Jones (1991) lacks sufficient statistical power to detect earnings management of "plausible magnitude" (Dechow et al. 1995), a plethora of papers identify various causes and consequences of earnings management (Dechow et al. 2010). This research has widespread and long-standing currency in top-tier accounting journals, and these journals continue to publish incremental refinements in the methods used for this purpose (Collins et al. 2017; Owens et al. 2017). Innovations to models used to estimate unexpected accruals are largely confined to methodological modifications controlling for firm performance

[2] Hopwood (2007) argues that in this respect, social sciences compare unfavourably to the natural sciences.

[3] Although Moehrle et al. (2009) cite earnings management research as an example of accounting research with professional impact, they offer no specific examples or citations of the use of this research in practice.

Electronic copy available at: https://ssrn.com/abstract=3015043

(Kothari et al. 2005), non-linear growth (Collins et al., 2017) and the match between accruals and cash flows from operation (Dechow and Dichev 2002; McNichols 2002).[4] It is striking that the evidence offered for such innovations almost universally focuses on a reduction in type I errors. There is practically no attention given to explicitly improving the power of these models.[5]

Why are accounting researchers so concerned with type I errors? Given the near-absence of top-tier publications that demonstrate an absence of earnings management behavior (i.e., which conclude that the null is true), it would seem obvious that researchers face strong incentives to detect results that reject the null hypothesis.[6] Hence, there is a natural concern that such research may reflect a bias towards rejecting the null, and referees and editors take great care to try and avoid publication of papers where the results can be subsequently shown to reflect a type I error. One way researchers address this concern is by using methods which are recognized as being most appropriate for reducing the likelihood of type I errors.

However, at the practical level, regulators, auditors, and accounting information users are likely far more concerned with the power of methods to detect earnings management (type II errors) than they are with wrongly concluding that some firms have engaged in earnings management (type I errors). In fact, we expect that most practical concerns about earnings management revolve around the overstatement of earnings, and those concerned are far more concerned with minimizing the costs of

[4] For example, Dechow et al. (1995) outline several possible extensions to the Jones model. Kothari et al. (2005) subsequently recommend this approach be implemented via a performance-matched control sample. Most recently, Collins et al. (2017) propose further refinements to Jones-type models that deal with non-linear growth and performance effects.

[5] A notable exception is Dechow et al. (2012), who focus on improved power to detect earnings management when the expected reversal effect is included in the model of expected accruals. However, while they document substantial power improvement, their method requires ex ante specification of the reversal period, which Gerakos (2012) identifies as both theoretically difficult and impossible to implement in real-time surveillance settings.

[6] A possible exception arises when examining alleged effects of regulatory intervention, or even the extent to which an effect used as the basis for justifying regulatory intervention is actually evident. Examples include the effects on accounting quality of the provision of non-audit services

Electronic copy available at: https://ssrn.com/abstract=3015043

type II errors. An obvious example is the auditing profession. It is extremely rare to find an auditor subject to litigation resulting from the understatement of earnings; just as regulatory actions are similarly rare. For example, investors rarely sue auditors over earnings understatements, except where the resulting undervaluation of the firm can be shown to have had a direct economic cost, such as in a management buyout of external stockholders. On the other hand, the cost of an auditor failing to recognize a material earnings overstatement is severe. This is especially so when compared to the costs of some additional analysis which ultimately shows that the "problem" is either minor or otherwise simply not present. Such type I errors likely have a very low cost relative to a type II error. In short, the practical costs of type II errors are significantly higher than those associated with type I errors.

Put simply, the research community and the potential users of the research in practice face fundamentally different incentives with regard to the minimization of total expected error costs. This problem is especially evident when considering the earnings management literature and its (non)relevance to practice. Given the trade-off between type I and type II errors (Sheskin 2003), modification of unexpected accruals models improves model specification and reduces type I errors at the expense of increasing type II errors. For example, the performance-matched approach in Kothari et al. (2005) is effective in mitigating type I errors that arise from the correlation between the accrual model residual and firms' performance. However, the performance-matched approach based on *ROA* detects upwards earnings management equivalent to 1%, 2% and 4% of assets, at a rate of 12.8%, 26.8%, and 60.0% respectively.[7] This is substantially lower than that reported by Kothari et al. (2005) for

[7] To put these figures in perspective, note that for the Compustat population as a whole, net income divided by total assets is around 4%. Hence, unexpected accruals equal to 4% of total assets is actually equivalent to total income, on average. Such massive earnings management would surely be self-evident.

Electronic copy available at: https://ssrn.com/abstract=3015043

the modified Jones model, which detects these positive unexpected accruals at a rate of 21.2%, 38.0%, and 88.4% respectively.[8] Perhaps ironically, both sets of results reinforce the earlier conclusion by Dechow et al. (1995) that unexpected accruals models "lack power in detecting earnings management of plausible magnitudes".

In this study, we contrast the power of a simple financial ratio-type analysis for identifying instances of significant earnings overstatements, with analysis based on several widely-used measures of unexpected accruals. We develop a measure (which we label as an EM-score) based on simple accruals, supplemented by several red flag variables derived from financial statement analysis techniques. Importantly, our study considers the relative cost of type I and type II errors, and conducts a direct comparison between the EM-score and unexpected accruals measures when the total expected misclassification costs of type I and type II errors are minimized. Our approach is intended to highlight just how important the relative costs of these errors are in determining the practical usefulness of the different approaches to identifying earnings overstatements.

Our primary concern is not the development of an improved, let alone novel method for detecting earnings overstatements. Beneish (1999) has previously demonstrated the detection of a small sample of GAAP violations disclosed in Accounting and Auditing Enforcement Releases (AAERs) using red flag variables. Likewise, others have shown the ability of a broad selection of financial ratios to detect earnings management. For example, Dechow et al. (2011) develop an F-score model based on accruals quality, financial performance, non-financial performance, off-balance-sheet activates, and market-related variables. The F-score model correctly classifies about 64% of the sample firms when detecting earnings overstatements with

[8] See Table 4 of Kothari et al. (2005).

Electronic copy available at: https://ssrn.com/abstract=3015043

an average type I and type II error rates of 36% and 32 % respectively. While Dechow et al. (2011) (figure 2) show how the trade-off between type I and type II errors can be considered by accounting information users, they do not directly compare the detection power of their F-score with "standard" unexpected accruals measures, let alone conduct a comparison when the total misclassification costs are minimized. Jansen et al. (2012) propose a new diagnostic for earnings overstatements based on discretionary changes in a firm's profit margin and asset turnover ratio. However, the resulting measure is only able to correctly classify about 20% of earnings overstatements, implying a type II error rate of 80%. Moreover, the authors do not report any comparison between the power of their new diagnostic for identifying earnings overstatements with unexpected accruals measures. Nor do they consider the relative costs of type I and type II errors.

In contrast to these studies, our fundamental concern is the trade-off between type I and type II errors, as highlighted by the extent to which our simple EM-score based on ratio analysis outperforms various unexpected accrual models when the total misclassification cost is minimized. Relative to unexpected accrual models that have been a mainstay of financial accounting research, we demonstrate the superiority of using simple red flags based on financial statement analysis for detecting earnings overstatements when realistic trade-offs between the costs of type I and type II errors are considered, and the misclassification cost is minimized by identifying the optimal cut-off point.

We calibrate our EM-score model based on a comparison of firms subject to AAERs identifying earnings overstatements and a set of control firms. Firms subject to AAERs are expected to reflect a very high likelihood of relatively substantial

Electronic copy available at: https://ssrn.com/abstract=3015043

earnings overstatement (Dechow et al. 1995; 1996; 2011).[9] Indeed, the AAERs typically reflect earnings manipulation outside the boundaries of GAAP. We calibrate a model that distinguishes between AAER cases and a set of control firms, which is also relatively robust to a plausible range of assumed values for the cost of type I and type II errors. In particular, we first match each upwards earnings manipulator with five control firms, selected based on industry membership, firm size, and time-period. Then, we estimate the EM-score model using a logistic regression of the indicator of AAERs on accruals and a set of red flag variables including changes in sales, the divergence between accruals and cash flows, inventory changes, changes in bad debts reserves and changes in asset quality. Our analysis is therefore focused on indicators that are expected to reflect attempts to manage earnings beyond the boundaries of GAAP, particularly those that result in the overstatement of earnings. Finally, we use the predicted probability derived from the EM-score model as the EM-score for each observation.

We then conduct our EM-score classification analysis based on the optimal cut-off point, namely the cut-off EM-score. The cut-off EM-score is determined by identifying where the total expected misclassification costs are minimized, namely the sum of the costs associated with type I and type II errors. For this purpose, we assume the prior probability of fraudulent financial reporting is in the 1-4% range, and assume the relative cost of type I and type II errors ranges from 10:1 to 50:1. The expected cost of misclassification is a function of the prior probability of fraudulent financial accounting, the prior probability of non-fraudulent financial accounting, the observed type I and type II error rates, and the relative cost of type I and type II errors.

[9] However, despite being instances of relatively extreme earnings management, the extent of the accounting manipulation does not appear to be fully anticipated prior to the announcement of SEC action (Feroz et al. 1991; Dechow et al. 1996; Dechow et al. 2012; Files 2012)

Electronic copy available at: https://ssrn.com/abstract=3015043

Following our identification of optimal EM-scores, we then similarly examine the power of several commonly-used unexpected accrual measures using the same sample of AAER firm-years. The unexpected accrual measures we consider include the modified Jones model (Dechow et al. 1995), the modified Dechow and Dichev (2002) approach suggested by McNichols (2002) (hereafter modified DD model) and the performance-matched model based on *ROA* (Kothari et al. 2005). To compare the power of our EM-score model with these unexpected accruals measures, we first compare their marginal effects for identifying earnings overstatements using logistic regressions. The marginal analysis shows that, in general, the power of our EM-score model is about five times that of typical unexpected accruals, or a simple total accruals measure.

Subsequently, we choose the cut-off point for unexpected accruals when the total expected misclassification costs are minimized. We thus directly compare the power of unexpected accruals measures to our EM–score model. We find that, when the misclassification costs are minimized, unexpected accruals measures can correctly classify on average 90% of the sample with a type I error of 7% and type II error of 84%. Most importantly, our results display a consistent pattern whereby the high overall accuracy of unexpected accruals models primarily reflects a significantly lower rate of type I errors. In contrast, unexpected accruals are only able to separate the most extreme examples of manipulation, such that the type II error rate for the unexpected accruals measures typically exceeds 75%, while those for our EM-score are consistently around 25% or less. These results therefore reinforce the lesson that commonly-used unexpected accruals measures reduce the type I error rate at the expense of increased type II errors. Importantly, we also find that the unexpected

Electronic copy available at: https://ssrn.com/abstract=3015043

accrual measures do not perform significantly better than a simple measure of total accruals.

Overall, the results support our view that academic research directed at detecting earnings management using extensions of the Jones (1991) model gives insufficient weight to the issue of greatest practical concern, namely the power of these methods to detect substantial earnings overstatements. While the focus of researchers on minimizing type I errors is entirely understandable given their relatively high cost to the researchers, reviewers and editors concerned, our results also highlight a fundamental tension between the academic "rigor" of this research and its practical relevance. Although Dechow et al. (1995) clearly identified how the Jones (1991) model lacked power, researchers have continued to place far greater weight on their findings and those subsequent (e.g., Kothari et al. 2005) addressing how type I errors can be minimized.

Our study makes two important contributions. First, the approach we take is in stark contrast to recent attempts at further enhancing extant measures of unexpected accruals, either by the addition of further explanatory variables (Dechow and Dichev 2002; McNichols 2002; Collins et al. 2017) or via the use of matching procedures (Kothari et al. 2005). In many respects, our approach is closer in spirit to studies that have concentrated on a specific accrual adjustment.[10] Unlike many prior studies that examine firms subject to AAERs, our concern is with the power of methods used to identify earnings manipulation, rather than the causes or consequences (Beneish 1999; Dechow et al. 1996, 2011). Our results highlight how the assumed cost of type I

[10] Examples where a single accrual adjustment is examined include Miller and Skinner (1998), Marquardt and Wiedman (2004), Petroni (1992) and Beaver and McNicholls (1998). Relatedly, Bowen et al. (2002) and Davis (2002) examine the use of grossed-up revenue and barter, and Rasmussen (2013) examines the implications of revenue recognition methods for earnings management and earnings informativeness, but only for a specific group of firms (i.e., internet firms or the semiconductor industry).

Electronic copy available at: https://ssrn.com/abstract=3015043

versus type II errors substantially impacts how such research is conducted, and particularly the relevance of the research beyond the academy.

Second, our study contributes to the extent earnings management literature, especially at the practical level, by directly comparing the performance of our EM-score and other commonly used unexpected accruals models. In contrast to prior research, we compare the performance of these measures after explicitly considering the relative cost of type I and type II errors and hence, the need to minimize overall misclassification costs. Our results demonstrate that, when misclassification costs are minimized, unexpected accruals models reduce type I error rates at the expense of type II errors. Since accounting information users and auditors likely face drastically higher costs associated with type II errors relative to type I errors, unexpected accruals models are inevitably far less useful than a simple financial statement analysis approach.

The remainder of this paper proceeds as follows. The next section describes the estimation of our EM-score, while the results of our attempt to distinguish between AAER and control firms using the EM-score are summarized in section three. In the fourth section, we consider the relative ability of several unexpected accrual measures to distinguish between the same two groups of firms. The fifth section discusses additional tests for the robustness check, while section six concludes.

2. EM-score estimation

The essence of our approach to identifying instances of earnings overstatements is to supplement a simple measure of operating accruals with some relatively straightforward financial ratio-type analysis. In this section, we discuss the motivation and the selection of the financial ratio-type variables that we hypothesize to be

Electronic copy available at: https://ssrn.com/abstract=3015043

associated with upwards earnings manipulations. Each financial ratio-type variable can be viewed as a red flag, with the overall result an earnings management score (EM-score). The variables employed in our model include a selection of financial statement ratios directed at the detection of either premature revenue recognition or increased cost deferral. These variables are commonly used in practice (Melumad and Nissim 2009). We choose these variables based on a trade-off between comprehensiveness and our primary interest being to highlight the relative lack of power for detecting earnings overstatements of accruals-based measures used to identify earnings management.[11] Most financial ratios are constructed so as to identify time series variation consistent with the financial statement effects of earnings overstatements.

However, time series changes provide no indication of the absolute level of aggression or the significance of past distortions. To the extent that past distortions are significant, manipulation may be better detected by comparing differences in the magnitudes of certain financial statement ratios relative to comparable firms. In many cases, however, peer firm differences in financial statement ratios are an unreliable indicator of manipulation, often-reflecting firm and industry specific factors. Accordingly, the five financial ratio-type variables which we combine with operating accruals to form a model to predict the likelihood of earnings overstatements (i.e., to form an EM-score) all capture the time series change from the year before the accounting fraud (i.e., the most recent 10-K filing) to the first year of GAAP violation.[12] Briefly, the five financial ratio type variables and the accruals variable which we consider are as follows:

---

[11] We have also considered the financial ratio variables used in Beneish (1999) and Dechow et al. (2011), and chosen the variables that are most commonly used in practice and which have significant differences between upwards earnings manipulators and non-manipulators.
[12] See the notes to Table 2 for an abbreviated definition of each variable.

Electronic copy available at: https://ssrn.com/abstract=3015043

Operating accrual magnitude (ACC): *ACC* is the value of operating (i.e., net working capital) accruals deflated by total assets in the year of GAAP violation. Operating accruals are measured as net income before extraordinary items plus depreciation and amortization expense less cash flow from operations. Higher values of *ACC* are expected to be associated with a greater likelihood of earnings overstatement.

Sales index (SLSI): *SLSI* is computed as the ratio of reported net revenue relative to a notional estimate of unmanipulated net revenue. Stubben (2010) finds that revenue models are less biased, better specified, and more powerful than commonly used accrual models. To estimate unmanipulated net revenue, the time series change in the ratio of accounts receivable to net revenue is examined. As detailed in Appendix 2, the estimate for unmanipulated net revenue assumes that all times series changes in the value of this ratio are the result of manipulation, and measures its effect accordingly. An *SLSI* value in excess of one is assumed to reflect aggression (or an increase in aggression) in the firm's revenue recognition policy.

Accruals index (ACCI): A divergence between earnings and cash flow is often identified by practitioners as a prime indicator of earnings manipulation (Schilit 2010). *ACCI* is measured as one plus the current value of operating accruals deflated by the average of current and lagged total assets in the year of GAAP violation, divided by one plus the lagged value of operating accruals scaled by the aforementioned deflator (the average of total assets for the year of GAAP violation and the prior year). This index seeks to capture the time series change in the magnitude of total accruals. A value in excess of one reflects a growing divergence between operating earnings and cash flows.

Electronic copy available at: https://ssrn.com/abstract=3015043

Inventory index (INVI): Management has considerable discretion with respect to the timing of inventory write-offs. Production decisions can also be used to inflate inventory levels and thereby decrease the associated cost of goods sold expense. Evidence highlighting the possible role of inventory accounting techniques as methods to manipulate reported earnings is reported by Marquardt and Weidman (2004), Roychowdhury (2006), Summers and Sweeney (1998) and Zang (2011). *INVI* is measured as one plus the ratio of current inventory to net revenue, all deflated by one plus the ratio of lagged inventory to lagged net revenue. An index value in excess of one is expected to be associated with a greater probability of upwards earnings manipulation.

Reserve index (RESI): Considerable subjectivity is involved in the estimation of expense provisions. Audit partners surveyed by Nelson et al. (2002) report that "cookie jar reserves" are the most popular method for manipulating reported earnings. Marquardt and Weidman (2004) and Teoh et al. (1998) report context-specific results consistent with this claim. Accordingly, *RESI* measures the relationship between the reserve for bad debts receivables and the current accounts receivables balance. *RESI* is measured as one plus the ratio of the lagged value of the bad debts reserve relative to the lagged receivables balance, all deflated by one plus the ratio of the current bad debts reserve relative to the current receivables balance. A value in excess of one is consistent with earnings manipulation.[13]

Asset quality index (AQI): An increase in so-called soft asset balances may be indicative of aggressive cost capitalization. Beneish (1999) highlights the significance of soft assets in an analysis of SEC Enforcement Releases. In addition, Dechow et al. (2011) find that, when firms have more soft assets on their balance sheet, there is

[13] Of course, there are many other types of provisions that could be used to manipulate earnings, such as reserves for future health care benefits, and periodic maintenance reserves. However, these require more context-specific analysis than our relatively generalist red flag approach.

Electronic copy available at: https://ssrn.com/abstract=3015043

more discretion for management to change assumptions to meet short-term earnings goals.

In our study, *AQI* refers to one plus the current value of soft assets deflated by the average of the current and lagged value of total assets in the year of GAAP violation, divided by one plus the lagged value of soft assets scaled by the aforementioned deflator (the average value of total assets for the year of GAAP violation and the prior year). Soft assets include the Compustat items “other current assets”, “other non-current assets” and intangibles. Goodwill is excluded from intangibles to remove the distorting effects of merger and acquisition activity. An *AQI* in excess of one may indicate increased cost deferral and a resulting upwards manipulation of earnings.

3. EM-score evidence

Upwards earnings manipulators are identified from the SEC’s AAERs. Although not all AAERs pertain to fraudulent financial reporting, they are the best place to find a fairly complete sample of SEC actions concerning violations of GAAP. AAERs have been previously used to identify samples where earnings manipulation can be reasonably assumed (Feroz et al. 1991; Dechow et al. 1995; Beneish 1999; Dechow et al., 2011; Schrand and Zechman, 2012). We obtain our AAERs dataset from the University of California, Berkeley in 2015. The dataset, dated at 21/10/2014, includes 1,554 AAERs listed on the SEC website between 1989 and 2011. Forty-three non-financial companies were identified as having understated a prior 10-K filing.[14] We exclude observations that have insufficient Compustat data. The final sample yielded 573 earnings manipulators that have overstated their earnings. Each sample firm is

[14] The dataset includes AAERs observations up to 2011. This is because the SEC needs to average three years to identify and investigate the alleged GAAP violations.

Electronic copy available at: https://ssrn.com/abstract=3015043

then matched with five control firms, selected based on industry membership (two digit SIC code), firm size (measured in total assets), and time-period (year of GAAP violation). Control firms are also required to have sufficient data on Compustat. Panel A of Table 1 provides a summary of the sample selection procedure, while Panel B reports summary statistics for key variables of 573 earnings manipulators and 2,865 control firms used in our research.

<INSERT TABLE 1>

We use logit analysis to model the differential financial statement characteristics observed between the two sample cohorts (earnings manipulators and control firms). An earnings management score (EM score) is calculated as:

$$EM = W_1X_1 + W_2X_2 + \ldots\ldots + W_jX_j \qquad (1)$$

where:

EM = the indicator of earnings manipulators, equals 1 if the firm is an earnings manipulator listed in AAER, zero otherwise

X = the j^{th} attribute or independent variable (red flag accounting ratios)

W = the estimated coefficient or weight for the j^{th} attribute

The logit model is estimated with the dependent variable equal to one if the observation is an earnings manipulator, or zero otherwise. The independent variables are the red flag ratios identified above. We subsequently use the predicted probabilities that are derived from the EM-score model (Exp(EM)/(1+Exp(EM)) and scale to the unconditional probability to evaluate several cut-off values to determine

Electronic copy available at: https://ssrn.com/abstract=3015043

how well the variables employed in the model distinguish financial statement distortions resulting from earnings management.

3.1 Univariate statistics

Table 2 reports univariate test for AAERs firms and control firms. Panel A of Table 2 compares our six red flag accounting ratios and firm's performance indicators estimated in the year of GAAP violation (fraud year) and the previous year (control year) between earnings manipulators and non-manipulators. The results of the univariate test in Panel A suggest that earnings manipulators have significantly higher *ACC* (t=3.29), *SLSI* (t=4.65), *ACCI* (t=3.02) and *AQI* (t=4.00) than non-manipulators in the fraud year. In the control year, all these red flag ratios, except *ACCI*, are still significantly higher in earnings manipulators, compared with non-manipulators. However, difference in difference tests only show a significant difference in *SLSI* (t=1.87). On the other hand, in Panel B, we are unable to find any significant difference between earnings manipulators and non-manipulators in other performance indicators. These results confirm that our red flag accounting ratios have better predictability to detect the earnings overstatements, compared with other performance indicators.

<INSERT TABLE 2>

3.2 EM-score validity

Table 3 extends the analysis to examine average values for each of the EM-score components between certain EM-score ranges. The frequency distribution shows that most upwards earnings manipulators have high EM-scores. For the most part, higher EM-score values are associated with increases in each of the six red flag accounting

Electronic copy available at: https://ssrn.com/abstract=3015043

ratios. However, the *ACC*, *SLSI*, and *ACCI* variables provide the most contribution. This is particularly the case for higher EM-scores. Overall, the results from the analysis in Table 3 show that each of our red flag variables contributes to increased EM-scores. In addition, the results highlight the usefulness of the EM-score model as a screening tool with practical relevance.

<INSERT TABLE 3>

3.3 EM-score calibration

Table 4 reports univariate and multivariate estimations of the relation between the EM-score components and the likelihood of an earnings overstatement. Univariate logistic regressions are reported as models 1 through 6. As expected, all coefficient estimates are positive. The univariate coefficients and logistic regression statistics for the *ACC* (coefficient=1.265, t = 3.24) and *AQI* (coefficient=1.684, t=4.38) models are significant at the 1% level, while the coefficient on *SLSI* (coefficient=0.749, t=1.92) is significant at the 10% level. However, the estimated coefficients for the *ACCI*, *INVI* and *RESI* variables are insignificant.

Multivariate logistic regressions are presented as models 7 through 11 of Table 4. Model 7 only includes the *ACC* and *SLSI* variables. We find that only the coefficient on *ACC* (coefficient=1.200, t=3.11) is significantly associated with the indicator variable of earnings overstatements. Additional variables are added to the logistic regression in models 8 through to 11. We find that *ACC* outperforms other red flag variables in detecting earnings overstatements, while *SLSI* and *RESI* have marginal power in identifying earning manipulators. Finally, we include all our six red flag variables in the model 11. We find coefficients on *ACC* (coefficient=1.344,

t=2.84), *RESI* (coefficient=1.649, t=1.91) and *AQI* (coefficient=1.523, t=3.84) are positive and significant. These results suggest that the operating accrual magnitude (*ACC*), the relationship between the reserve for bad debts receivables and the current receivables balance (*RESI*) and the soft assets balance (*AQI*) have the most power for detecting earnings overstatements.

<INSERT TABLE 4>

3.4 Classification accuracy of the red flag accounting ratios

As the next step in considering how well the proposed EM-score approach discerns manipulative from non-manipulative financial reporting, the classification accuracy derived from the predicted probabilities of the EM-score model are analyzed according to several optimal cut-off points. The cut-off probability estimate used to measure classification accuracy is determined by finding the point where the expected misclassification costs are minimized. Misclassification costs encompass type I and type II errors. Type I errors occur where the EM-score falsely classifies a company as a manipulator. Type II errors refer to the failure to distinguish earnings manipulators from non-earnings manipulators. As we have already argued, the costs associated with these two error types are likely to differ very substantially in practice. The optimal EM-score to measure classification accuracy is the point where the expected misclassification costs are minimized.

$$EC = q_1(M_{12}/N_1)C_1 + q_2(M_{21}/N_2)C_2 \quad (2)$$

where :

EC =	the expected costs of misclassification
q_1 =	the prior probability of fraudulent financial accounting

Electronic copy available at: https://ssrn.com/abstract=3015043

q_2 =	the prior probability of non-fraudulent financial accounting
M_{12}/N_1 =	observed type I errors relative to the sample of non-earnings manipulators
M_{21}/N_2 =	observed type II errors relative to the sample of earnings manipulators
C_1 =	cost of type I errors
C_2 =	cost of type II errors

Equation (2) requires estimates for the prior probabilities of fraudulent and non-fraudulent financial reporting. However, it is difficult to measure the true proportion of firms that manipulate earnings beyond the levels permissible by GAAP. Even a sample comprising all AAER firms would not be an exhaustive list of all GAAP violators. The SEC is resource constrained, and is unable to screen all financial reporting activity. Further, identifying the precise line that distinguishes fraudulent financial reporting from aggressive (but legitimate) application of GAAP is also difficult.

To be consistent with past empirical research, the prior probability of fraudulent financial reporting is arbitrarily estimated to be in the 1-4% range. The estimated cost of type II errors relative to type I errors is also guided by extant empirical analysis. Beneish (1999) estimates relative error costs from a portfolio perspective, and documents a negative 40% return in the quarter containing the discovery of earnings manipulation. Beneish (1999) uses this as an estimate of the cost resulting from the failure to detect fraudulent financial accounting (type II errors) and compares it to the 1-2% average stock return earned by listed firms (type I errors). Utilizing these

Electronic copy available at: https://ssrn.com/abstract=3015043

approximations, Beneish (1999) estimates the cost of type II errors to be 20-40 times greater than type I errors.[15]

<INSERT TABLE 5>

Table 5 presents results of estimating the EM-score associated with the lowest statistical cost of misclassification. It is apparent that this score is easily affected by the choice of prior probability estimates and the relative error costs of misclassifications. This is particularly the case for high probability estimates of manipulative reporting. Following Beneish (1999), we estimate the relative cost of type I and type II errors is ranging from 10:1 to 50:1. The probability of upwards earnings manipulators from the sample is set from 1% to 4%. Relatively, the probability of non-manipulators is from 99% to 96%. Overall, results show that the EM-score model correctly classifies about 45% of all sample observations with about 25% type II errors and approximate 55% type I errors. This suggests that although the model has some ability to distinguish between sample cohorts, it does so with significant classification error.

4. Analysis of unexpected accruals

We benchmark the power of our EM-score approach with several unexpected accrual measures. We use three methods for estimating expected accruals listed in Dechow et al. (2010). The first method is the modified cross-sectional Jones (1991) model suggested by Dechow et al. (1995). This is estimated as follows:

[15] Of course, this approach measures relative misclassification costs from a shareholder perspective. Other decision makers (such as lenders) may have different objective functions and therefore would assign quite different estimates of relative error costs. However, to take our auditor example in the introduction, the cost of a type II error is expected to be far larger than for a type I error.

Electronic copy available at: https://ssrn.com/abstract=3015043

$$Total\ Accruals_{it} = \alpha + \beta_1(\Delta REV_{it} - \Delta REC_{it}) + \beta_2(PPE_{it}) + \varepsilon_{it} \qquad (3)$$

where *Total Accruals* is the difference between income before extraordinary items and operating cash flows, all deflated by the lagged value of total assets. ΔREV is the change in net revenue divided by the lagged value of total assets. ΔREC is the change in accounts receivables deflated by the lagged value of total assets, and *PPE* is the current value of total property, plant and equipment divided by the lagged value of total assets.

The second method for estimating unexpected accruals uses the modified version of the Dechow and Dichev (2002) model as suggested by McNichols (2002). Unexpected accruals are estimated as:

$$Current\ Accruals_{it} = \alpha + \beta_1 CFO_{t-1} + \beta_2 CFO_t + \beta_3 CFO_t + \beta_4 \Delta Sales + \beta_5(PPE_{it}) + \varepsilon_{it} \qquad (4)$$

where Δ*Sales* is computed as the difference between the one period lagged value of *Sales* and the current period of *Sales,* all deflated by total assets and scaled by 100.

The third approach to estimate unexpected accruals is the modified Jones model matched on *ROA*, following Kothari et al. (2005):

$$DisAcc_t - Matched\ firm's\ DisAcc_t \qquad (5)$$

Electronic copy available at: https://ssrn.com/abstract=3015043

Equations (3) through (5) are each estimated in industry cross section, based on two-digit SIC code and calendar year. Following Dechow et al. (2003), a minimum of ten observations are required. Eleven earnings manipulators were excluded due to the additional data requirements, while three earnings manipulators were excluded from the sample due to an insufficient number of two-digit SIC industry year peer firm observations to compute a valid expected accrual model.

The summary statistics of unexpected accruals measures for AAERs and control firms are reported in Panel A of Table 6. The summary statistics shows that the average unexpected accruals calculated by the modified Jones model, the modified DD model, the modified Jones model matching on *ROA* and the total accruals measure are 0.014, 0.006, -0.006 and -0.058 respectively. Panel B compares these measures between earnings manipulators and non-manipulations. Univariate tests suggest that unexpected accruals are significantly higher in manipulators than non-manipulators. For example, the two-sample *t*-test shows that the average unexpected accruals of manipulators measured by the modified Jones model, the modified DD model and the modified Jones model matching on *ROA* are significantly higher than those from non-manipulators (t=5.92, 5.99 and 4.05 respectively). Tests of differences in medians reported in Panel C consistently show that the median values of unexpected accruals are typically higher for manipulators than non-manipulators (χ^2=14.55, 9.90 and 17.53 respectively). Consistent with the results reported by Dechow et al. (2011), these results provide some evidence of the ability of the unexpected accrual measures to distinguish earnings attributes consistent with aggressive financial reporting.

<INSERT TABLE 6>

Electronic copy available at: https://ssrn.com/abstract=3015043

Since our results suggest both EM-score and unexpected accruals models have the ability to identify instances of earnings manipulation, a natural question is whether the EM-score outperforms unexpected accruals. Thus we regress the indicator variable of earnings overstatements on the *EM-score_dummy* and each unexpected accruals measure. The *EM-score_dummy* equals 1 if the EM-score is equal to or greater than 1.00, otherwise zero.

Results in Panel A of Table 7 report that the coefficients on unexpected accruals calculated by the modified Jones model, the modified DD model, and the modified Jones model matching on *ROA* are 1.703, 3.075, and 0.829 respectively. All of these coefficients associated with the unexpected accruals measures are significant at 1% levels. Similarly, the *EM-score_dummy* is also positively and significantly associated with the presence of an earnings overstatement.

Since our results suggest that the unexpected accruals models have some power to detect earnings manipulators, we subsequently conduct a battery of tests to distinguish the power of our EM-score model from the unexpected accruals models. First, we examine the marginal effects based on the logistic regressions. Marginal effects measure the expected instantaneous change in the dependent variable as a function of a change in a certain explanatory variable, while keeping all the other covariates constant.

We first calculate the discrete change of the indicator variable of earnings overstatements from 0 to 1 (*dy/dx*) for our *EM-score_dummy* and each unexpected accruals measure. We then compute the difference between the lower quartile (37.5 percentile) and the upper quartile (75 percentile) for the *EM-score_dummy* and each unexpected accruals measures. We choose the 37.5 percentile and the 75 percentile

Electronic copy available at: https://ssrn.com/abstract=3015043

because the distribution of the *EM-score_dummy* is concentrated in this range. Finally, we multiply *dy/dx* by the difference between the lower and upper quartiles, and label it as the percentage of increase (*poi*), indicating the variation from 0 to 1 of the earnings overstatement indicator, when the variable of interest moves from its lower quartile value to its upper quartile value.

The results in Panel B of Table 7 show that our EM-score model has significantly greater power to detect instances of earnings overstatements than unexpected accruals, and simple total accruals. In particular, the power of the EM-score model (*poi*=5.72%) is about four times higher than for the modified Jones model (*poi*=1.63%). Similarly, our EM-score model (*poi*=5.00%) is more than three times better than the modified DD model (*poi*=1.73%), and six times stronger than the modified Jones model matching on *ROA* (*poi*=1.15%). For comparison, we also benchmark the power of the EM-score model against the total accruals model, and find that our EM-score model (*poi*=6.60%) is about six times better than the total accruals model (*poi*=0.78%).

<INSERT TABLE 7>

In the second test, we assess the accuracy of unexpected accruals models by using the same approach as in Table 5. We estimate the 2% and 4% prior probability for manipulated earnings for every unexpected accruals model. We then identify the cut-off point for each model that is associated with the lowest overall cost of misclassification.

The results in Table 8 reveal that the modified DD model has the highest total accuracy rate among unexpected accruals measures. Its total accuracy rate

Electronic copy available at: https://ssrn.com/abstract=3015043

consistently exceeds 90%, while the modified Jones model and the modified Jones model matching on *ROA* have total accuracy rates between 80% and 90%. Comparing the type II error rates, the modified Jones model outperforms other unexpected accruals measures, except when the relative cost of type I and type II errors are 40:1 and 50:1, and where the prior probability of manipulated earnings is 4%.

Overall, the results in Table 8 display a consistent pattern. Although all unexpected accrual models have high total accuracy rates, this is largely attributable to the low type I error rates (i.e., the range is from 1.38% to 16.18%). However, the trade-off is that unexpected accrual models have high type II error rates (i.e., the range is from 75.50% to 97.50% type II errors). It is also noteworthy that unexpected accrual measures do not significantly outperform the total accruals measure in detecting earnings overstatements. For example, the overall accuracy rate of the simple total accruals measure ranges from 91.44% to 94.03%, while the modified Jones model has accuracy ranging from 85.68% to 94.65%. These results are also consistent with the evidence provided by Dechow et al. (1995) that the Jones model lacks power.

The results in Table 8 are consistent with researchers focusing on reduction, so far as possible, of the likelihood of making type I errors via the identification and use of "improved" methods for identifying unexpected accruals. In comparison with our EM-score model, the type I error rates from unexpected accruals measures are significantly lower. The type I error rates from the EM-score model range between 51.79% and 58.73%, while type I error rates from unexpected accruals measures range from 1.38% to 16.18%. The higher type I error rates from our EM–score model lead to the relatively lower total accuracy rates, compared with unexpected accruals measures.

Electronic copy available at: https://ssrn.com/abstract=3015043

However, it is equally clear that unexpected accruals models reduce type I error rates by sacrificing the type II error rate. For instance, for the modified Jones model, type II errors range from 75.50% to 93.25%, while the highest type II error rate from our EM score model is only 25%. As unexpected accruals measures are associated with a far higher level of type II errors in detecting earnings overstatements, the results in Table 8 support our perspective that unexpected accruals models of the type that dominate earnings management research are not that useful in practical settings where the failure to identify instances of earnings overstatements has a far higher cost than wrongly concluding that an overstatement has occurred.

<INSERT TABLE 8>

5. Additional tests

We perform three additional tests (untabulated) to assess the robustness of our results.[16] First, we extend our EM-score model to include three additional measures suggested by Dechow et al. (2011). These are the change in receivables (*ch_rec*), the change in inventory (*ch_inv*), and the change in cash sales (*ch_cs*).[17] However, our results are similar, in that we find the indicator for AAERs is still positively associated with *ACC*, *RESI* and *AQI*. The inclusion of these additional measures does not change our conclusion that our simple red flag variables are more powerful in detecting earnings overstatements than measures of unexpected accruals.

Second, we also compare the power of our EM-score model with alternative unexpected accruals measures. We re-estimate unexpected accruals using only current accruals rather than total accruals. The results suggest that our EM-score model is

[16] The tabulated results of the additional tests are available from the authors upon request.
[17] Following Dechow et al. (2011), we calculate *ch_rec* as (ΔAccount Receivable/Average total assets), *ch_inv* as (ΔInventory/Average total assets), *ch_cs* as (Percentage change in cash sales=(Sales-ΔAccount Receivable)).

Electronic copy available at: https://ssrn.com/abstract=3015043

significantly more powerful in detecting earnings overstatements than these alternative unexpected accruals measures. For example, unexpected current accruals calculated by the modified Jones model can correctly classify about 90% of the sample firms, with a type I error of 7% and type II error rate of 81%. Similarly, the type I and type II error rates for the modified DD model and the modified Jones model matching on *ROA* are 4% and 88%, and 11% and 81% respectively. The far higher type II errors rates further confirm our argument that Jones-type models reduce type I errors at the expense of sacrificing type II errors, especially when the total misclassification costs are minimized.

In the third additional test, we set the EM-score equal to 1.00 as the cut-off point, instead of choosing an EM-score that minimizes total misclassification costs (i.e., we implement a "naïve" benchmark approach). An EM-score of 1.00 indicates that the firm has the same probability of earnings overstatement as the unconditional expectation, while EM-scores greater than one indicate higher probabilities of earnings overstatement. For comparison, we choose 2% (of total assets) as the naïve cut-off point for unexpected accruals and total accruals measures. Examining the distribution of EM-scores reveals that an EM-score of 1.00 is located around the 60th percentile, and when a cut-off point for unexpected accruals of 0.02 is applied, this also is around the 60th percentile of these distributions.

The results of this "naïve" approach result in the EM-score model correctly classifying about 50% of all observations, with type I error rates of about 50% and type II error rates of 30%. Relative to the results reported in Table 5, this result is indicative of how minimizing the total misclassification cost is associated with the lowest rate of type II errors. In contrast, the results for measures of unexpected accruals, as well as the simple total accruals measure, continue to have lower type I

Electronic copy available at: https://ssrn.com/abstract=3015043

error rates, but at the cost of higher type II error rates. Hence, even when using a naïve strategy with simple cut-off benchmarks, the EM-score approach significantly outperforms commonly used measures of unexpected accruals.

6. Conclusion

Our paper is motivated by recognition that academic researchers face increasing pressure to demonstrate the wider impact (i.e., practical relevance) of their research. Yet there are relatively few examinations of factors that likely restrict accounting research from having such impact. Following the argument of Hopwood (2007), we explicitly recognize that accounting researchers face strong incentives to undertake research that has a low probability of making type I errors (i.e., falsely rejecting the null hypothesis). This is in marked contrast to potential users of research who often have type II error costs that are far higher than those associated with type I errors.

We illustrate this dilemma by considering the limited relevance of earnings management research for identifying relatively egregious instances of earnings overstatements. Although Moehrle et al. (2009) point to earnings management research as an example of how accounting research has practical relevance, they offer no examples of how it is actually used in practice. In contrast, Ball (2013) argues that there is little evidence of methods used in earnings management research being used in practice. We focus on methods used to estimate the extent of earnings management because they have had widespread application and, despite longstanding recognition of issues associated with a lack of power, leading journals continue to publish extensions of these models which have as their primary focus a further reduction in type I errors (Collins et al. 2017).

Electronic copy available at: https://ssrn.com/abstract=3015043

In contrast to research examining extensions of the Jones (1991) model of expected accruals that typically focus on the extent to which type I errors are reduced, our focus is on the power of these models relative to a simple red flag, accounting ratio-based model for detecting earnings overstatements of a typically large magnitude. We demonstrate that a combination of a simple measure of accruals combined with some straightforward financial ratio analysis can successfully distinguish between firms alleged to engage in quite serious earnings overstatements, relative to a set of matched control firms. The most important components of our EM-score model are measures of operating accruals and estimated revenue manipulation. This result is not surprising, as we rely on a sample of firm-year subject to SEC enforcement action due to earnings overstatements as our benchmark indicator of upwards earnings management.

However, the primary contribution of the paper lies neither in the recognition that financial statement analysis techniques are useful for identifying earnings overstatements, nor suggesting a model that is better than that outlined by Dechow et al. (2011). Our focus is on highlighting how low test power becomes critical when the expected relative cost of type I and type II errors are explicitly considered, and the resulting misclassification costs are minimized. Despite early evidence on this point in Dechow et al. (1995), there has been little if any evidence on how power becomes increasingly important when the relative costs of type II errors far exceed those associated with type I errors. Although it is entirely appropriate that researchers are concerned with avoiding type I errors, it is equally apparent that users of financial reports (including auditors) are far more concerned with the need to avoid type II errors. Our findings display a consistent pattern that commonly used unexpected accruals models reduce type I errors by sacrificing type II errors. In this respect,

Electronic copy available at: https://ssrn.com/abstract=3015043

advances to methods for detecting earnings management that reflect improvements via reduced type I errors have limited practical relevance.

In conclusion, the results reported in this paper highlight the tension between the search for better-specified methods by which to test hypotheses about the factors giving rise to accounting manipulation (or the consequences of manipulation) versus the practical interest in getting the most powerful tools for identifying instances of earnings overstatements where. Although this tension is not necessarily a bad thing, we argue that it needs to be explicitly recognized and that future research directed at practical solutions for identifying earnings overstatements should have a primary focus on increased power, so as to try and minimize type II errors. While accounting researchers continue to emphasize the minimization of type I error rates, we expect that broader impact of such research to be severely limited.

Electronic copy available at: https://ssrn.com/abstract=3015043

References

Ball, R. 2013. Accounting informs investors and earnings management is rife: two questionable beliefs. *Accounting Horizons* 27 (4):847-853.

Beaver, W., and M. McNichols. 1998. The characteristics and valuation of loss reserves of property-casualty insurers. *Review of Accounting Studies* 3 (1): 73-95.

Beneish, M.D. 1999. The Detection of Earnings Manipulation. *Financial Analysts Journal* 55 (5): 24-36.

Bowen, R.M., A.K. Davis and S. Rajgopal. 2002. Determinants of revenue reporting practices for internet firms. *Contemporary Accounting Research* 19 (4): 523-562.

Collins, D. W., R. S. Pungaliya, and A. M. Vijh. 2017. The effects of firm growth and model specification choices on tests of earnings management in quarterly settings. *The Accounting Review* 92 (2): 69-100.

Davis, A. K. 2002. The Value Relevance of Revenue for Internet Firms: Does Reporting Grossed‐up or Barter Revenue Make a Difference? *Journal of Accounting Research* 40 (2): 445-477.

Dechow, P. M., and I. D. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77 (s-1): 35-59.

Dechow, P., S. Richardson and I. Tuna. 2003. Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies* 8 (2-3): 355-384.

Dechow, P., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193-225.

Dechow, P., R. Sloan, and A. Sweeney. 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13 (1): 1-36.

Dechow, P., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50 (2): 344-401.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28 (1):17-82.

Dechow, P. M., A. P. Hutton, J. H. Kim, and R. G. Sloan. 2012. Detecting earnings management: A new approach. *Journal of Accounting Research* 50 (2): 275-334.

Dyckman, T.R., and S.A. Zeff. 2015. Accounting research: Past present and future. *Abacus* 51 (4): 511-524.

Feroz, H., K. Park, and V. Pastena. 1991. The financial and market effects of the SECs accounting and auditing enforcement releases. *Journal of Accounting Research* 29 (Supplement): 107-142.

Fields, T.D., T.Z. Lys and L. Vincent. 2001. Empirical research on accounting choice. *Journal of Accounting and Economics* 31 (1-3): 255-308.

Files, R. 2012. SEC enforcement: Does forthright disclosure and cooperation really matter? *Journal of Accounting and Economics* 53 (1): 353-374.

Electronic copy available at: https://ssrn.com/abstract=3015043

Gerakos, J. 2012. Discussion of Detecting Earnings Management: A New Approach. *Journal of Accounting Research* 50 (2): 335–347.

Hoang, K., S.E. Salterio, and J. Sylph. 2017. Barriers to transferring accounting and auditing research to standard setters. International Symposium on Audit Research, Sydney.

Hopwood, A.G. 2007. Whither accounting research? *The Accounting Review* 82 (5): 1365-1374.

Jansen, I. P., S. Ramnath, and T. L. Yohn. 2012. A diagnostic for earnings management using changes in asset turnover and profit margin. *Contemporary Accounting Research* 29 (1): 221-251.

Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29 (2): 193-228.

Kaplan, R.S. 2011. Accounting scholarship that advances professional knowledge and practice. *The Accounting Review* 86 (2): 367-383.

Kothari, S., A. Leone, and C. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39 (1): 163-197.

Marquardt, C.A. and C.I. Wiedman, C.I. 2004. How are earnings managed? An examination of specific accruals. *Contemporary Accounting Research* 21 (2): 461-490.

McNichols, M.F. 2002. Discussion of The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors. *The Accounting Review* 77 (s-1): 61-69.

Melumad, N. D., and D. Nissim. 2009. Line-item analysis of earnings quality. *Foundations and Trends® in Accounting* 3 (2–3): 87-221.

Miller, G. S., and D.J. Skinner. 1998. Determinants of the valuation allowances for deferred tax assets under SFAS No. 109. *The Accounting Review* 73 (2): 213-233.

Moehrle, S.R., K.L. Anderson, F.L. Ayres, C.E. Bolt-Lee, R.S. Debreceny, M.T. Dugan, C.E. Hogan, M.W. Maher and E. Plummer. 2009. The impact of academic accounting research on professional practice: An analysis by the AAA research impact task force. *Accounting Horizons* 23 (4): 411-456.

Nelson, M.W., J.A. Elliott and R.L. Tarpley. 2002. Evidence from auditors about managers' and auditors' earnings-management decisions. *The Accounting Review* 77 (Supplement): 175-202.

Ohlson, J.A. 2015. Accounting research and common sense. *Abacus* 51 (4): 525-535.

Owens, E.L., J.S. Wu and J. Zimmerman. 2017. Idiosyncratic shocks to firm underlying economics and abnormal accruals. *The Accounting Review* 92 (2): 183-219.

Petroni, K.R. 1992. Optimistic reporting in the property casualty insurance industry. *Journal of Accounting and Economics* 15 (4): 485-508.

Rasmussen, S. J. 2013. Revenue recognition, earnings management, and earnings informativeness in the semiconductor industry. *Accounting Horizons* 27 (1): 91-112

Roychowdhury, S. 2006. Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42 (3): 335-370.

Electronic copy available at: https://ssrn.com/abstract=3015043

Schilit, H. 2010. *Financial Shenanigans: How to Detect Accounting Gimmicks & Fraud in Financial Reports,* Third Edition. New York, NY: McGraw Hill.

Schrand, C. M., and S. L. Zechman. 2012. Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics* 53 (1): 311-329.

Sheskin, D. 2003. *The Handbook of Parametric and Nonparametric Statistical Procedures*, Third Edition. Boca Raton, FL: Chapman and Hall/CRC

Stubben, S. R. 2010. Discretionary revenues as a measure of earnings management. *The Accounting Review* 85 (2): 695-717.

Summers, S.L., and J.T. Sweeney. 1998. Fraudulent Misstated Financial Statements and Insider Trading: An Empirical Analysis. *The Accounting Review* 73 (1): 131-146.

Teoh, S., T. Wong, and G. Rao. 1998. Are accruals during initial public offerings opportunistic? *Review of Accounting Studies* 3 (1–2): 175–208.

Zang, A. Y. 2011. Evidence on the trade-off between real activities manipulation and accrual-based earnings management. *The Accounting Review* 87 (2): 675-703.

Electronic copy available at: https://ssrn.com/abstract=3015043

Appendix 1: Variables Measurement

Variable	Measurement
Panel A: Earnings management variables	
EM	*EM* is the indicator variable equals to one, if the firm is listed in the AAER as the upwards earnings manipulators, otherwise zero.
EM-Score	*EM-Score* is the predicted probability that is derived from the EM-Score model (EXP(*EM*)/(1+Exp(*EM*)) and deflated by the unconditional probability of earnings overstatements.
EM-Score_dummy	*EM-Score_dummy* equals to one if the *EM-score* is equal to or greater than 1.00, otherwise zero.
Panel B: Red flag variables	
ACC	The operating accruals magnitude is measured by $(EBEI_t+D\&A_t-CFO_t)/Assets_t$, where *EBEI* is earnings before Extraordinary Items; *D&A* refers to the aggregate Depreciation and Amortisation Expense; *Assets* refer to the total value to all assets.
SLSI	The sales index is measured by $Net\ Revenue_t / Net\ Revenue^{\dagger}$, where *Net Revenue* is the net sale of the firm. *Net Revenue*† is the estimated value of non-manipulated net revenue (see Appendix 2).
ACCI	The accruals index is measured as one plus the current value of operating accruals deflated by the average of current and lagged total assets in the year of GAAP violation: $\frac{1+((EBEI_t + D\&A_t - CFO_t)/((Assets_t + Assets_{t-1})/2))}{1+((EBEI_{t-1} + D\&A_{t-1} - CFO_{t-1})/((Assets_t + Assets_{t-1})/2))}$
INVI	The inventory index is measured as one plus the ratio of current inventory to net revenue, all deflated by one plus the ratio of lagged inventory to lagged net revenue: $\frac{1+(Inventory_t / Cost\ of\ Goods\ Sold_t)}{1+(Inventory_{t-1} / Cost\ of\ Goods\ Sold_{t-1})}$
RESI	The reserve index is measured as one plus the ratio of the lagged value of the bad debts reserve relative to the lagged receivables balance, all deflated by one plus the ratio of the current bad debts reserve relative to the current receivables balance: $\frac{1+(BDR_{t-1} / Receivables_{t-1})}{1+(BDR_t / Receivables_t)}$, where *BDR* is the provision for doubtful receivables
AQI	The asset quality index is measured as one plus the current value of soft assets deflated by the average of the current and lagged value of total assets in the year of GAAP violation, divided by one plus the lagged value of soft assets scalded by the aforementioned deflator (the average value of total assets for the year of GAAP violation and the prior year): $\frac{1+(SA_t/((Assets_t + Assets_{t-1})/2))}{1+(SA_{t-1}/((Assets_t + Assets_{t-1})/2))}$, where *SA* is the soft assets measured as the sum of *Other Current Assets*, plus Intangibles, and *Other Non-Current Assets*.

Electronic copy available at: https://ssrn.com/abstract=3015043

Appendix 1 (continued)

Variable	Measurement
Panel C: Performance indicators variables	
RECD_RECT	The ratio of *Receivables* to *Sales*
INV_COGS	The ratio of *Inventory* to *Cost of Goods Sold*.
SA_AT	The ratio of *Soft Assets* to *Total Assets*. *Soft Assets* are defined as the sum of *Other Current Assets*, plus *Intangibles*, and *Other Non-Current Assets*.
SG	The growth rate(%) in reported *Net Revenue* from the control year to the fraud year
ASSET_GR	The growth rate (%) of *Total Assets*
ASSETTURN	The asset turnover is measured as *Net Revenue* divided by *Total Assets*.
ROA	The return on assets is measured as *Earnings before Extraordinary Items* deflated by *Total Assets*.
GM	The gross margin is computed as *Gross Profit* divided by *Net Revenue*.
SGA	The margin for Selling, General and Administrative Expenses (*SGA*) is computed as the ratio of *SGA* expenses divided by *Net Revenue*.

Electronic copy available at: https://ssrn.com/abstract=3015043

Appendix 2

Net Revenue† is a notional estimate of non-manipulated net revenue. It seeks to measure the effect of manipulation that specifically results from opportunistic distortions in the timing of its recognition. This is determined by analyzing the time series change in the ratio of receivables-to-sales.

To begin, revenue should be recognised in the period that it is *earned*. However the derivation of revenue is often a continuous process and ascertaining an appropriate point at which it is *earned* is difficult to determine and requires some subjective judgment. This subjective judgment facilitates the use opportunistic manipulation. Outlined below is a set of mathematical formulas to derive *Net Revenue*†.

To determine current non-manipulated net revenue (Net Revenue†), reported revenue must be reduced by the amount by which it is overstated. This can be mathematically expressed as

$$\text{Net Revenue}^{\dagger} = \text{Sales Revenue} - \Delta S$$

Where *Net Revenue*† equals revenue before the effect of prematurely recognized credit sales, *Sales Revenue* is the amount reported in financial statements, and ΔS is the dollar volume of overstated revenue. Due to the fixed accounting relations that tie the balance sheet to the profit and loss statement, any intervention in the timing of credit sales will be accrued on the balance sheet, namely the accounts receivable balance. In essence, Net Revenue overstatements resulting from expedited credit transactions are simultaneously and uniformly reflected as overstatements in accounts receivables. Hence, the overstatement in sales revenue can be expressed as

$$\Delta S = \Delta R$$

Electronic copy available at: https://ssrn.com/abstract=3015043

Where ΔR is the dollar value of overstated receivables. The amount of aggressively accrued receivables can also be expressed as

$$\Delta R = \text{Accounts Receivable} - \text{Accounts Receivable}^{\dagger}$$

Where *Accounts Receivable* is the amount reported in the financial statements, and *Accounts Receivable*† is the dollar amount of non-manipulated receivables. To estimate a value for *Accounts Receivable*†, several assumptions need to be made. Typically, timing manipulation that seeks to capture premature sales transactions is detected from time series increases in the ratio of receivables to sales. Assuming last year's ratio is a clean/non-manipulated ratio, *Accounts Receivable*† can be expressed as

$$\text{Accounts Receivable}^{\dagger} = (\text{Accounts Receivable}_{t-1} / \text{Net Revenue}_{t-1}) * \text{Net Revenue}^{\dagger}$$

These mathematical relationships provide the basis for a simultaneous equation approach to detecting and estimating the magnitude of revenue manipulation. The accuracy of the model is dependent on the extent to which revenue manipulation is observed and reflected in changes in the ratio of receivables to sales. Putting all these equations together and solving for *Net Revenue*† gives

$$\text{Net Revenue}^{\dagger} = \text{Sales Revenue}_{t} - [\text{Accounts Receivable}_{t} - ((\text{Accounts Receivable}_{t-1} / \text{Net Revenue}_{t-1}) * \text{Net Revenue}_{t}{}^{\dagger})]$$

$$\text{Net Revenue}^{\dagger} = \frac{\text{Net Revenue}_{t} - \text{Accounts Receivable}_{t}}{(1 - \text{Accounts Receivable}_{t-1} / \text{Net Revenue}_{t-1})}$$

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 1: Sample selection criteria and summary statistics

Panel A: Sample selection criteria

AAERs obtained from the SEC website that are dated between 1989 and 2011	1,554
Less: Earnings Manipulators identified as understated from the AAER search	(43)
Less: Earnings Manipulators with insufficient data on Compustat	(938)
Final Sample	573

Panel B: Summary Statistics

	N	Mean	Std	Q1	Median	Q3
ACC	3424	-0.019	0.154	-0.039	-0.005	0.032
SLSI	3424	0.027	0.099	0.000	0.003	0.014
INVI	3424	1.000	0.066	0.985	1.000	1.010
RESI	3424	1.004	0.054	0.992	1.000	1.009
AQI	3424	1.032	0.105	0.994	1.005	1.036

Table 1 reports the summary of sample selection criteria and the summary statistics of key variables used in our paper. *ACC* is the operating accruals magnitude, *SLSI* is the sales index, *ACCI* is the accruals index, *INVI* is the inventory index, *RESI* is the reserve index, and *AQI* is the asset quality index. All variables have been winsorized at percentile bands one and ninety-nine. All variables are defined in the Appendix 1.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 2 Univariate test – AAERs and control firms

	Manipulators		Non-manipulators		Manipulator vs. non-Manipulators		
	Control Year	**Fraud Year**	**Control Year**	**Fraud Year**	**Control Year**	**Fraud Year**	**Difference in difference**
PANEL A: The red flag accounting ratio							
ACC	0.009	0.010	-0.012	-0.022	**0.021****	**0.032*****	0.014
					(2.24)	(3.29)	(1.33)
SLSI	0.069	0.056	0.025	0.026	**0.044*****	**0.030*****	**0.014***
					(5.86)	(4.65)	(1.87)
ACCI	1.015	1.043	1.015	1.003	0.000	**0.040*****	0.015
					(0.01)	(3.02)	(0.97)
INVI	0.995	1.005	1.002	0.999	-0.007	0.006	0.004
					(-1.30)	(1.43)	(0.80)
RESI	1.001	1.007	1.002	1.004	-0.001	0.003	-0.000
					(-0.32)	(0.88)	(-0.09)
AQI	1.062	1.055	1.029	1.028	**0.033*****	**0.027*****	0.004
					(4.32)	(4.00)	(0.51)
PANEL B: Performance indicators							
RECD_RECT	0.058	0.064	0.063	0.065	-0.005	-0.001	0.002
					(-0.62)	(-0.20)	(0.22)
INV_COGS	0.221	0.226	0.208	0.202	0.013	0.024	0.015
					(0.66)	(1.47)	(0.74)
SA_AT	0.175	0.197	0.180	0.187	-0.005	0.010	-0.012
					(-0.37)	(0.86)	(-0.86)
SG	0.666	0.431	0.351	0.378	0.315	0.053	0.032
					(1.53)	(0.20)	(0.12)
ASSET_GR	0.683	0.647	0.543	0.476	0.140	0.171	0.149
					(0.45)	(0.60)	(0.48)
ASSETTURN	1.245	1.111	1.183	1.168	0.062	-0.057	0.062
					(0.91)	(-0.91)	(0.88)
ROA	-0.009	-0.038	0.000	-0.021	-0.009	-0.017	0.006
					(-0.60)	(-1.03)	(0.33)
GM	0.367	0.387	0.375	0.369	-0.008	0.018	-0.002
					(-0.41)	(1.07)	(-0.12)
SGA	0.357	0.366	0.327	0.346	0.030	0.020	0.016
					(0.93)	(0.68)	(0.47)

The two-sample t-test is used to examine the differences in the mean value of financial statement characteristics between earnings manipulators and non-earnings manipulators in their *control* and *fraud* years. *Fraud* year refers to the year of GAAP violation (10-K filing) under investigation. *Control* refers to the year prior to the 10-K filing under SEC investigation.
For the red flag accounting ratios, *ACC* is the operating accruals magnitude, *SLSI* is the sales index, *ACCI* is the accruals index, *INVI* is the inventory index, *RESI* is the reserve index, and *AQI* is the asset quality index.
For the Control and Fraud year, *RECD_RECT* is the ratio of account receivables to sales, *INV_COGS* is the ratio of inventory to cost of goods sold, *SA_AT* is the ratio of soft assets to total assets, *SG* is the sales growth rate, *ASSET_GR* is the growth rate of total assets, *ASSETTURN* is the assets turnover, *ROA* is return on assets, *GM* is the gross margin, and *SGA* is the margin for selling, general and administrative expenses. All variables have been winsorized at percentile bands one and ninety-nine. The t-stats are reported in brackets directly beneath each coefficient and test statistic. *** (**, *) indicates significant at the 1% (5%, 10%) level for two-tailed test. All variables are defined in the Appendix 1.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 3: Analysis of red flag ratios for various EM-score values

EM-score Range			*FREQ*	*FREQ (Manipulator)*	*FREQ (non-Manipulator)*	*ACC*	*SLSI*	*ACCI*	*INVI*	*RESI*	*AQI*
0.125	<=EM-Score<	0.15	4	0	4	-1.309	0.003	0.504	0.992	0.883	0.877
0.15	<=EM-Score<	0.175	3	0	3	-1.333	0.002	0.328	1.000	1.083	0.772
0.175	<=EM-Score<	0.2	3	0	3	-1.238	0.004	0.448	0.883	1.075	0.772
0.2	<=EM-Score<	0.225	2	2	0	-1.145	0.000	0.531	1.025	1.088	0.783
0.225	<=EM-Score<	0.25	1	0	1	-1.379	0.000	0.118	1.014	1.306	0.772
0.25	<=EM-Score<	0.275	1	0	1	-1.379	0.458	-0.215	0.994	1.018	0.947
0.275	<=EM-Score<	0.3	5	0	5	-1.100	0.042	0.036	1.052	1.103	0.868
0.3	<=EM-Score<	0.325	4	0	4	-0.799	0.006	0.205	0.968	0.980	0.839
0.325	<=EM-Score<	0.35	2	1	1	-0.727	0.002	0.377	1.066	0.955	0.884
0.35	<=EM-Score<	0.375	1	1	0	-1.259	0.003	0.248	0.717	1.381	0.914
0.375	<=EM-Score<	0.4	3	1	2	-0.512	0.005	1.207	1.061	0.982	0.829
0.4	<=EM-Score<	0.425	4	0	4	-0.944	0.058	0.280	0.977	1.174	0.938
0.425	<=EM-Score<	0.45	4	1	3	-0.421	0.005	0.832	0.985	0.996	0.785
0.45	<=EM-Score<	0.475	4	1	3	-0.604	0.027	0.628	0.993	1.093	0.854
0.475	<=EM-Score<	0.5	3	1	2	-0.518	0.060	0.753	0.872	1.038	0.880
0.5	<=EM-Score<	0.525	4	0	4	-0.366	0.011	0.948	0.997	0.969	0.888
0.525	<=EM-Score<	0.55	5	1	4	-0.533	0.015	0.703	0.982	1.104	0.898
0.55	<=EM-Score<	0.575	6	0	6	-0.485	0.056	0.688	0.917	0.993	0.989
0.575	<=EM-Score<	0.6	11	0	11	-0.424	0.012	0.649	0.928	1.015	0.959
0.6	<=EM-Score<	0.625	12	0	12	-0.302	0.025	0.911	0.990	0.943	0.984
0.625	<=EM-Score<	0.65	13	0	13	-0.171	0.032	1.213	0.975	0.911	0.956
0.65	<=EM-Score<	0.675	23	1	22	-0.205	0.022	0.889	0.946	0.931	0.965
0.675	<=EM-Score<	0.7	15	4	11	-0.217	0.024	0.921	0.961	1.007	0.926
0.7	<=EM-Score<	0.725	29	4	25	-0.222	0.029	0.907	0.970	1.011	0.949
0.725	<=EM-Score<	0.75	36	5	31	-0.148	0.014	1.046	0.996	0.986	0.953
0.75	<=EM-Score<	0.775	41	6	35	-0.144	0.025	1.004	0.969	0.997	0.955
0.775	<=EM-Score<	0.8	40	7	33	-0.162	0.017	0.886	0.994	1.004	0.978
0.8	<=EM-Score<	0.825	57	2	55	-0.079	0.010	1.009	0.995	0.988	0.964
0.825	<=EM-Score<	0.85	72	12	60	-0.084	0.025	0.988	1.002	0.985	0.988
0.85	<=EM-Score<	0.875	137	13	124	-0.063	0.013	0.952	0.994	0.987	0.989
0.875	<=EM-Score<	0.9	168	25	143	-0.047	0.012	0.994	0.988	0.994	0.995
0.9	<=EM-Score<	0.925	329	40	289	-0.026	0.008	0.996	0.991	0.997	0.997
0.925	<=EM-Score<	0.95	431	69	362	-0.017	0.010	0.992	0.998	1.001	1.004
0.95	<=EM-Score<	0.975	404	46	358	0.002	0.014	1.005	0.994	0.999	1.010
0.975	<=EM-Score<	1	312	60	252	0.015	0.013	1.021	1.002	1.001	1.018
1	<=EM-Score<	1.025	235	40	195	0.024	0.025	1.007	1.011	1.005	1.019
1.025	<=EM-Score<	1.05	887	198	689	0.057	0.040	1.052	1.009	1.014	1.079
1.05	<=EM-Score<	1.075	123	31	92	0.045	0.200	1.086	1.026	1.053	1.364
1.075	<=EM-Score<	1.1	2	1	1	0.151	0.496	0.309	1.043	0.993	1.534

Table 3 reports mean values for the six red flag accounting ratios for non-financial firm-year observations within certain EM-score ranges. *ACC* is the operating accruals magnitude, *SLSI* is the sales index, *ACCI* is the accruals index, *INVI* is the inventory index, *RESI* is the reserve index, and *AQI* is the asset quality index. All variables have been winsorized at two standard deviation points from the mean. The column *FREQ* measures the frequency of observations within certain EM-score ranges. All variables are defined in the Appendix 1.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 4 Results of the EM score models

	VARIABLES EMPLOYED IN THE LOGIT MODEL							
	INT	*ACC*	*SLSI*	*ACCI*	*INVI*	*RESI*	*AQI*	Pseudo R^2
Model 1	-1.596*** (-34.75)	**1.265***** (3.24)	-	-	-	-	-	0.0040
Model 2	-1.631*** (-34.31)	-	**0.749*** (1.92)	-	-	-	-	0.0011
Model 3	-1.918*** (-8.67)	-	-	0.307 (1.44)	-	-	-	0.0007
Model 4	-2.210*** (-3.21)	-	-	-	0.601 (0.88)	-	-	0.0002
Model 5	-2.562*** (-3.11)	-	-	-	-	0.949 (1.16)	-	0.0004
Model 6	-3.356*** (-8.31)	-	-	-	-	-	**1.684***** (4.38)	0.0058
Model 7	-1.615*** (-33.89)	**1.200***** (3.11)	0.628 (1.60)	-	-	-	-	0.0048
Model 8	-1.383*** (-4.87)	**1.402***** (3.05)	**0.666*** (1.68)	-0.229 (-0.83)	-	-	-	0.0050
Model 9	-1.594** (-2.12)	**1.386***** (3.00)	**0.660*** (1.67)	-0.230 (-0.83)	0.212 (0.30)	-	-	0.0051
Model 10	-3.300*** (-2.85)	**1.492***** (3.21)	0.644 (1.62)	-0.191 (-0.69)	0.200 (0.29)	**1.673*** (1.93)	-	0.0062
Model 11	-4.643*** (-3.83)	**1.344***** (2.84)	0.585 (1.46)	-0.153 (-0.54)	-0.054 (-0.08)	**1.649*** (1.91)	**1.523***** (3.84)	0.0107

Table 4 presents the results of the Logit analysis. *ACC*, *SLSI*, *ACCI*, *INVI*, *RESI* and *AQI* are the individual 'red flag' forensic accounting variables used to distinguish between the group of 573 earnings manipulators and the 2,865 control firms. These variables are outlined in the notes to Table 2. The t-stats are reported in brackets directly beneath each logit coefficient and test statistic. *** (**, *) indicates significant at the 1% (5%, 10%) level for two-tailed test. All variables are defined in the Appendix 1.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 5: Classification accuracy of the EM-score model

RELATIVE COSTS OF TYPE I AND TYPE II ERRORS	PRIOR PROBABILITY OF MANIPULATED EARNINGS	PRIOR PROBABILITY OF NON-MANIPULATED EARNINGS	CUT-OFF EM-SCORE	ACCURACY (%)	TYPE II ERRORS (%)	TYPE I ERRORS (%)
10/1	0.01	0.99	0.9999	43.53	21.00	56.83
20/1	0.01	0.99	0.9999	43.53	21.00	56.83
30/1	0.01	0.99	0.9999	43.53	21.00	56.83
40/1	0.01	0.99	0.9999	43.53	21.00	56.83
50/1	0.01	0.99	0.9999	43.53	21.00	56.83
10/1	0.02	0.98	0.9459	48.75	25.00	51.79
20/1	0.02	0.98	0.9459	48.75	25.00	51.79
30/1	0.02	0.98	0.9459	48.75	25.00	51.79
40/1	0.02	0.98	0.9459	48.75	25.00	51.79
50/1	0.02	0.98	0.9459	48.75	25.00	51.79
10/1	0.03	0.97	0.9345	44.69	22.67	56.32
20/1	0.03	0.97	0.9347	44.73	22.67	56.28
30/1	0.03	0.97	0.9399	45.91	23.67	55.03
40/1	0.03	0.97	0.9399	45.91	23.67	55.03
50/1	0.03	0.97	0.9399	45.91	23.67	55.03
10/1	0.04	0.96	0.9422	42.67	23.75	58.73
20/1	0.04	0.96	0.9422	42.67	23.75	58.73
30/1	0.04	0.96	0.9479	43.94	25.00	57.35
40/1	0.04	0.96	0.9479	43.94	25.00	57.35
50/1	0.04	0.96	0.9479	43.94	25.00	57.35

Table 5 investigates the classification accuracy of the EM-Score Model. Accuracy is measured as the percentage of sample firms correctly classified. Type I errors are the misclassification of non-earnings manipulators as earnings manipulators (expressed as a percentage). Type II errors are the misclassification of earnings manipulators as non-earnings manipulators (expressed as a percentage). The prior probability of manipulated earnings refers to the proportion of firms, relative to all firms, that are expected to manipulate earnings beyond the levels permissible by GAAP. The prior probability of non-manipulated earnings is one minus the prior probability of manipulated earnings. The cut-off EM-score is the EM-score associated with the lowest statistical cost of misclassification.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 6: Summary statistics and univariate test for unexpected accruals

Panel A: Summary Statistics of Accruals Measures for AAER and control firms

	N	Mean	Std	Q1	Median	Q3
Unexpected Total Accruals: Modified Jones Model	3342	0.014	0.124	-0.032	0.019	0.068
Unexpected Current Accruals: Modified DD Model	3100	0.006	0.074	-0.026	0.003	0.032
Unexpected Total Accruals: Modified Jones Model matching on ROA	3342	-0.006	0.154	-0.076	-0.004	0.067
Total Accruals	3382	-0.058	0.137	-0.106	-0.050	-0.007

Panel B: T-test

	Manipulators	**Non-manipulators**	**Difference: Mean**	**t-statistic**
Unexpected Total Accruals: Modified Jones	0.043	0.009	0.034	5.92
Unexpected Current Accruals: Modified DD	0.024	0.002	0.022	5.99
Unexpected Total Accruals: Modified Jones matching on ROA	0.019	-0.010	0.029	4.05

Panel C: Median test

	Manipulators	**Non-manipulators**	**Difference: Median**	χ^2
Unexpected Total Accruals: Modified Jones	0.034	0.016	0.018	14.55
Unexpected Current Accruals: Modified DD	0.011	0.001	0.010	9.90
Unexpected Total Accruals: Modified Jones matching on ROA	0.014	-0.009	0.023	17.53

Table 6 reports univariate tests for unexpected accrual metrics (UA), as well as tests of the difference between the 562 available firm-year observations identified as earnings manipulators relative to the 2,810 matched non-earnings manipulators. All accrual metrics have been winsorized at percentile bands one and ninety-nine.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 7: Unexpected accruals as an identifier of AAERS

Panel A: Multivariate analysis

	(1)	(2)	(3)	(4)
VARIABLES	EM	EM	EM	EM
EM-score_dummy	**0.407*** **	**0.370*** **	**0.491*** **	**0.377*** **
	(4.02)	(3.50)	(5.05)	(3.71)
Unexpected Total Accruals: Modified Jones Model	**1.703*** **	-	-	-
	(4.11)			
Unexpected Current Accruals: Modified DD Model	-	**3.075*** **	-	-
		(4.50)		
Unexpected Total Accruals: Modified Jones Model matching on ROA	-	-	**0.829*** **	-
			(2.69)	
Total Accruals	-	-	-	**1.729*** **
				(4.61)
Constant	-1.824***	-1.855***	-1.817***	-1.682***
	(-28.95)	(-28.05)	(-28.71)	(-24.42)
Pseudo R^2	0.0173	0.0174	0.0138	0.0185
Observations	3,342	3,100	3,342	3,382

Panel B: Marginal effects using the interquartile range

	Percentage of increase			
EM-score_dummy	5.72%	5.00%	6.98%	6.60%
Unexpected Total Accruals: Modified Jones	2.31%	-	-	-
Unexpected Current Accruals: Modified DD	-	2.34%	-	-
Unexpected Total Accruals: Modified Jones matching on ROA	-	-	1.60%	-
Total Accruals	-	-	-	1.16%

Panel C: Marginal effects using 37.5th to 75th percentiles

	Percentage of increase			
EM-score_dummy	5.72%	5.00%	6.98%	6.60%
Unexpected Total Accruals: Modified Jones	1.63%	-	-	-
Unexpected Current Accruals: Modified DD	-	1.73%	-	-
Unexpected Total Accruals: Modified Jones matching on ROA	-	-	1.15%	-
Total Accruals	-	-	-	0.78%

Three different unexpected accrual estimations are used to distinguish between a group of 562 earnings manipulators and 2,810 control matched non-earnings manipulators. *EM-score_dummy* equals to 1 if the EM-score is equal to or greater than 1, 0 otherwise. All accrual estimates have been winsorized at percentile bands one and ninety-nine. The t-stat is reported in brackets directly beneath each logistic coefficient and test statistic. *** (**, *) indicates significant at the 1% (5%, 10%) level for two-tailed test. All variables are defined in the Appendix 1.

Electronic copy available at: https://ssrn.com/abstract=3015043

Table 8 Classification accuracy of unexpected accruals

Panel A: Unexpected Total Accruals: Modified Jones Model

RELATIVE COSTS OF TYPE I AND TYPE II ERRORS	PRIOR PROBABILITY OF MANIPULATED EARNINGS	PRIOR PROBABILITY OF NON-MANIPULATED EARNINGS	CUT-OFF POINT	ACCURACY (%)	TYPE II ERRORS (%)	TYPE I ERRORS (%)
10/1	0.02	0.98	0.1112	87.21	75.50	11.51
20/1	0.02	0.98	0.1112	87.21	75.50	11.51
30/1	0.02	0.98	0.1112	87.21	75.50	11.51
40/1	0.02	0.98	0.1112	87.21	75.50	11.51
50/1	0.02	0.98	0.1112	87.21	75.50	11.51
10/1	0.04	0.96	0.1118	85.68	79.50	11.60
20/1	0.04	0.96	0.1118	85.68	79.50	11.60
30/1	0.04	0.96	0.1192	86.61	80.50	10.59
40/1	0.04	0.96	0.3151	94.65	93.25	1.69
50/1	0.04	0.96	0.3151	94.65	93.25	1.69

Panel B: Unexpected Current Accruals: Modified DD Model

10/1	0.02	0.98	0.1116	93.48	85.00	4.92
20/1	0.02	0.98	0.1116	93.48	85.00	4.92
30/1	0.02	0.98	0.1116	93.48	85.00	4.92
40/1	0.02	0.98	0.1116	93.48	85.00	4.92
50/1	0.02	0.98	0.1116	93.48	85.00	4.92
10/1	0.04	0.96	0.1116	91.85	87.75	4.83
20/1	0.04	0.96	0.1116	91.85	87.75	4.83
30/1	0.04	0.96	0.1181	92.17	88.25	4.48
40/1	0.04	0.96	0.1121	92.17	88.25	4.48
50/1	0.04	0.96	0.1151	92.17	88.25	4.48

Panel C: Unexpected Total Accruals: Modified Jones Model matching on ROA

10/1	0.02	0.98	0.1116	82.55	79.50	16.18
20/1	0.02	0.98	0.1116	82.55	79.50	16.18
30/1	0.02	0.98	0.1116	82.55	79.50	16.18
40/1	0.02	0.98	0.1116	82.55	79.50	16.18
50/1	0.02	0.98	0.1638	88.43	85.00	10.07
10/1	0.04	0.96	0.1117	81.28	79.75	16.18
20/1	0.04	0.96	0.1117	81.28	79.75	16.18
30/1	0.04	0.96	0.3713	93.93	96.00	2.32
40/1	0.04	0.96	0.4411	94.78	97.50	1.38
50/1	0.04	0.96	0.4411	94.78	97.50	1.38

Panel D: Total Accruals

10/1	0.02	0.98	0.1122	92.62	84.50	5.81
20/1	0.02	0.98	0.1122	92.62	84.50	5.81
30/1	0.02	0.98	0.1122	92.62	84.50	5.81
40/1	0.02	0.98	0.1122	92.62	84.50	5.81
50/1	0.02	0.98	0.1122	92.62	84.50	5.81
10/1	0.04	0.96	0.1116	91.44	86.75	5.30
20/1	0.04	0.96	0.1116	91.44	86.75	5.30
30/1	0.04	0.96	0.1116	91.44	86.75	5.30
40/1	0.04	0.96	0.2146	94.03	91.50	2.41
50/1	0.04	0.96	0.2146	94.03	91.50	2.41

Table 8 investigates the classification accuracy of the unexpected accruals models, including Unexpected Total Accruals: Modified Jones Model, Unexpected Current Accruals: Modified DD Model, Unexpected Total Accruals: Modified Jones Model matching on ROA and Total Accruals. Accuracy is measured as the percentage of sample firms correctly classified. Type I errors are the misclassification of non-earnings manipulators as earnings manipulators (expressed as a percentage). Type II errors are the misclassification of earnings manipulators as non-earnings manipulators (expressed as a percentage). The prior probability of manipulated earnings refers to the proportion of firms, relative to all firms, that are expected to manipulate earnings beyond the levels permissible by GAAP. The prior probability of non-manipulated earnings is one minus the prior probability of manipulated earnings. The cut-off point for unexpected accruals is the level of unexpected accruals associated with the lowest statistical cost of misclassification.

Electronic copy available at: https://ssrn.com/abstract=3015043