Database Challenges in Financial Misconduct Research

Jonathan M. Karpoff
Professor of Finance
University of Washington
karpoff@uw.edu

Allison Koester
Assistant Professor of Accounting
Georgetown University
apk29@georgetown.edu

D. Scott Lee
Professor of Finance
University of Nevada, Las Vegas
scott.lee@unlv.edu

Gerald S. Martin
Associate Professor of Finance
American University
gmartin@american.edu

Current draft: May 30, 2014

Abstract: More than 150 research papers that examine the causes and effects of financial misconduct compile samples use one of four popular databases. We use a hand-collected sample of 1,099 cases identified by the SEC for financial misrepresentation to document, calibrate, and measure the importance of four features in each of these four databases that pose challenges for researchers. First, the initial public revelations of financial misconduct occur months before the initial coverage in these databases. Second, these databases collect just one type of event, so they omit other relevant announcements that affect a researcher's use of the events. Third, most of the events captured by these databases are unrelated to financial fraud. Fourth, the databases miss large numbers of events they were designed to capture. Event study tests and comparisons of firm characteristics show that these four database features can lead to economically meaningful biases in many applications. We show which database is most susceptible to each potential challenge, and propose remedies for researchers seeking to use these four databases.

JEL classifications: G38; K22; K42; M41
Keywords: Financial misconduct, restatement, class action lawsuit, Securities and Exchange Commission

Acknowledgements: We thank Emre Carr, Matthew Cedergren, Douglas Cumming, Patricia Dechow, Weili Ge, Karen Hennes, Shane Johnson, Andy Leone, Brian Miller, Tom Omer, Michael Pakaluk, Cathy Schrand, Nate Sharp, Richard Sloan, Wei-Ling Song, Jaron Wilde, Jide Wintoki, and workshop participants at Georgetown University, the University of Kansas, Fordham University, the University of Washington, Texas A&M University, Southern Methodist University, and the University of Alberta, We also thank conference participants at the 2012 CFA-FAJ-Schulich Conference on Fraud Ethics and Regulation, 2012 American Accounting Association (AAA) annual meeting, 2013 AAA Financial Accounting and Reporting (FARS) mid-year meeting, 2012 Investor Protection, Corporate Governance, and Fraud Prevention (IPCGFP) Conference, the 2012 European Finance Association annual meeting, 2012 Financial Management Association annual meeting, and 2013 Financial Markets and Corporate Governance Conference for helpful comments. Finally, we thank Jun Zhang, Lucas Perin, Chris Yost-Bremm, and Harvey Cheong for research assistance and Simi Kedia for data on the restated periods for the restatement announcement events in the Government Accountability Office (GAO) database.

1. Introduction

The electronic availability of data on financial restatements, class action lawsuits, and regulatory actions has facilitated an explosion of research on the causes and effects of financial misconduct.[1] Four databases figure prominently in this line of research: (1) the Government Accountability Office (GAO) and (2) Audit Analytics (AA) databases of restatement announcements, (3) the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and (4) the Securities and Exchange Commission's (SEC's) Accounting and Auditing Enforcement Releases (AAERs). The research that uses these databases has produced new insights about corporate governance, executive compensation, financial reporting, market efficiency, and regulatory policy.

This paper provides a systematic analysis of four features in each of these four popular databases that, if not recognized and properly addressed, can impart large biases to empirical tests. In particular, we illustrate and measure the frequency and economic importance of (1) late initial revelation dates, (2) scope limitations, (3) potentially extraneous events, and (4) complete and partial data omissions. We explain the research design implications of each feature, document which databases are least subject to concern about each feature, and offer suggestions to mitigate data-related biases when using these databases.

To document the severity of the database features and their economic importance, we compare the events identified by each database to 1,099 hand-collected histories for all cases in which the SEC brought regulatory action for financial misrepresentation from 1978 through 2011. The 1,099 case histories comprise a thorough record of 10,415 unique information events – an average of 9.5 events per case – that includes news articles, press releases, Form 8-K filings, restatement announcements, lawsuit filings and settlements, and regulatory releases from the SEC and Department of Justice. We refer to these cases as the HC (hand-collected) database throughout the paper.

[1] For examples, see Dechow et al. (1996), Hribar and Jenkins (2004), McNichols and Stubben (2008), Graham et al. (2008), Dyck et al. (2010), Dechow et al. (2011), and more than 150 other papers listed in Appendix A.

We report four primary findings about the severity and economic significance of the four database features. First, the initial dates associated with the events included in each of the four popular databases occur an average of 150 to 1,017 calendar days (depending on the database) after the initial public disclosure of the financial misconduct. As a result, event studies that rely on the dates in these databases understate the initial one-day market-adjusted stock price reaction to news of financial misconduct by 56% to 73% (depending on the database). Second, each database (by design) captures only one type of misconduct-related event (e.g., the GAO and AA databases capture only restatements, the SCAC database captures only securities class action lawsuits, and the AAER database captures only SEC enforcement actions receiving a secondary AAER designation). As a result, each captures only 6% to 36% (depending on the database) of the value-relevant announcements associated with the cases of misconduct they identify. Furthermore, the key informational events missed by each database have substantially larger impacts on firm value than the events captured by each database. Third, between 46% and 98% of the events in these databases are unrelated to charges of financial fraud. We show that the events that are related to financial fraud are systematically different from the other events in each database, indicating a need for substantial and systematic culling for researchers seeking samples of financial fraud. And fourth, these databases omit from 17% to 80% of the events they seek to capture during their sample periods. We show that firm and misconduct violation characteristics are statistically and economically different for included versus omitted cases in each database, indicating that empirical tests that rely exclusively on these databases are conducted on non-representative samples.

None of the four databases scores consistently well or poorly across these four features. AAERs, for example, suffer less than the other databases from scope limitations (feature #2) and potentially extraneous events (feature #3), but perform very poorly in identifying when investors first learn of the misconduct (feature #1). The SCAC database has relatively few errors of omission (feature #4), but also captures the fewest number of relevant informational events that pertain to each case of misconduct (feature #2). Table 8 at the end of this paper provides a summary ranking of the four databases' performance along each of the four database features. The fact that no single database consistently

performs well across all four features indicates that researchers may wish to consider these features when choosing their data source and designing empirical tests. For example, researchers seeking to measure stock price reactions to news of misconduct should be particularly concerned about late initial revelation dates (feature #1). Tests that rely on untainted control samples will be more affected by scope limitations (feature #2) and data omissions (feature #4), whereas the construction of samples of financial fraud will be particularly affected by potentially extraneous events (feature #3).

Broadly stated, the most important inference from our empirical results is that financial misconduct typically prompts a sequence of public announcements that can stretch over several years. Relying on the information related to only one or two of these announcements, as provided by these popular databases, can cause researchers to misclassify data and misinterpret empirical results. To avoid such problems, it typically is necessary to (i) research the full sequence of announcements that pertain to each case of misconduct that the researcher identifies, and (ii) be aware of and sensitive to data omissions in each of these databases.

We acknowledge that the AA, GAO, AAER, and SCAC databases were not compiled with the intent of (1) identifying the initial revelation date of financial misconduct; (2) identifying all important informational events related to the misconduct, or (3) culling extraneous events. We do not argue that these four databases *should* score well on these first three features. Rather, our analysis seeks to demonstrate how these features become important when the databases are used to examine such research questions as: the effects of the revelation of financial misconduct on a firm's value, costs, or characteristics; the characteristics of firms or managers that engage in or are caught engaging in financial misconduct; the likelihood that firms engage in misconduct; the penalties related to earnings misstatements; and other misconduct-related research questions.[2] We also hasten to emphasize that our analysis is not intended as a criticism of previous papers. Rather, we hope that a more complete

[2] Given that each database seeks to capture a comprehensive number of similar-type events for publicly traded firms within its specific time period, our fourth feature – complete and partial data omissions – is a shortcoming of each database per se, implying that researchers should be aware of potential selection biases when using each database.

understanding of the databases – what each includes and excludes – can facilitate more precise and powerful empirical tests in future research.

In documenting and calibrating the four database features, our paper adds to a long tradition of empirical research that helps to improve our understanding and use of popular databases. Rosenberg and Houglet (1974), Bennin (1980), and Elton et al. (2001), for example, examine the accuracy of CRSP, Compustat, and Morningstar data. Kahle and Walkling (1996) and Ali et al. (2009) document significant discrepancies in industry concentration and classifications when using data from Compustat, CRSP, or the U.S. Census. Anderson and Lee (1997) examine the accuracy of stock ownership databases. Chuk et al. (2013) document omission rates in Thomson Reuter's Company Issued Guidance database of management earnings forecasts. Ljungqvist et al. (2009) expose an ex post revision bias in analyst stock recommendations in the I/B/E/S database. Gillan et al. (2013) discuss empirical challenges that arise from backfilled data in the Execucomp dataset, and Aiken et al. (2013), Canina et al. (1998), Shumway (1997), Shumway and Warther (1999), and Harris et al. (2014) examine biases in other databases or their use. The paper that is closest to ours is Hennes et al. (2008), which partitions restatements in the GAO database into intentional misreporting "irregularities" versus simple "errors." This partition is similar to our third database feature, potentially extraneous events. However, our paper measures the extent and economic importance of four distinct features that affect four different databases, rather than a single database feature for a single database.

This paper proceeds as follows. Section 2 describes the four database features and provides an overview of each feature's prominence within each database. Section 3 describes our hand-collected data and our procedure for measuring the prominence of each feature in each database. Sections 4 through 7 are devoted to the four database features (one section per feature). We illustrate each feature with an example, document its severity in each database, and examine its economic significance. Section 8 concludes by ranking the four databases in terms of their overall susceptibility to each feature and

providing suggestions for mitigating the potential biases that can arise if the database features are not explicitly addressed in research designs.[3]

2. Brief descriptions of the four database features

The four databases examined in this paper are described in greater detail in Appendix B, but brief descriptions are as follows: The Government Accountability Office (GAO) database of restatement announcements consists of three reports issued between 2002 and 2006 that include a total of 2,705 restatement announcements from January 1, 1997 through June 30, 2006 (GAO 2002, 2003, 2006a, 2006b). Observations were identified by Lexis-Nexis "US Newspapers and Wires" database keyword searches for variants of "restate," "adjust," "amend," and "revise" within 50 words of "financial statement" or "earning." The observations were screened to keep only restatements that correct previous mistakes in applying financial reporting standards. The Audit Analytics (AA) database of restatement announcements is a commercial database that contains restatements and non-reliance filings made by all SEC registrants from January 1, 2000 through the present.[4] AA extracts its data principally from SEC Form 8-K or required amended periodic reports (Forms 10-K/A, 10-Q/A, 10KSB/A, 20-F/A, and 40-F/A), and analyze all 8-K and 8-K/A filings that contain "Item 4.02 - Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review" (an item required by the SEC since August 2004). The Securities Class Action Clearinghouse (SCAC) database contains information about federal civil securities class action lawsuits filed by a firm's shareholders from 1996 through the present. The Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs) refers to the subset of SEC administrative proceedings and litigation releases that receive a secondary designation as AAERs for "… releases involving accountants … Henceforth,

[3] In addition, an Internet Appendix for this paper summarizes (i) the financial misrepresentation statutes of the 1933 Securities Act and 1934 Securities Exchange Act; (ii) differences in the definition of "fraud" used by legal, finance, and accounting scholars; and (iii) additional information on our construction of the 1,099 case histories in our hand-collected data.

[4] Per email correspondence with an AA analyst on November 15, 2011, AA defines a restatement as "an adjustment to previously issued financial statements as a result of an error, fraud, or GAAP misapplication [and] does not include restatements caused by adoption of new accounting principles or revisions for comparative purposes as a result of mergers and acquisitions."

interested persons will be able to easily distinguish enforcement releases involving accountants from releases in which the Commission announces the adoption or revision of rules related to financial reporting or discusses its interpretive views on financial reporting matters."[5]

2.1. Late initial revelation dates

Researchers frequently seek to measure share price reactions to news of misconduct, or to compare firm characteristics before and after investors learn of the misconduct (e.g. Cheng et al. 2010; Hribar and Jenkins 2004; Kravet and Shevlin 2010). But none of the four databases were assembled with the intent of identifying the initial date upon which investors learn about financial misconduct. As a result, the restatement announcements in the GAO and AA databases, the lawsuit filing dates in the SCAC database, and the releases in the AAER data series all tend to lag the initial public revelation of the misconduct. Our hand-collected data include information on the initial event through which investors learned of the misconduct.[6] Using these data, we find that the earliest dates supplied by each of the four databases lag the actual initial revelation of misconduct by between 150 to 1,017 days, on average, depending on the database (see Table 3, Panel A). Empirical tests that use the initial GAO, AA, SCAC, or AAER dates to measure the share value impact of misconduct will understate the loss in share values by between 56% and 73%, depending on the database.

2.2. Scope limitations

Scope limitations arise because most cases of financial misconduct are revealed through a wide variety of announcements, whereas each database, by design, includes only one type of announcement: restatements (GAO and AA), securities class action lawsuits (SCAC), or SEC regulatory releases that

[5] This description of AAERs is provided in the first AAER issued on May 17, 1982. See Securities Act of 1933, Release Nos. 33-6396; 34-18649; 35-22457; IC-12377; AAER-1, 1982 SEC LEXIS 2565. Our database of AAERs consists of all AAERs issued from May 17, 1982 through December 31, 2011 (through AAER-3350). We note that a parallel database of AAERs through September 1, 2010 (AAER-3180) recently has been made available to researchers by the University of California, Berkeley Center for Financial Reporting & Management.

[6] A potential concern is that our identification of the initial revelation dates is affected by a look-ahead bias. In section 4.4 we demonstrate that this concern is unwarranted. In the majority of cases (72%), the initial revelation event is identified by the SEC in its enforcement releases and requires no interpretation or judgment on the part of the researcher. In the 28% of cases in which we identify the initial event date, the identification requires no controversial interpretation or judgment, as they all involve (i) restatement announcements, (ii) class action lawsuit filings, (iii) SEC inquiries or investigations into financial misconduct, or (iv) SEC regulatory sanctions for misconduct. Section 4.4 and Table 4 report a detailed breakdown of all 1,099 initial revelation events.

receive designation as an AAER. Thus, a researcher relying on the events in any one of these databases will miss most of the value-relevant information events for the cases of misconduct that the database identifies. We examine all cases in the GAO database that involve SEC enforcements of financial misrepresentation and find that the GAO events capture 9.8% of the information-relevant events, on average. The corresponding percentages for the AA, SCAC, and AAER databases are 8.7%, 5.9%, and 36.2% (Table 5, Panel A). Furthermore, the shareholder value losses associated with the events identified by these databases are small relative to the losses associated with the events they do not capture. For example, the average shareholder wealth effect of the events identified by the GAO database is -7.82%, compared to an average wealth effect of -50.36% for the other relevant announcements for these same cases that are not included in the GAO database. Stated differently, more than 84% of the shareholder losses related to financial misconduct escape detection when analysis is confined to the restatement announcements provided by the GAO database. The analogous averages for the AA, SCAC, and AAER databases are 88%, 90%, and 83% (Table 5, Panel B). We show below that such scope limitations affect how the misconduct is characterized and even whether a researcher will cull the case from his or her sample.

2.3. Potentially extraneous events

The third database feature arises because each database contains many events that researchers interested in financial misconduct seek to remove, or cull, from their samples. Whether a particular event should or should not be included in a sample depends on the researcher's specific objective, so we cannot provide exact counts of the magnitude of the potentially extraneous events feature for all research questions. We can, however, document the percentage of extraneous observations for two objective criteria appropriate for many research questions: (i) SEC-identified financial misrepresentation and (ii) SEC-identified financial fraud. First, the percentage of observations that do not prompt SEC enforcement for financial misrepresentation is 84.2% in the GAO database, 97.8% in the AA database, 88.6% in the SCAC database, and 19.7% in the AAER database (Panel A, Table 6). Second, the percentage of observations that do not prompt SEC or DOJ charges of financial fraud is 89.4% in the GAO database,

98.1% in the AA database, 90.4% in the SCAC database, and 46.2% in the AAER database (Panel C, Table 6). Stated differently, only one in 10 events identified by the GAO and the SCAC databases involve charges of financial misrepresentation and fraud. The "hit" rate for the AA database is even lower, as researchers using the AA database to identify misrepresentation and fraud must cull 49 of every 50 events. The events that we identify as potentially extraneous have mean one-day stock price reactions that approach zero, while the complementary set of events that involve financial misrepresentation and fraud have one-day stock price reactions that are negative, large in magnitude, and statistically significant (Table 6, Panels D – F). This indicates that how a researcher culls her sample can have an economically meaningful effect on the type and impact of the events that remain in the sample.[7]

Many researchers cull their samples manually to remove extraneous events and identify meaningful instances of misconduct.[8] However, unless researchers document their culling methodologies or make their culled samples publicly available, their research studies cannot be replicated. The culling criteria that we use reflect SEC and Department of Justice deliberations that are readily available in public records. In Section 6.3 we show that our criteria yield samples that are larger than the culled samples used in many papers, and are associated with stock price reactions of smaller magnitudes than reported in many papers. These findings suggest that, while many researchers are aware of the potential for extraneous observations in their research designs, it is common to overcorrect for this problem and construct samples that are weighted toward extreme cases of misconduct.

2.4. Complete and partial data omissions

The fourth database feature arises because each database omits both cases and events that should be included, given the database's own scope, sampling design, and time period. Whereas the first three

[7] An important qualification regards the SCAC cases. Most security class action lawsuits are brought under Section 10(b) of the Securities Exchange Act, and allege fraud by the firm or individuals associated with the firm. A user of these data would justifiably describe his or her sample as consisting of cases of (alleged) fraud. For these cases, our criteria help to sort out the more serious cases of fraud from minor or frivolous lawsuits – a task that Dyck et al. (2010) and others emphasize as important in constructing a useful database of financial misconduct.

[8] For an example of data culling using the GAO database, see Hennes et al. (2008), who categorize 73.6% of the events in the GAO database as restatements due to unintentional misapplications of GAAP ("errors") versus intentional misreporting ("irregularities"). See http://sbaleone.bus.miami.edu/. For the SCAC database, see Gande and Lewis (2009) and Dyck et al. (2010). For the AAER data series, see Dechow et al. (1996); Dechow et al. (2011); Erickson et al. (2004; 2006); and Schrand and Zechman (2012).

features arise because the databases were not designed for ways in which researchers use them, this fourth feature is a problem with the databases themselves. For example, the GAO database identifies 427 separate restatement announcements related to misconduct that prompted enforcement action by the SEC. During the GAO's sampling window, however, there were a total of 1,124 such restatement announcements – an omission rate of 62%. The analogous omission rates for the AA, SCAC, and AAER databases are 79.9%, 17.4%, and 49.8%.

Errors of omission arise in two ways: the database completely omits a case of misconduct that has one or more *same-type* events (e.g., when the GAO or AA databases miss all restatement announcements for a given case), or the database provides partial coverage of a misconduct case by identifying some but not all same-type events pertaining to the case. The GAO database, for example, identifies 427 of 1,124 restatement announcements pertaining to cases of misconduct that prompted SEC enforcement activity. Of the 697 omitted restatement announcements, 219 relate to cases of misconduct that the GAO database completely omits, and 478 relate to cases of misconduct in which the GAO identifies some but not all of the relevant restatement announcements.

Such high rates of data omissions pose at least two concerns. First, control samples assembled from firms that do not appear in a chosen database (e.g., GAO, AA, SCAC, or AAER) may include firms that actually belong in the treatment sample, thus biasing tests that use control samples. Second, data omissions shrink sample sizes, which can affect test power and result in non-representative samples. When confounded by potentially extraneous events in these databases (feature #3) and scope limitations that make it difficult to accurately characterize or classifying the events in the databases (feature #2), researchers who rely on the databases face significant measurement challenges.

Researchers who work with noisy data sometimes claim that noise works against finding a significant relation, so the true relation must be even stronger than the measured relation. As Roberts and Whited (2012) point out, however, this claim is not in general true. Rather, if measurement error in either the dependent or the independent variable is correlated with other independent variables, the coefficients of all independent variables are biased, and "… it is safe to say that bias is not necessarily

toward zero and that it can be severe" (Roberts and Whited 2012, p. 16). Similarly, Burgstahler (1987, p. 203) points out that, "… hypothesis tests with low power are not only undesirable *ex ante* (because of the low probability of observing significant results) but also *ex post* (because little probability revision should be induced even when significant results are observed)."

3. Detailed database comparisons

To implement our analysis, we first compile case histories for all 1,099 cases in which the SEC initiated an enforcement action between 1978 and 2011 for financial misrepresentation. To be included in our hand-collected (HC) sample, the case must include a violation of one or more of following three "13(b)" provisions of the Securities and Exchange Act of 1934: [9]

(i) Section 13(b)(2)(a), a.k.a. 15 U.S.C. §§ 78 m(b)(2)(A) – which requires firms to keep and maintain books and records that accurately reflect all transactions;

(ii) Section 13(b)(2)(b), a.k.a. 15 U.S.C. §§ 78 m(b)(2)(B) – which requires firms to devise and maintain a system of internal accounting controls; and

(iii) Section 13(b)(5), a.k.a. 15 U.S.C. §§ 78 m(b)(5) – which prohibits knowingly circumventing or failing to implement a system of internal accounting controls, or knowingly falsifying any book, record, or account.

To construct the case histories for all 1,099 cases, we hand-collected data from seven primary sources: (i) the SEC website (www.sec.gov), which contains SEC press and selected enforcement releases related to enforcement actions since September 19, 1995; (ii) the Department of Justice, which provides information on enforcement activity through a network of related agencies with particular emphasis on high-profile enforcement actions available at www.usdoj.gov; (iii) the Wolters Kluwer Law & Business Securities (Federal) electronic library, which contains all SEC releases and other materials as reported in the SEC Docket since 1973 and select Federal Securities Law Reporter releases from 1940 to 1972; (iv) Lexis-Nexis' FEDSEC:SECREL and FEDSEC:CASES library, which contains information on securities

[9] Most enforcement actions for financial misrepresentation include other charges as well, including Section 13(a) charges for late or incomplete filings, and fraud charges. When we limit our analysis to cases that involve such related violations, or delete such cases, the results are similar to those reported here. These results are available upon request.

enforcement actions; (v) the PACER database, which contains lawsuit-related information from federal appellate, district and bankruptcy courts; (vi) the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system; and (vii) Lexis-Nexis' All News and Dow Jones' Factiva news source, which includes news releases that reveal when firms are subject to private civil suits and regulatory scrutiny. The resulting case histories include events that are in the GAO, AA, SCAC, and AAER databases, along with many other informational events that are not in any of the four databases. The Internet Appendix to this paper provides additional detail on the data collection process.

We use the term "case" to describe the group of related events that identify a potential instance of misconduct. To assemble related events into cases for the GAO, SCAC, and AAER databases, we read all the events in these databases and manually matched all related events into cases. Yielding to the sheer number of events (11,001 restatement announcements) in the AA database, we used a less precise two-step procedure to map AA events into cases. First, we use our hand-collected data to identify all AA events that correspond to cases that include 13(b) violations. When the AA database identifies multiple restatement announcements associated with one firm's 13(b) violation, we assemble these restatements into a single case that is identified by the AA database. Second, for the remainder of the AA events that are not associated with a case with a 13(b) violation, we assemble restatement announcements by company name and combine multiple AA-reported restatements made by a firm into a single case if the restated periods overlap or are contiguous.

Table 1 provides an overview of the four databases (GAO, AA, SCAC, and AAER) used in much of the financial misconduct literature, and Panel A of Table 2 reports on the numbers of events and cases in each of the databases and our hand-collected (HC) sample. The GAO database consists of 2,707 restatements that pertain to 2,321 separate instances of alleged misconduct (i.e., cases). Counting through the end of 2010, the AA database has 11,001 restatements that pertain to 8,358 separate cases, the SCAC database has 3,421 lawsuit filings that pertain to 3,116 separate cases, and the AAER database has 3,568

SEC releases that pertain to 1,356 unique cases.[10] The hand-collected sample has 10,415 events related to 1,099 unique cases.

To examine three of the database features (late initial revelation dates, scope limitations, and complete and partial data omissions), we focus on the subset of each database's unique cases in which regulators initiated an enforcement action for a 13(b) financial misrepresentation violation. That is, we examine the intersection of each database's cases with the 1,099 cases in our hand-collected data. As reported in Panel A of Table 2, the GAO database identifies 290 cases (which include 427 GAO-identified restatement announcement events) that overlap with our hand-collected data. The AA database contains 188 cases (which include 239 AA-identified restatement announcement events) that overlap with the hand-collected data; the SCAC database contains 346 such cases that include 389 SCAC-identified securities class action lawsuits; and the AAER database has 939 such cases that include 2,865 individual AAERs.

Our documentation of these three features assumes a researcher using any one of these databases accurately combines multiple events that pertain to a single case and removes instances of misconduct that did not trigger SEC enforcement action for misrepresentation. If a researcher does not make such adjustments to the data, the remaining feature (i.e., potentially extraneous events) becomes important. To calculate the importance of this feature, we use all observations in the GAO, AA, SCAC, and AAER databases and not just the events and cases that intersect with our hand-collected sample.

Panel B of Table 2 further describes the wide array of event types in the hand-collected data. The histories for these 1,099 cases of financial misrepresentation consist of 10,415 distinct informational events. These events include 1,442 restatement announcements, 615 announcements of securities class action filings, 630 securities class action settlements, 3,066 SEC enforcement releases that receive a

[10] The last AAER in 2011 is numbered AAER-3350. Some AAERs, however, refer to misconduct at more than one firm. Our count of AAER releases includes all firm-AAER combinations. In counting AAER firm-events, we exclude 66 AAERs that reinstate previously disbarred or suspended accountants, eight AAERs that were never issued (AAER-372 through AAER-379), two interpretive AAERs that provide financial reporting guidance (AAER-82 discusses the significance of oral guarantees to the financing reporting process and AAER-120 discusses accounting for loan losses by registrants engaged in lending activities), and five AAERs that were "intentionally omitted" according to the SEC Docket (AAER-1029, 1092, 1400, 1941, and 2579). See the Internet Appendix for a detailed description of our mapping of AAERs to the creation of the AAER database used in our calculations.

secondary designation as an AAER, and 1,445 SEC enforcement releases that do not receive AAER designation. In addition, there are 1,298 "Other regulatory events," which consist primarily of press releases of enforcement activities undertaken by the DOJ. The 1,919 "Other press releases and material announcements" primarily include press releases by the target firms that reveal SEC or DOJ investigations of financial misconduct. This category also includes news articles, earnings forecast revisions, and announcements of employee turnover or auditor changes related to allegations of misconduct. Multiple events often occur on the same calendar date, and the 10,415 events in the hand-collected data occur on 8,787 unique calendar dates. To illustrate, Panel B of Table 2 reports there are 1,104 unique event dates with only restatement announcements and an additional 274 dates with both a restatement announcement and an "Other press releases and material announcements" event (i.e., the 'a+g' row at the bottom of Panel B). Overall, the average case involves 9.5 announcement events spread over 8.0 unique event dates.

As Figure 1 illustrates, one reason the databases differ is that they cover different time intervals. The GAO database covers the shortest period (1997 – June 30, 2006), followed by the AA (2000-2011), SCAC (1996 – 2011), and AAER (April 15, 1982 – 2011) databases and our hand-collected (1978 – 2011) data.[11] To measure the extent of each database feature in Sections 4 through 7 below, we use each database's unique sample period. For example, we count the GAO database's omissions of restatement announcements only during the GAO sample period. Note that for purposes of our study we truncate the AA and SCAC databases at December 31, 2010 but continue to gather data on SEC and DOJ enforcement activities (including AAERs) through 2011. Because restatements and class action lawsuits typically precede regulatory enforcement proceedings, we want to assure that we do not mistakenly count AA or

[11] While AAER-1 was issued on April 15, 1982, AAER-1 indexes and classifies 20 previous cases of financial misconduct that effectively extend its coverage period backwards. The earliest of these 20 previous cases had a related regulatory proceeding issued on September 24, 1971.

SCAC cases as extraneous observations when regulators initiate enforcement action in 2011 for cases in which there was a restatement or class action lawsuit filing before 2011.[12]

4. Feature #1: Late initial revelation dates

4.1. An illustration

From 1998 to 2000, Computer Associates (CA), a publicly traded software company, misreported more than $500 million in revenues in an effort to inflate its stock price. In particular, CA accelerated the recognition of revenues by fraudulently extending its reporting periods and misclassifying maintenance contracts to allow the firm to book future expected payments as current revenue. A *New York Times* article first exposed the fraud on April 29, 2001 with a detailed description of CA's fraudulent reporting practices and supporting claims from former CA employees and industry analysts (Berenson 2001).[13] On the day of the New York Times article, CA's stock price fell by 8.8% (one-day market-adjusted return) and the Securities and Exchange Commission (SEC) soon began an investigation of CA's financial reporting practices. Over the ensuing nine years, CA restated its earnings nine times, was sued in a securities class action lawsuit for financial reporting fraud, faced SEC administrative sanctions and civil penalties, and saw key employees targeted by the Department of Justice (DOJ) for criminal prosecution. We identify 49 separate days stretching from April 29, 2001 through October 14, 2010 on which new information about the nature, scope, and consequences of CA's financial fraud was publicly revealed via press coverage, SEC filings, and lawsuit filings. Figure 2 displays a chronology of these events.

[12] Among all cases in which regulators take action for financial misrepresentation and there is a restatement or class action lawsuit event, the initial regulatory proceeding occurs within one year of the restatement or class action lawsuit in three-quarters of the cases. Only rarely does an initial restatement (20 out of 1099 cases) or first security class action lawsuit filing (33 out of 1,099 cases) occur after the initial regulatory action. None of the regulatory enforcement actions initiated in 2011 contain a restatement or class action filing event in 2011, so it is unlikely that using a December 31, 2011 cutoff for the regulatory enforcement data causes any meaningful bias to our counts of data omissions, partial coverage, and potentially extraneous observation issues.

[13] See Berenson (2001). To illustrate the nature and specificity of the information in the New York Times article, the fifth paragraph cites analysts and former CA employees: "Computer Associates, they say, has used accounting tricks to systematically overstate its revenue and profits for years. The practices were so widespread that employees joked that C.A. stood for 'Creative Accounting,' and that March, June, September and December, when fiscal quarters end, had 35 days, giving the company extra time to close sales and book revenue."

The CA case is a good candidate for empirical investigations of the causes or consequences of financial misconduct. In fact, this case appears to be included in samples used by Hribar and Jenkins (2004), Fich and Shivdasani (2007), Graham et al. (2008), and many other important papers in the accounting and finance literatures. Some of these papers measure the stock return to the public discovery of financial misconduct, while others examine how firm characteristics change around the initial revelation of misconduct. These investigations require a simple, if fundamental, starting point: when did investors first learn of the misconduct? In CA's case, the first public revelation clearly occurred with the April 29, 2001 *New York Times* article – indeed, this is the event that subsequently is referenced in DOJ, SEC, and class action lawsuit filings as bringing attention to CA's misconduct.[14] A researcher relying only on the computerized databases, however, would miss the April 29 article. According to the GAO database, CA's first restatement announcement occurred on February 9, 2002, which is 296 calendar days after the *New York Times* article. According to the AA database, CA's first restatement announcement was on May 26, 2005, which is 1,488 days after the initial article. The SCAC database records the filing date of the class action lawsuit as February 25, 2002, or 302 days after the *New York Times* article. And the first AAER associated with the CA case is on January 22, 2004, which is 998 days after the initial article.

In each case, relying upon the database would cause a researcher to misidentify when investors initially learned of the financial misconduct, by a period that ranges from 296 to 1,488 days. To be sure, each of the events identified by the databases conveyed some new information about CA's misconduct or its consequences. But since investors already knew about the misconduct, it would be erroneous to identify any of these dates as the first public revelation of CA's misconduct. Rather, the databases' events conveyed new information about the exact nature of CA's restatements, lawsuits, or SEC actions that investors could partially but not fully anticipate from prior disclosures about CA's misconduct.

[14] References to the centrality of the New York Times article are available in the report of the Special Litigation Committee that was formed by the CA Board of Directors in response to a derivative lawsuit regarding CA's accounting fraud. See http://online.wsj.com/public/resources/documents/20070413_CA.pdf.

4.2. Late initial revelation dates in each of the four databases

The CA example is not unusual. Panel A of Table 3 shows the distribution of the lag between the initial date in each of the databases and the initial revelation date in the HC sample. The median initial GAO restatement occurs 14 calendar days after the initial revelation of misconduct. The distribution of the lag is skewed, as the mean is 187 days and the maximum is 2,242 days. The maximum occurred at Aspen Technology, which restated its financial statements more than six years after the initial public revelation of its misconduct. The minimum, -3 days, reflects an initial revelation that happened to be a restatement announcement issued on a Friday after the stock market was closed.[15]

The AA database identifies restatements that are associated with 188 of the 1,099 cases that involve SEC charges for financial misrepresentation. For these 188 cases, the median lag between initial revelation and the initial AA restatement date is 66 days, with a mean of 242 days. Class action lawsuit filings also typically lag the initial revelation of the misconduct. For the 346 unique cases identified by the SCAC database, the median lag of the lawsuit filing date is 23 days, with a mean lag of 150 days and a maximum of 2,118 days. Late initial announcements are a pronounced characteristic of the AAER database because the releases that the SEC designates as AAERs tend to occur toward the end of the case history. If we focus on the date of the earliest AAER associated with each of the 939 cases of misconduct identified, the median lateness is 991 days and the mean is 1,017 days.[16]

[15] We classify revelation dates according to the earliest date that a price reaction can be observed, i.e., the following Monday if the announcement is time-stamped after the close of markets on Friday. There are very few cases in which a database's initial date precedes the actual initial revelation date, so the average absolute deviation is very close to the summary numbers reported in Panel A of Table 3.

[16] To repeat, these calculations refer only to the subset of AAERs that are associated with a Section 13(b) financial misrepresentation violation. To better understand how AAERs are used, we collected data on all other AAERs released through 2010 (untabulated). These include an additional 722 AAERs that are associated with 412 additional instances of misconduct that did not involve charges of financial misrepresentation (13(b) violations). They include such charges as insider trading and malfeasance by broker-dealers or mutual fund managers – as long as the misconduct involved an accountant. A total of 77.2% of these additional 412 actions relate to equity funds, 7.5% relate to broker-dealers, 3.6% relate to mutual funds, and the remaining observations are spread across several smaller categories. This tabulation helps to illustrate the reason that some SEC releases receive a secondary designation as an AAER. An AAER itself is not necessarily an indication of financial misrepresentation or fraud.

4.3. Relevance and economic importance of late initial revelation dates

Researchers frequently seek to examine the valuation impact of financial misconduct by measuring share price reactions around the misconduct's disclosure (e.g., Feroz et al. 1991; Dechow et al. 1996; Fich and Shivdasani 2007; Cheng et al. 2010). Other researchers measure changes in such firm attributes as managerial turnover or the cost of equity or debt capital around the initial disclosure of financial misconduct (e.g., Helland 2006; Hribar and Jenkins 2004; Kravet and Shevlin 2010; Graham et al. 2008; Chen et al. 2013). The use of late initial revelation dates will affect a researcher's assessment of the timing of the misconduct and could affect empirical results. In this section we measure the size of the potential bias in one specific setting – measures of the valuation impact of financial misconduct.

Panel B of Table 3 **r**eports on the impact of late revelation dates in a typical event study using one-day market-adjusted returns. The GAO database identifies 290 instances of misconduct that triggered SEC sanctions for financial misrepresentation. CRSP data to measure stock returns for the initial GAO restatement date are available for 260 of these cases. The mean one-day market-adjusted share return for these 260 GAO dates is -7.06%, and the median is -2.13%. While these are significant declines in value, they substantially underestimate the actual impact on share values from the initial revelation of the misconduct that resulted in the restatements. Using the earliest initial revelation dates in the hand-collected data for these 260 cases, the mean market-adjusted return is -16.17% and the median is -9.31%. That is, the mean abnormal return as measured by the initial GAO restatement date understates the magnitude of the mean return on the initial revelation date by 56%. Using medians, the GAO data understates the impact on share values by 77%. The differences between the measures based on the GAO dates and the actual initial revelation dates are statistically significant at the 1% level.

Comparisons using the AA, SCAC, and AAER databases are similar. CRSP returns are available for 137 of the initial events related to the 188 cases of misconduct identified by the AA database. For these 137 cases, the mean one-day abnormal return is -4.83%, with a median of -1.67%. This is significantly smaller than the mean one-day abnormal return of -13.59% that results from using the actual initial revelation dates for these 137 cases. Thus, a researcher using the AA dates would understate

market reaction to the initial revelation of misconduct by 64% (79% using medians). CRSP returns are available on the class action lawsuit filing date for 300 of the 346 cases identified by the SCAC database that involved regulatory penalties for financial misrepresentation. The mean one-day abnormal return is -5.43% with a median of -1.21%. Using the actual initial revelation date for these 300 cases, the mean abnormal return is -18.64% with a median of -13.55%. The SCAC dates would lead a researcher to understate the mean initial share price reaction by 71% (91% using medians). Using the initial AAER events, the mean understatement is 73% (85% using medians).

In some applications, it is sufficient to identify merely the month, quarter, or year in which investors learn of financial misconduct (e.g., see Kravet and Shevlin 2010; Graham et al. 2008). In such applications the use of late initial revelation dates probably introduces relatively small measurement error. For research questions that require a precise event date, however, using the dates contained in the AA, GAO, SCAC, and/or AAER datasets introduces large measurement error. Our results indicate that the average magnitude of this measurement error in one common application – estimating the valuation impact of news of misconduct – ranges from 64% (for AA database events) to 73% (for AAER dataset events).

4.4. Assessing the potential for look-ahead bias in initial revelation dates

A potential concern about our tabulations is that the initial revelation dates in the hand-collected (HC) sample suffer from a look-ahead bias. A look-ahead bias would occur if the existence of misconduct is not apparent in the initial revelation and becomes apparent only with subsequent revelations. The data in Table 4, however, show that such a concern is unwarranted. Most of the initial revelation dates for the 1,099 cases (795, or 72%) are explicitly identified by the SEC in its enforcement releases and do not require any discretionary judgment by the researcher. The SEC refers to these as "trigger dates." We identify the other 304 (28%) initial revelation dates using the data sources described in Section 3. None of the 304 initial revelation dates we identify are likely to be controversial. Of these 304 dates, 81 include restatement announcements, 54 include class action filing dates, and 55 are regulatory actions taken by the SEC. These are the exact types of events that the GAO, AA, SCAC, or

AAER databases seek to capture. An additional 36 are firm revelations (via press reports or 8-K filings) that the SEC has initiated an inquiry into financial irregularities, and an additional 57 are firm announcements that the SEC has initiated a formal investigation of misconduct. Eight more are firm announcements that the firm and/or senior managers have received Wells Notices from the SEC regarding financial improprieties, and the final 13 are announcements of trading suspensions by the SEC related to the revelation misconduct. Thus, all 1,099 initial revelation dates in the hand-collected data are either: (i) identified by the SEC, or (ii) involve announcements that clearly indicate the possibility or likelihood of financial misconduct.

Column (3) of Table 4 catalogs the types and frequency of revelations that the SEC identifies as "trigger dates" that lead to its investigation of financial misconduct. It is possible that some SEC-identified revelations suffer from a look-ahead bias, but close scrutiny of the SEC releases indicates look-ahead bias is uncommon. Of the 795 initial revelation dates identified by the SEC, 203 involve restatement announcements, seven are class action filings, and 12 involve restatement announcements and class action filings on the same trading day. Most other categories in Table 4 are self-explanatory: 146 initial revelations are firm announcements about internal investigations for financial misconduct, while fewer numbers involve auditor changes, management changes, or bankruptcies in which accounting irregularities play a prominent role. The 'Combination: multiple firm revelations' category includes several types of adverse revelations by the firm, including accounting irregularities or internal investigations of accounting irregularities, impairment charges, severe earnings shortfalls, inadequate financial reserves, and debt covenant violations associated with accounting irregularities. As an example, Rent-Way, Inc. issued a press release on October 30, 2000 revealing an internal investigation of "certain accounting matters," the likelihood that its financial reports would have to be revised, and the suspension of its Corporate Controller. Rent-Way subsequently announced the restatement of its 1998 and 1999 financial statements on July 2, 2001. An additional 50 initial dates are third-party revelations, such as *The New York Times* article that broke the story on Computer Associates' financial fraud.

One category of initial revelations may suffer some look-ahead bias – the 35 initial revelations in the "Management change related to misconduct" category. As an example, on February 28, 2001 Rica Foods Inc. announced the resignation of its CFO, Randall Piedra. The firm also revealed that it "may have not met several negative covenants" in a loan agreement, but there is no explicit mention of accounting irregularities in the press announcement. As another example, on May 10, 2004 Buca Inc. announced the resignation of its Chairman and CEO. The firm also revealed a second quarter loss, the possible violation of debt covenants, and that it was re-evaluating its insurance reserves, but the connection between these revelations and financial misconduct did not become apparent until a February 7, 2005 press release announcing an SEC investigation. These isolated examples notwithstanding, most initial announcements in the "Management change related to misconduct" category contain explicit references to accounting irregularities. As a typical example, Phar-Mor Inc. announced the dismissal of co-founder and Vice-President Michael I. Monus on August 4, 1992, and also revealed that Federal agents had launched a criminal investigation into possible financial misconduct at the firm. To conclude, it is possible that a small number of the initial revelation dates identified by the SEC are subject to a look-ahead bias. But this number is so small as to be immaterial for our results. None of the results reported in this paper are materially affected if we omit the 35 cases in the "Management change…" category.

Table 4 also shows a primary reason for the late initial revelation date feature: the four popular databases rarely identify the actual initial revelation date. For the 1,099 cases in the hand-collected sample, the GAO database identifies the initial revelation event in 128 cases, the AA database identifies 50, the SCAC database identifies 30, and the AAER database identifies only 21.

5. Feature #2: Scope limitations

5.1. An illustration

Our earlier discussion of the financial misconduct case of Computer Associates identified 49 discrete dates upon which the market received new information (Figure 2). By design, each of the four databases captures only one type of event. So each database misses most of the 49 events in the

Computer Associates case. For example, the GAO database captures 8 of the 49 events (8 of CA's 9 restatement announcements related to its misconduct). The AA database contains only one of Computer Associates' nine relevant restatement announcements, so it covers an even smaller fraction of the 49 information events that describe the firm's fraud and its consequences. The SCAC database contains the filing and settlement dates for Computer Associates' class action lawsuit, representing two of the 49 information event days, and the AAER database captures ten of the 49 information event days. We call these scope limitations because each database's scope, by design, is limited to one type of event. Below, we show that a broader examination of the full sequence of events that reveal information about the misconduct frequently is required to accurately characterize the misconduct and decide whether it is appropriate for the research question at hand. For example, restatement or lawsuit announcements by themselves frequently do not reveal the consequences from SEC sanctions or the severity of the misconduct, and AAERs by themselves frequently do not reveal information about the consequences of any related class action lawsuits.

5.2. Extent of scope limitations in each of the four databases

Panel A of Table 5 summarizes the extent of the scope limitations for each database. The GAO database identifies 427 restatements that correspond to 290 unique cases in which 13(b) violations were prosecuted. This is an average of 1.5 GAO events per case. When we examine the full sequence of events in the 290 case histories associated with the GAO database, we detect 4,336 events – an average of 15.0 events per case. Thus, the GAO database misses 90.2% of the information events related to the GAO-detected cases of financial misrepresentation.

The AA database identifies 239 restatements that are associated with 188 unique cases of financial misconduct that involve 13(b) violations, an average of 1.3 restatements per case. In total, however, there are 2,738 relevant information events in these 188 case histories, an average of 14.6 events per case. This implies that the AA database misses 91.3% of the incremental information events associated with AA-detected cases of financial misrepresentation. Similarly, the SCAC database contains 389 lawsuit filings that are associated with 346 unique cases of financial misconduct that involve 13(b)

violations, an average of 1.1 SCAC lawsuit filings per case. There are a total of 6,556 relevant information events pertaining to these 346 cases, or an average of 19.0 events per case, implying that the SCAC database misses 94.1% of the relevant information events associated with the SCAC-detected misrepresentation cases. There are 2,865 AAERs involving 939 unique 13(b)-related cases (an average of 3.1 AAERs per case). There are a total of 7,919 information events related to these 939 cases, indicating the AAER database misses 63.8% of the information events relevant for understanding the AAER-detected cases of financial misrepresentation.

5.3. Relevance and economic importance of scope limitations

The additional information contained in each database's missing events can be important for researchers who seek to understand the nature of the misconduct, the full consequences to the firm, whether the misconduct is associated with other types of misconduct or charges (e.g., fraud), the number of respondents that are involved in the misconduct, or the penalties imposed on the various respondents. Such limitations can bias research inferences. Karpoff et al. (2008), for example, show that previous conclusions regarding managerial turnover around financial misconduct – which are based on restatement announcements, class action lawsuits, or AAERs – are changed if one considers the full sequence of informational events by which news of the misconduct is conveyed.

In this section we provide a broader measure of the economic importance of scope limitations by extending our previous event study application from section 4.3. For each firm, we compute the one-day market adjusted stock returns for all independent event dates pertaining to each case of misconduct, and then we cumulate the abnormal returns over all such dates. We find that, in terms of their impacts on firm value, the events missed by each database are more important than the events captured by the database. These returns results are reported in Panel B of Table 5. Beginning with the GAO database, returns data are available to compute the cumulated abnormal return for 265 of the 290 cases in the GAO database that trigger regulatory enforcement action for financial misrepresentation. The mean cumulated abnormal return is -7.82% with a median of -2.45%. By comparison, the mean cumulated abnormal return summing over all event dates in each of these 265 cases is -50.36%, with a median of -35.54%. The

differences are significant at the 1% level.[17] These results indicate the importance of the scope limitations. A conscientious researcher who relies upon the GAO data might carefully cull immaterial restatements and combine information from multiple restatements for each case of financial misconduct. Nonetheless, this researcher would not have information about the sequence of related announcements that, on average, constitute 84% of the value-related information about the misconduct and its discovery. Stated differently, even a conscientious user relying solely on the GAO data would capture an average of only 16% of the value-relevant information pertaining to the GAO database's cases of misconduct.

The results are similar for the other databases. The mean cumulated abnormal return over the restatement announcements in the AA database is -4.64%, with a median of -1.87%. The mean cumulated abnormal return summing over all event dates in each AA-identified case is -38.38% with a median of -26.79%. This implies that the AA database captures on average only 12% of the value-relevant information pertaining to the cases of misconduct that it identifies. The SCAC database captures on average only 10% of the value-relevant information pertaining to the cases of misrepresentation it identifies, and the AAER database captures, on average, 17% of the value-relevant information for the cases of misrepresentation it identifies.

Some researchers mitigate the scope limitation problem by augmenting their data with additional (typically hand-collected) data about the cases of misconduct they identify (e.g., see Dechow et al. 1996; Burns and Kedia 2006; Gleason et al. 2008; Dyck et al. 2010). Our results indicate that such effort frequently is required to avoid missing most of the value-relevant information that helps to characterize and classify the misconduct.

[17] An alternative measure of the cumulated abnormal return would compound the one-day returns, rather than add them. The results using such an alternative measure are similar to those reported.

6. Feature #3: *Potentially extraneous observations*

6.1. An illustration

On May 1, 2006 Samsonite Corporation, a distributor of luggage and computer cases, filed Form 12b-25 ("Notice of inability to timely file a Form 10-K, 10-KSB, or 10-KT") with the SEC. In the form, Samsonite announced its inability to timely file its Form 10-K for the fiscal year ended January 31, 2006 because it needed additional time to complete its analysis of its deferred tax asset valuation allowance. The firm also states that, "…depending upon the final results of this analysis…the Company may be required to restate its financial statements for 2003, 2004, and 2005." On May 9, 2006 Samsonite filed Form 8-K with the SEC in which it announced the firm had in fact discovered an error with its deferred tax asset valuation allowance and that its financial statements related to 2003 through 2005 should not be relied upon. The cumulative effect of the error resulted in a $5.7 million increase to the firm's accumulated deficit as of January 31, 2006. Both the GAO and Audit Analytics appropriately include this restatement in their databases – the GAO records the restatement announcement date as the Form 8-K filing date of May 9, 2006, while AA records the restatement "disclosure date" as the Form 12b-25 filing date of May 1, 2006.

Should the Samsonite restatement be included in samples of financial misconduct or fraud? This is a matter of judgment, whose answer depends on the particular research question. Many researchers conclude that the GAO, AA, SCAC, and AAER databases include events and cases that are inappropriate for their particular research questions, and cull their samples to remove extraneous events. For example, Hennes et al. (2008) identify 73.6% of the GAO restatements as "errors" and 26.4% as "irregularities" (corrections of intentional misreporting).[18] For a researcher seeking cases of misrepresentation, the restatements that are "errors" might be considered extraneous, i.e., events that should not be included in the sample. Similarly, Dyck et al. (2010) begin with 2,171 SCAC database cases, but cull their sample

[18] Specifically, Hennes et al. (2008) classify a restatement as an irregularity (as opposed to an error) if the restatement is announced using any variety of the words "fraud" or "irregularity"; the Securities and Exchange Commission or Department of Justice is conducting an investigation related to the restatement; or there are other investigations into the accounting matter such as the hiring of a forensic accounting firm (p. 1489).

down to 216 cases to eliminate small firms (assets less than $750 million), small settlements (less than $3 million), and other cases judged not to represent instances of meaningful misconduct – thus culling 90% of the initial sample. Dechow et al. (2011) motivate their investigation of AAERs in part by pointing out that the GAO and SCAC databases may contain many extraneous observations, and Dechow et al. (2010, p. 371) note that external indicators of financial misconduct such as restatements can reflect both intentional and unintentional misstatements. Other examples of papers in which researchers manually cull their samples to identify cases of misconduct or fraud include Erickson et al. (2006), Brazel et al. (2009), and Johnson et al. (2009).

6.2. Potentially extraneous observations in the four databases

Again, whether any individual case should or should not be included in a researcher's sample depends on the research question. To help guide such judgment, we propose two criteria that researchers can use to cull their samples, and report the fraction of cases that must be culled using each criterion. For the first criterion, we classify a case as relevant if it is associated with any SEC regulatory action for financial misrepresentation (i.e., a Section 13(b) violation). Cases not associated with a SEC sanction for financial misrepresentation are classified as extraneous. We recognize that some instances of financial misconduct may prompt restatements or lawsuits without triggering any SEC sanctions. Nevertheless, this screen has the advantages of objectivity and replicability, as it relies on SEC and DOJ classifications rather then the researcher's personal judgment.

The rates of extraneous observations using this criterion are presented in Panel A of Table 6. The GAO database identifies 2,707 events, but only 427 (15.8%) of these events are associated with cases of alleged misconduct in which regulators initiated an enforcement action for financial misrepresentation. By this screen, 2,280 (84.2%) of the GAO events are extraneous, implying that a researcher must remove 5.3 GAO events for every event that does identify a case of financial misrepresentation.

The AA database identifies 11,001 restatements, only 239 of which are related to cases in which regulators take enforcement action for financial misrepresentation. This implies that 97.8% of the cases are extraneous (e.g., researchers need to sift through 45 AA restatement announcement events for every

event that identifies a case of misrepresentation). Using the SCAC database, 3,032 of the 3,421 lawsuit filing events are not associated with cases that involve actions for financial misrepresentation by regulators, yielding an extraneous event rate of 88.6%. This measure is slightly lower than the culling rate of 90% reported by Dyck et al. (2010). By comparison, the percentage of extraneous events in the AAER database is only 19.7%. This does not mean that 80.3% of the AAERs directly refer to charges of financial misrepresentation. Rather, 80.3% of the AAERs are associated with SEC enforcement actions that include financial misrepresentation charges at some point in their case histories.[19]

Many researchers seek to examine not just instances of financial misrepresentation, but cases of financial *fraud*.[20] Fraud implies that the perpetrator intended to deceive stakeholders, and intent to deceive requires a higher burden of proof than is necessary to bring charges of financial misrepresentation under Section 13(b) of the Securities Exchange Act. We define a case as involving fraud if the SEC or DOJ file charges alleging the violation of: (i) Section 17(a) of the 1933 Securities Act for fraudulent interstate transactions related to the issuance of a security; or (ii) Section 10(b) of the 1934 Securities Exchange Act for manipulative and deceptive devices related to the trading of an already issued security. As reported in the last column of Panel C of Table 6, financial fraud charges are included in 821 (74.7%) of the 1,099 hand-collected cases.

Because financial fraud is less common than financial misrepresentation, the percentage of extraneous cases is even higher if a researcher uses the GAO, AA, SCAC, or AAER databases to identify financial misrepresentation *and* fraud. As reported in Panel C of Table 6, 89.4% of the cases identified by the GAO database is unrelated to financial fraud. The corresponding rate is 98.1% for the AA database, 90.4% for the SCAC database, and 46.2% for the AAER database. We acknowledge that financial fraud charges could accompany other types of misconduct that do not include Section 13(b) violations but

[19] The Samsonite case, discussed above, did not prompt financial misrepresentation charges by the SEC, so by our criterion this is an extraneous case. The Hennes et al. (2008) criteria also classify the Samsonite case as extraneous (or, using their terminology, as an "error").

[20] The term "fraud" has both colloquial and technical meanings. It is possible that researchers who use the term "fraud" intend merely to say that their samples consist of activities that may be illegal. We argue, however, that such informality adds ambiguity to the nature of the events in the sample. The Internet Appendix highlights several differences in the definition of fraud as used in the legal, finance, and accounting literatures.

might show up in the GAO, AA, or SCAC databases, and our fraud screen would inaccurately classify such cases as extraneous.

6.3. Relevance and economic importance of culling potentially extraneous observations

Panels D through F of Table 6 report on the economic importance of including non-misconduct or non-fraud events in one's sample when measuring market reactions. In Panel D, we report the one-day market-adjusted return using all events in each database. The GAO database, for example, has a total of 2,707 events, and only 427 of these events are associated with a 13(b) violation. Returns data are available for 389 of these 427 events, with a mean one-day abnormal return of -5.34%. The remaining 2,280 events (with returns data available for 2,017) are classified as extraneous events. The mean one-day abnormal return for these extraneous events is -1.36%. In Panel E, we repeat this experiment, but eliminate all follow-on events for each case of misconduct. This mimics the approach taken by researchers who discard follow-on observations for each case of misconduct (e.g., Graham et al. 2008; Chen et al. 2013). In Panel F, cases are flagged as extraneous if they do not involve charges of financial *fraud* by the SEC, as opposed to the less restrictive financial misrepresentation category used in Panels A and B. In all three panels, a similar pattern emerges for all four databases. The average abnormal return for the relevant events is significantly larger (at the 1% level) than the corresponding return for extraneous events. In each panel, however, the abnormal return for the extraneous sample is still negative and statistically significant, indicating that events in the extraneous sample are value-relevant to investors.

Lest our criteria for identifying extraneous cases (i.e., screening for the presence of 13(b) or financial fraud charges) seem strict, we emphasize two important points. First, these criteria are objective, easily replicable, and based on SEC and DOJ deliberations rather than the individual researcher's assessment. Second, these criteria appear to be *less* restrictive than the culling methods used in prior research. For example, Francis et al. (1994) use a sample of 45 class action filings from the SCAC database and report a mean one-day abnormal return of -17.2%. Ferris and Pritchard (2001) examine 89 lawsuit filings from the SCAC database and report a three-day abnormal stock return of -25.0%. In contrast, we find that the 301 of the 346 SCAC cases that correctly identify financial

misrepresentation that triggers SEC action are associated with a mean one-day abnormal return of only -5.42% (Panel E of Table 6). Similarly, Beneish (1999) reports a -20.2% three-day abnormal return using data from 54 firms identified by AAERs plus 10 additional firms identified from a media search. In contrast, our culled sample of AAERs has a much larger sample size of 652, and the associated mean one-day market-adjusted return is only -3.98% (Panel E of Table 6). These examples illustrate that ad hoc culling methods frequently lead to small samples that are biased toward more extreme cases of financial misconduct. For research designs that use control samples, such aggressive culling also increases the potential for misconduct cases to be erroneously classified as observations in the control firm pool.

7. Feature #4: Complete and partial data omissions

7.1. An illustration

Figure 3 illustrates the sequence of informational events for Professional Transportation Group, Ltd., Inc., a Georgia-based interstate trucking business that created fictitious sales in the late 1990s. On May 15, 2000, the firm announced a restatement of its 1999 first three quarterly earnings. The announcement attributes the restatement to staff resignations and a change in the firm's data processing system, which "…caused certain errors to occur in recording certain revenues…" On November 9, 2000, the firm announced that it also would restate earnings from 2000, and revealed that its auditors were "…reviewing certain revenue recognition matters…" On November 30, 2000, Professional Transportation removed its CEO and Board Chair and filed for bankruptcy. The first SEC regulatory action related to this financial fraud occurred nearly three years later on October 2, 2003. The SEC formally revoked the firm's securities registration on December 1, 2003, and the SEC issued two additional regulatory releases on April 7 and June 30, 2004.

The Professional Transport case illustrates a surprisingly pervasive attribute of the GAO, AA, SCAC, and AAER databases: complete and partial data omissions. For example, the GAO database completely misses this case of misconduct even though two restatements were announced in 2000 – during the GAO's coverage period from January 1997 through June 30, 2006. By its own sampling

design and sampling interval, the GAO database should have captured the Professional Transport case. Because it does not, we call this a completely omitted case.

Even when the databases do identify a case of misconduct, frequently the coverage is only partial. In the Professional Transportation case, the AA database identifies the May 15, 2000 restatement announcement but misses the November 9, 2000 restatement. We call this a partial omission of the Professional Transport case. Partial omissions can be important. In the Professional Transport case, the May 15, 2000 restatement would lead at least some researchers to conclude that the restatement was an "error" and not a case of fraud; many would even discard this case from their sample. The information in the second restatement announcement on November 9, 2000 is important for accurately classifying this as a case of financial fraud.

As the Professional Transportation case illustrates, complete and partial data omissions can affect how researchers classify and use their data. Researchers using the GAO or SCAC data would incorrectly treat Professional Transportation as a potential control firm, i.e., one without misconduct. Researchers using AA data would consider the Professional Transportation case, but because the May 15, 2000 restatement appears benign, probably would conclude that it is not a case of material misconduct or fraud. Researchers using the AAER data would identify this as an important fraud case. But because the first AAER was issued when the firm's stock was no longer actively traded, Professional Transportation might be dropped from the sample even though there is share price information when the firm's misconduct initially was revealed to the public.

7.2. Complete and partial data omissions in the four databases

Panel A of Table 7 summarizes the rates at which each of the four databases omits cases of financial misconduct that the database was designed to capture. As a benchmark, we use events associated with the 1,099 cases in the hand-collected sample. Again, all of these cases prompted SEC and DOJ enforcement action, so these are serious cases of misconduct most researchers in this area would like their databases to include. The GAO database aims to include all material restatement announcements for the January 1, 1997 through June 30, 2006 period. During this period there were 1,124 restatement

announcements pertaining to 417 unique cases of financial misconduct in the hand-collected sample. The GAO database includes 427 of these announcements and omits 697 – an omission rate of 62.0%. The omission rates for the other databases range from 17.4% (for the SCAC database) to 79.9% (for the AA database).

Panel A of Table 7 also breaks down the two main reasons for these omissions. Of the 697 restatement announcements that the GAO database omits, 219 are omitted because the GAO database omits all restatement announcements related to 127 cases of misconduct, and 478 are omitted because the GAO identifies some but not all of the material restatement announcements pertaining to 290 other cases of misconduct.

Panel B reports on the cases that each database completely omits. During the GAO's sampling period there were 417 cases of financial misconduct that had one or more restatement announcement. The GAO at least partially identifies 290 of these cases and completely omits 127 (and these 127 omitted cases contain 219 individual restatement announcements). This result indicates that researchers who rely on the GAO database could identify 290 (69.5%), of the financial misconduct cases with a restatement announcement during the GAO's sampling period. Even if a researcher were to carefully research the histories of these 290 cases, he or she still would miss 127 (30.5%) of the total number of serious financial misconduct cases that had a restatement announcement during this period. Notice that the rates of complete omission – using each database's sampling period and screen to identify misconduct – range from a low of 9.4% for the SCAC database to 53.9% for the AA database.

7.3. Relevance and economic importance of complete and partial data omissions

Many researchers identify their empirical tests by comparing the characteristics of firms identified in one of the four databases (i.e., treatment firms) with those of other (i.e., control) firms (e.g., Dechow et al. 1996; Erickson et al. 2006; Davidson et al. 2014). Such comparisons will be affected by which firms are placed into the treatment and control groups. As illustrated in the Professional Transport case, partial data omissions can affect how individual cases are characterized – a researcher using the AA database likely would classify Professional Transport as a candidate for the control group, when in fact it

is a case of financial fraud. An AAER user might simply delete this case from the sample because the first AAER was issued long after the firm was delisted. The likelihood of such misclassification is even greater when the case is completely omitted. For example, a user of the GAO database would almost certainly include Professional Transport as a candidate for a control group because the GAO database completely omits this case – even though there are two restatement announcements during the GAO's sampling period.

The results above in Section 7.2 provide evidence on the likelihood and potential importance of such misclassifications, as the complete omission rates are highest for the AA database and lowest for the SCAC database. In this section we provide additional evidence on the potential biases that can arise from complete omissions, by describing the characteristics of the cases that each database omits and includes. The results are summarized in Panels C and D of Table 7. The GAO database omits cases that involve smaller, less visible, and more financially troubled firms than the cases it captures. Relative to firms in GAO-included cases, firms in GAO-omitted cases are smaller and less likely to be covered by the CRSP and I/B/E/S databases, and they are almost twice as likely to have going concern qualifications from their auditors. The data also suggest the GAO-omitted cases involve less serious cases of misconduct (i.e., lower median regulatory fines and class action penalties and likely to include a fraud charge). The median initial stock price reaction of -9.3% to news of the misconduct is more negative for the included cases, but the median abnormal stock price drop of -6.1% for the omitted cases indicates that the omitted cases also involve meaningful misconduct.

Like the GAO database, the AAER database tends to omit cases that involve smaller and less visible firms than the cases it includes, and the omitted violations, while consequential, are not as costly as the included violations. The SCAC database has relatively few omitted cases, but Panels C and D reveal that omitted cases are less likely to be covered by the CRSP and I/B/E/S databases and face fewer regulatory proceedings and smaller securities class action lawsuit awards. In contrast, none of the firm or violation-related characteristics of the cases the AA database includes are significantly different from the

cases excluded, suggesting that coverage by the AA database is incomplete but not systematically biased according to firm characteristics.

Many researchers are aware of omissions in these databases and seek to augment their samples by searching for cases that are not included in the database they use (e.g., see Gleason et al. 2008; Burns and Kedia 2006). Our results indicate that such augmentation is essential because the data omissions and partial coverage problems are large. One potential problem is that data omissions can lead to biased tests, or tests on non-representative samples. Samples constructed from the GAO, SCAC, or AAER databases, for example, are slanted toward relatively large firms that appear in standard databases, and toward relatively large or costly cases of misconduct. For some research applications, of course, such a slant may be exactly what the researcher desires. For other research applications, this selection bias may yield inferences that do not generalize to the broader population of misconduct firms. Our findings should help researchers choose which database to use, as well as whether and how to augment the database with additional hand-collected data. A second problem posed by data omissions and partial coverage is the prospect of low power tests. For example, control samples may be constructed from pools of firms that include both misconduct and non-misconduct firms or firm-years. The low inclusion rates in the GAO, AA, and AAER databases indicate that this may be a particular concern for these databases. As Burgstahler (1987) demonstrates, low power tests increase the probability of observing seemingly significant test results even when the null hypothesis holds. Roberts and Whited (2012) show that measurement error, which can be introduced by complete and/or partial data omissions, typically result in biased coefficients and standard errors.

8. Conclusion

Research on the causes and consequences of financial misconduct has facilitated significant advances in our understanding of corporate governance, financial reporting, and the optimal regulation of financial markets. This paper makes several contributions to the field of financial misconduct research. We first document the complex nature of financial misconduct cases and show how four popular

databases used in most research capture only small parts of the total picture. We also explain how the complexity of most misconduct cases leads to four database challenges that researchers must navigate to avoid biased empirical tests: (1) late initial revelation dates, (2) scope limitations, (3) potentially extraneous events, and (4) complete and partial data omissions. To enable researchers to assess the potential severity of biases that can arise, we measure the magnitude and examine the economic relevance of each feature for all four databases.

The exact importance of each feature depends on the specific research question and the extent to which the required database is subject to each feature. Table 8 provides a summary of the extent to which each database is affected by each of these four features. Regarding the first feature (late initial revelation dates), the SCAC database fares the best – the database's first event occurs an average of 150 days after the initial public revelation of the misconduct. The AAER database fares the worst, as the first event in the database is an average of 1,017 days after the initial public revelation. When focusing on median lag values, the GAO database does the best (median 14 day lag) while the AAER database remains the worst (median 991 day lag). With respect to the second feature (scope limitations), the AAER database fares the best (64% of value-relevant events are omitted), while the SCAC database fares the worst (94% of value-relevant events are omitted). With respect to the rate at which the databases include potentially extraneous non-fraud events (the third feature), we find AAER database has lowest rate of non-fraud events (46%), while the AA database has the highest rate of non-fraud events (98%). The fourth database feature regards these databases' tendencies to omit events that each claims to capture during its sampling period. Here, the SCAC database omits the least same-type events (17%), while the AA database omits the most same-type events (80%).

These findings yield several insights for how these four important databases can be used effectively. First, simple awareness of the complexity of a typical financial misconduct case can guide the interpretation of empirical tests. For example, the share price reaction to a restatement (e.g., Burns and Kedia 2006) or class action lawsuit (e.g., Gande and Lewis 2009), or information transfer from a restatement (Gleason et al. 2008) does not reveal unconditionally new information about the potential

misconduct. Rather, a restatement or securities class action lawsuit is one event in a series of events that comprise a case of misconduct. Each event reveals information that partially resolves the uncertainty that remains from earlier revelations about the misconduct and the consequences to the firm from its discovery.

Second, our results underscore the importance of supplementing the data from any of these four databases with additional hand-collected data. Each database contains a large sample of a specific type of event (e.g., restatements, lawsuits, or AAERs) useful in identifying instances of possible financial misconduct. Additional information frequently is required to classify a case as involving financial fraud, identify important related events such as managerial turnovers, or identify the consequences to the firm.

Third, we advocate a more systematic and replicable approach to culling the events in these databases to identify instances of material financial misconduct or fraud.[21] Our results indicate that ad hoc culling to eliminate potentially extraneous events from a data sample can create its own biases, as our sample sizes after eliminating non-fraud observations substantially exceed the sample sizes that appear in many published papers using GAO, AA, SCAC, and/or AAER data. This finding is especially noteworthy because there are significantly fewer SEC enforcement action cases related to 13(b) violations and financial fraud relative to the number of observations included in each of these four datasets (see Panel A of Table 2). Our event study results indicate that the datasets used in many papers tend to be culled aggressively in ways that select extreme cases of misconduct with large stock price effects.

Fourth, the scope limitations and high rates of omission in these databases indicate that researchers must exercise care when constructing control samples. This is because control samples based on firms or firm-years that do not appear in the GAO, AA, SCAC, or AAER databases may include firms and firm-years that did, in fact, have restatements, lawsuits, or SEC enforcement activity. Among the four databases, the SCAC database has the lowest omission rate while the AA database has the highest

[21] Researchers also should be aware that the term "fraud" has multiple meanings, as discussed more thoroughly in the Internet Appendix to this paper. As shown in Panel C of Table 6, one-fourth of cases that prompt SEC enforcement action for financial misrepresentation do not include any fraud charges. Restatement announcements are associated with fraud charges even less frequently.

omission rate. Researchers can further avoid tainted control samples and biased test results by accessing other data sources to verify that their candidate control firms do not have contemporaneous restatements, lawsuits, and/or SEC enforcement actions.

Research on financial misconduct is compelling and insightful for many reasons. It informs our understanding of how markets, firm governance, and regulation work – or sometimes do not work – to discipline and deter opportunistic behavior and fraud. As such, this research provides essential insight into when and how Adam Smith's Invisible Hand works to facilitate the use of external finance, the growth of financial markets, and economic growth. The main takeaway from this paper is that cases of financial misconduct typically involve a complex sequence of interrelated events, all of which frequently are necessary to accurately characterize and measure the consequences of the misconduct. We investigate the impact of failing to consider the fuller sequence of events when measuring the valuation impact of misconduct, and find that the resulting measurement error is economically significant.

We encourage researchers to continue investigating the causes and impacts of financial misconduct. We also hope that the documentation and calibration of the four database features in the GAO, AA, SCAC, and AAER databases, as well as the suggestions on ways to manage these features, will help researchers use these important data in well-specific empirical tests.

References

Aiken, A. L., C. P. Clifford, and J. Ellis, 2013. Out of the Dark: Hedge Fund Reporting Biases and Commercial Databases. *Review of Financial Studies* 26: 208-243.

Ali, A., S. Klasa, and E. Yeung, 2009. The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research. *Review of Financial Studies* 22: 3839-3871.

Anderson, R. C., and D. S. Lee, 1997. Ownership Studies: The Data Source Does Matter. *Journal of Financial and Quantitative Analysis*, 32: 311-329.

Beneish, M., 1999. Incentives and Penalties Related to Earnings Overstatements that Violate GAAP. *The Accounting Review* 74: 425-457.

Bennin, R., 1980, Error Rates in CRSP and COMPUSTAT: A Second Look. *Journal of Finance* 35: 1267-1271.

Berenson, A., 2001. "A Software Company Runs Out of Tricks: The Past May Haunt Computer Associates," *New York Times* (April 29, 2001), p. BU1 and 11.

Brazel, J., K. Jones, and M. Zimbelman, 2009. Using Nonfinancial Measures to Assess Fraud Risk. *Journal of Accounting Research* 47: 1135-1166.

Burgstahler, D., 1987. Inferences from Empirical Research. *The Accounting Review* 62: 203-214.

Burns, N., and S. Kedia, 2006. The Impact of Performance-Based Compensation on Misreporting. *Journal of Financial Economics* 79: 35-67.

Canina, L., R. Michaely, R. Thaler, and K. Womack, 1998. Caveat Compounder: A Warning about Using the Daily CRSP Equal-Weighted Index to Compute Long-Run Excess Returns. *Journal of Finance* 53: 403-416.

Chen, X., Q. Cheng, and A. Lo, 2013. Accounting Restatements and External Financing Choices. *Contemporary Accounting Research* 30(2): 750-779.

Cheng, C. S. A., H. Huang, Y. Li, and G. J. Lobo, 2010. Institutional Monitoring through Shareholder Litigation. *Journal of Financial Economics* 95: 356-383.

Chuk, E., D. Matsumoto, and G. Miller, 2013. Assessing Methods of Identifying Management Forecasts: CIG vs. Researcher Collected. *Journal of Accounting and Economics* 55: 23-42.

Davidson, R., A. Dey, and S. Smith, 2014. Executives' "Off-the-Job" Behavior, Corporate Culture, and Financial Reporting Risk. *Journal of Financial Economics* (forthcoming).

Dechow, P., W. Ge, C. Larson, and R. Sloan, 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28: 17-82.

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: 1-36.

Dyck, A., A. Morse, and L. Zingales, 2010. Who Blows the Whistle on Corporate Fraud? *Journal of Finance* 65: 2213-2253.

Elton, E. J., M. Gruber, and C. R. Blake, 2001. A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases. *Journal of Finance* 56: 2415-2430.

Erickson, M., M. Hanlon, and E. Maydew, 2004. How Much Will Firms Pay for Earnings That Do Not Exist? Evidence of Taxes Paid on Allegedly Fraudulent Earnings. *The Accounting Review* 79: 387-408.

Erickson, M., M. Hanlon, and E. Maydew, 2006. Is There a Link between Executive Equity Incentives and Accounting Fraud? *Journal of Accounting Research* 44: 113-143.

Feroz, E., Park, K., V. Pastena, 1991. The Financial and Market Effects of the SEC's Accounting and Auditing Enforcement Releases. *Journal of Accounting Research* 29: 107-142.

Ferris, S., and A. Pritchard, 2001. Stock Price Reactions to Securities Fraud Class Actions under the Private Securities Litigation Reform Act. Michigan Law and Economics Research Paper No. 01-009. Available at SSRN: http://ssrn.com/abstract=288216.

Fich, M., and A. Shivdasani, 2007. Financial Fraud, Director Reputation, and Shareholder Wealth. *Journal of Financial Economics* 86: 306–336.

Francis, J., D. Philbrick, and K. Schipper, 1994. Shareholder Litigation and Corporate Disclosures. *Journal of Accounting Research* 32: 137-164.

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

Gillan, S., J. Hartzell, A. Koch, and L. Starks, 2013. Getting the Incentives Right: Backfilling and Biases in Executive Compensation Data. Available at SSRN: http://ssrn.com/abstract=2247055.

Gleason, C., N. Jenkins, and W. Johnson, 2008. Financial Statement Credibility: The Contagion Effects of Accounting Restatements. *The Accounting Review* 83: 83-110.

Government Accountability Office (GAO), 2002. Financial Statement Restatements: Trends, Market Impacts, Regulatory Responses, and Remaining Challenges. Washington, D.C. GAO-03-138.

Government Accountability Office (GAO), 2003. Financial Restatement Database. Washington, D.C. GAO-03-395R.

Government Accountability Office (GAO), 2006a. Financial Restatements: Update of Public Company Trends, Market Impacts, and Regulatory Enforcement Activities. Washington, D.C. GAO-06-678.

Government Accountability Office (GAO), 2006b. Financial Restatement Database. Washington, D.C. GAO-06-1053R.

Graham, J., S. Li, and J. Qiu, 2008. Corporate Misreporting and Bank Loan Contracting. *Journal of Financial Economics* 89: 44-61.

Harris, R. S., T. Jenkinson, and S. N. Kaplan, 2014. Private Equity Performance: What Do We Know? *Journal of Finance*, forthcoming.

Helland, E., 2006. Reputational Penalties and the Merits of Class-Action Securities Litigation. *Journal of Law and Economics* 49(2): 365-395.

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

Hribar, P., N. Jenkins, 2004. The Effect of Accounting Restatements on Earnings Revisions and the Estimated Cost of Capital. *Review of Accounting Studies* 9: 337-356.

Johnson, S., H. Ryan, and Y. Tian, 2009. Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter. *Review of Finance* 13: 115-145.

Kahle, K. M., and R. A. Walkling, 1996. The Impact of Industry Classifications on Financial Research. *Journal of Financial and Quantitative Analysis* 31: 309-335.

Karpoff, J. M., D. S. Lee, and G.S. Martin, 2008. The Consequences to Managers for Financial Misrepresentation. *Journal of Financial Economics* 88: 193-215.

Kravet, T., and T. Shevlin, 2010. Accounting Restatements and Information Risk. *Review of Accounting Studies* 15: 264-294.

Ljungqvist, A., C. Malloy, and F. Marston, 2009. Rewriting History. *Journal of Finance* 64: 1935-1960.

McNichols, M., S. Stubben, 2008. Does earnings management affect firms' investment decisions? *The Accounting Review* 83(6): 1571-1603

Roberts, M. and T. Whited, 2012. Endogeneity in Empirical Corporate Finance, in George Constantinides, Milton Harris, and Rene Stulz, eds. *Handbook of the Economics of Finance* Volume 2, Elsevier.

Rosenberg, B. and M. Houglet, 1974. Error Rates in CRSP and Compustat Databases and Their Implications. *Journal of Finance* 29: 1303-1310.

Schrand, C., and S. Zechman, 2012. Executive Overconfidence and the Slippery Slope to Financial Misreporting. *Journal of Accounting and Economics* 53: 311-329.

Shumway, T., 1997, The Delisting Bias in CRSP Data. *Journal of Finance* 52: 327-340.

Shumway, T., and V. A. Warther, 1999. The Delisting Bias in CRSP's NASDAQ Data and Its Implications for Interpretation of the Size Effect. *Journal of Finance* 54: 2361-2379.

Table 1: Overview of four databases commonly used to study financial misconduct

This table describes the four major databases commonly used to identify financial misconduct: the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs). In our analysis, the AA and SCAC databases are truncated at December 31, 2010 and the AAER database is truncated at December 31, 2011.

Database	Type of Events	Who Maintains the Data	Description	Type of Trigger for Database Inclusion	Time Period Covered
GAO	Financial statement restatement announcements	Government Accountability Office	Restatement announcements compiled at the behest of Senator Sarbanes to determine the number of and reasons for financial statement restatement announcements	Lexis-Nexis keyword search for variants of "restate"	Jan 1, 1997 - June 30, 2006
AA	Financial statement restatement and non-reliance filings	Audit Analytics	Financial restatement and non-reliance filings disclosed in public filings by SEC registrants (primarily from SEC's EDGAR)	Financial restatement and non-reliance filings made by SEC registrants	Jan 1, 2000 - Dec 31, 2010
SCAC	Securities class action lawsuits	Stanford Securities Class Action Clearinghouse	Federal class action securities litigation, including prosecutions, defenses, and settlements, as well as supporting documents	Federal Court lawsuit for alleged violation of federal securities laws filed by shareholders	Jan 1, 1996 - Dec 31, 2010
AAER	Accounting and Auditing Enforcement Releases	Securities Exchange Commission	Secondary designation assigned by the SEC to administrative proceedings or litigation releases that involve, or are expected to be of interest to, accountants	SEC administrative and litigation releases announcing enforcement actions involving, or expected to be of interest to, accountants	July 2, 1975 - Dec 31, 2011

Table 2: Number of events and cases in each database and the hand-collected sample

Panel A reports the number of events and unique cases of misconduct in each database and the number of events and cases associated with 13(b) violations. The number of events exceeds the number of cases because some cases have more than one associated event. Panel B reports the composition of events in the hand-collected (HC) data. The "Number of events" columns associate each of the 10,415 events corresponding to 8,787 unique event dates in the hand-collected data. There are three types of SEC Enforcement Releases: (i) Litigation Releases concerning civil injunctive actions; (ii) Administrative Proceedings consisting of releases issued under authority granted by the Securities Act of 1933, Securities Exchange Act of 1934, Investment Company Act of 1940, Investment Advisors Act of 1940, or the Public Utility Holding Company Act of 1935; and (iii) Administrative Law Judges' Initial Decisions and Orders. The "Other regulatory events" category includes Self-Regulatory Organizations (SRO) trading halts, SEC news releases, and Department of Justice (DOJ) civil and criminal lawsuit filings and decisions. The "Other press releases and material announcements" category includes the initial announcements of the misconduct that prompt enforcement action and announcements of informal inquiries, formal investigation and Wells Notices from the SEC. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's series of Accounting and Auditing Enforcement Releases (AAERs). The hand-collected (HC) data includes all cases from 1978–2011 for which the SEC and/or Department of Justice brought action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. The Internet Appendix provides details regarding the event-to-case transformations and matching of the four commonly used databases with the hand-collected data.

Panel A: Events and cases in each database	**GAO**	**AA**	**SCAC**	**AAER**	**HC**
Number of events in the database	2,707	11,001	3,421	3,568	10,415
Number of unique cases in the database	2,321	8,358	3,116	1,356	1,099
Number of events associated with cases with a 13(b) violation for financial misrepresentation	427	239	389	2,865	10,415
Number of unique cases with a 13(b) violation for financial misrepresentation	290	188	346	939	1,099

Panel B: Composition of the hand-collected (HC) data		Number of events		Unique event dates	
Event type:		Total	Per case	Total	Per case
(a)	Restatement announcements	1,442	1.31	1,104	1.00
(b)	Securities class action lawsuit filings	615	0.56	574	0.52
(c)	Securities class action lawsuit settlements	630	0.57	617	0.56
	SEC Enforcement Releases				
(d)	- That include an AAER designation	3,066	2.79	2,118	1.93
(e)	- That do not include an AAER designation	1,445	1.31	1,179	1.07
(f)	Other regulatory events	1,298	1.18	934	0.85
(g)	Other press releases and material announcements	1,919	1.75	1,603	1.46
Event dates with more than one type of event:					
	(a + g)			274	0.25
	(d + f)			183	0.17
	(d + e)			75	0.07
	(e + f)			44	0.04
	(b + g)			22	0.02
	Other combinations of event types with the same date			60	0.05
Totals		10,415	9.48	8,787	8.00

Table 3: Feature #1 – Late initial revelation dates

This table reports the extent of the difference between the initial date reported by each database and first date the misconduct was publicly revealed (feature #1) and the effect of late initial revelation dates on abnormal stock return estimates associated with the revelation of financial misconduct. Panel A reports the difference measured in calendar days between the first date captured by the database and the initial revelation of the associated financial misconduct case. Panel B reports the effect of late initial revelation dates on one-day abnormal returns. The results are based on the subset of all events and cases in each database associated with an instance of misconduct that prompted enforcement action by the SEC and/or Department of Justice for financial misrepresentation under section 13(b) of the Securities Exchange Act of 1934. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs). The hand-collected (HC) data include all cases from 1978–2011 for which the SEC and/or Department of Justice brought action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. *** indicates statistical significance using two-tailed t-tests at the 1% level.

	GAO	**AA**	**SCAC**	**AAER**	**HC**
Panel A: Late initial revelation dates (feature #1)					
Number of cases with a 13(b) violation (from Table 2)	290	188	346	939	1,099
Number of days by which the initial event in the database lags the initial revelation of the misconduct					
Mean	187	242	150	1,017	–
Min	-3	-3	-3	-1	–
P25	0	0	2	594	–
P50	14	66	23	991	–
P75	218	310	153	1,399	–
Max	2,242	2,109	2,118	3,286	–

Table 3: Feature #1 – Late initial revelation dates (continued)

	GAO	AA	SCAC	AAER	HC
Panel B: Valuation effect of late initial revelation dates (feature #1) *One-day market-adjusted returns for the initial event date for cases that include SEC action for a 13(b) violation*					
Events identified by the database (from Table 2)	427	239	389	2,865	10,415
Cases identified by the database (from Table 2)	290	188	346	939	1,099
CRSP data available for the initial event in each case	260	137	300	637	944
One-day abnormal return using the initial event date provided by the database:					
Mean	-7.06%***	-4.83%***	-5.43%***	-4.03%***	-14.91%***
Median	-2.13%***	-1.67%***	-1.21%***	-1.13%***	-7.80%***
One-day abnormal return using the earlier initial event date identified by the hand-collected (HC) data:					
Mean	-16.17%***	-13.59%***	-18.64%***	-14.69%***	-14.91%***
Median	-9.31%***	-7.97%***	-13.55%***	-7.38%***	-7.80%***
Difference:					
Mean	9.11%***	8.76%***	13.22%***	10.67%***	–
Median	1.31%***	2.00%***	6.95%***	4.53%***	–
Percentage by which the database understates the initial revelation date share value reaction:					
Using means	56%	64%	71%	73%	–
Using medians	77%	79%	91%	85%	–

Table 4: Information contained in the initial revelation announcements

This table summarizes the type of information reported at the initial revelations of the 1,099 cases of financial misrepresentation in the hand-collected sample. A total of 795 initial revelation announcements are explicitly identified in SEC and DOJ regulatory proceedings (see column 3). The remaining 304 initial revelation announcements are described in column 4. "Restatement announcements" include all initial revelation events that include press announcements and/or 8-K filings that reveal restatements, except for the 12 events in the "Combination: restatement and class action filing" category, which have both a restatement and a class action lawsuit filing in the same trading day. "Combination: multiple firm revelations" are announcements of several types of adverse revelations by the firm, including accounting irregularities or internal investigations of accounting irregularities; and impairment charges, severe earnings shortfalls, inadequate financial reserves, and debt covenant violations associated with accounting irregularities. "Related regulatory activity, not by SEC or DOJ" includes announcements that arise from investigations or lawsuits by other federal or state agencies regarding non-financial misconduct, such as bribery or environmental violations. Some of these events also include an initial SEC regulatory action within the same trading day, including two SEC regulatory actions that also received AAER designations. "Unusual trading, trading suspension, or trading halt" includes unusual trading patterns and exchange-initiated trading halts (the 13 items in column 3) and trading suspensions initiated by the SEC (the 13 items in column 4). The 8 announcements in the "SEC trigger and regulatory response" category involve instances in which the SEC identified a trigger event and initiated an inquiry, investigation, or regulatory action in the same trading day as the trigger event. The databases in column (5) are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs).

(1)	(2)	(3)	(4)	(5)			
		Initial revelation identified in SEC and DOJ releases	Initial revelation, but not identified in SEC or DOJ releases	Number of initial revelation events captured by each database:			
	Total			GAO	AA	SCAC	AAER
Restatements and class action lawsuit filings:							
Restatement announcement	284	203	81	121	44		
Class action lawsuit filed	61	7	54			22	
Combination: restatement and class action filing	12	12		7	1	8	2
Firm and third-party revelations:							
Firm announces internal investigation	146	146			4		
Firm reveals an SEC inquiry into misconduct	36		36		1		
Firm reveals an SEC investigation of misconduct	57		57				
Firm discloses the receipt of a Wells Notice	8		8				
Auditor change or withdrawal related to misconduct	39	39					
Management change related to misconduct	35	35					
Bankruptcy	16	16					
Combination: multiple firm revelations	142	142					
Third-party revelation (e.g., newspaper)	50	50					
Whistleblower	19	19					
Regulatory activity:							
By-product of regulatory activity, not by SEC or DOJ	64	64					
Regulatory action taken by the SEC	55		55				19
Delayed filings or SEC review of filings	41	41					
Unusual trading, trading suspension, or trading halt	26	13	13				
SEC trigger and regulatory response	8	8					
Total	**1099**	**795**	**304**	128	50	30	21

Table 5: Feature #2 – Scope limitations

This table reports the extent to which the databases fail to include events associated with cases due to each database's scope limitation (feature #2) and the effect of scope limitations on abnormal stock return estimates associated with the revelation of financial misconduct. Panel A reports the extent to which each of the databases provide an incomplete record of the events associated with a financial misconduct case due to scope limitations. Panel B reports the effect of scope limitations on long-window abnormal returns cumulated over all relevant event dates. The results are based on the subset of all events and cases in each database associated with an instance of misconduct that prompted enforcement action by the SEC and/or Department of Justice for financial misrepresentation under section 13(b) of the Securities Exchange Act of 1934. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs). The hand-collected (HC) data include all cases from 1978–2011 for which the SEC and/or Department of Justice brought action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. *** indicates statistical significance using two-tailed t-tests at the 1% level.

	GAO	**AA**	**SCAC**	**AAER**	**HC**
Panel A: Extent of incomplete records due to scope limitations (feature #2)					
Number of cases with a 13(b) violation (from Table 2)	290	188	346	939	1,099
Number of events associated with these cases (from Table 2)	427	239	389	2,865	10,415
Average number of events per case	1.47	1.27	1.12	3.05	9.48
Number of all types of informational events related to these cases	4,336	2,738	6,556	7,919	10,415
Average number of events per case	14.95	14.56	18.95	8.43	9.48
% of all types of events the database captures	9.8%	8.7%	5.9%	36.2%	100.0%
% of all types of events the database misses	90.2%	91.3%	94.1%	63.8%	–
Total	100%	100%	100%	100%	100%

Table 5: Feature #2 – Scope limitations (continued)

	GAO	AA	SCAC	AAER	HC
Panel B: Valuation effect of scope limitations (feature #2) *Abnormal returns cumulated over all event dates for cases that include SEC action for a 13(b) violation*					
Cases identified by the database (from Table 2)	290	188	346	939	1,099
CRSP data available for at least one event per case	265	138	300	673	968
Cumulative abnormal return using all event dates for each case available in the database:					
Mean	-7.82%***	-4.64%***	-5.61%***	-7.49%***	-39.93%***
Median	-2.45%***	-1.87%***	-1.27%***	-1.89%***	-25.51%***
Cumulative abnormal return using all event dates for the same case available in the hand-collected (HC) database:					
Mean	-50.36%***	-38.38%***	-57.41%***	-44.38%***	-39.93%***
Median	-35.54%***	-26.79%***	-42.33%***	-29.36%***	-25.51%***
Difference:					
Mean	42.54%***	33.73%***	51.80%***	36.88%***	–
Median	28.03%***	21.26%***	38.22%***	22.89%***	–
Percentage by which the database understates the total change in share value:					
Using means	84%	88%	90%	83%	–
Using medians	93%	93%	97%	94%	–

Table 6: Feature #3 –Potentially extraneous observations

This table reports the number and frequencies of non-misconduct and non-fraud samples by database (feature #4) and the market implications of failing to exclude extraneous events. Panel A reports the *events* included in each database that are not associated with a case in which the SEC brings an enforcement action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. Panel B reports the *cases* included in each database that are not associated with the SEC bringing an enforcement action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. Panel C reports the *cases* that are not associated with a securities fraud charge under Section 17(a) of the Securities Act of 1933 or 10(b) of the Securities Exchange Act of 1934. Panel D reports the differing market implications of failing to cull *events* unassociated with a Section 13(b) violation of the Securities Exchange Act of 1934. Panel E reports the differing market implications of failing to cull *cases* unassociated with a Section 13(b) violation of the Securities Exchange Act of 1934. Panel F reports the differing market implications of failing to cull *cases* unassociated with a securities fraud charge under Section 17(a) of the Securities Exchange Act of 1934. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's series of Accounting and Auditing Enforcement Releases (AAERs). The hand-collected (HC) data include all cases from 1978–2011 for which the SEC and/or Department of Justice brought action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. ***, ** and * indicate statistical significance at $p < 0.001$, $p < 0.01$, and $p < 0.10$.

	GAO	**AA**	**SCAC**	**AAER**	**HC**
*Panel A: Identifying non-misconduct events using **all events** in each database*					
Events in the database (from Table 2)	2,707	11,001	3,421	3,568	10,415
Events associated with 13(b) violation cases (from Table 2)	(427)	(239)	(389)	(2,865)	(10,415)
Non-financial misconduct events	2,280	10,762	3,032	703	–
% of non-financial misconduct events	84.2%	97.8%	88.6%	19.7%	–
*Panel B: Identifying non-misconduct events using **unique cases** in each database*					
Cases in the database (from Table 2)	2,321	8,358	3,116	1,356	1,099
Cases associated with 13(b) violation cases (from Table 2)	(290)	(188)	(346)	(939)	(1,099)
Non-financial misconduct cases	2,031	8,170	2,770	417	–
% of non-financial misconduct cases	87.5%	97.8%	88.9%	30.8%	–
*Panel C: Identifying **non-fraud cases** in each database*					
Cases in the database (from Table 2)	2,321	8,358	3,116	1,356	1,099
Cases associated with a fraud charge	(246)	(155)	(300)	(729)	(821)
Non-fraud cases	2,075	8,203	2,816	627	278
% of non-fraud cases	89.4%	98.1%	90.4%	46.2%	25.3%

Table 6: Feature #3 –Potentially extraneous observations (continued)

	GAO	AA	SCAC	AAER	HC
*Panel D: Effect of extraneous events on market returns using **all events** in each database*					
Misrepresentation events (associated with a Section 13(b) violation case)					
Events in the database (from Table 2)	427	239	389	2,865	10,415
Events with CRSP data	389	166	334	1,571	6,545
Mean one-day abnormal return	-5.34%***	-3.86%***	-5.05%***	-3.21%***	-6.52%***
Median one-day abnormal return	-1.34%***	-1.38%***	-1.14%***	-0.96%***	-1.58%***
Extraneous events (unassociated with section 13(b) violation cases)					
Events in the database (from Table 2)	2,280	10,762	3,032	703	–
Events with CRSP data	2,017	4,352	2,426	136	–
Mean one-day abnormal return	-1.36%***	-0.69%***	-0.82%***	-2.83%*	–
Median one-day abnormal return	-0.39%***	-0.26%***	-0.25%***	-0.46%***	–
*Panel E: Effects of extraneous cases using the **first event** per case for all unique **cases** in each database*					
Misrepresentation cases (associated with a Section 13(b) violation)					
Cases in the database (from Table 2)	290	188	346	939	1,099
Cases with CRSP data	264	137	301	652	944
Mean one-day abnormal return	-6.93%***	-4.83%***	-5.42%***	-3.98%***	-14.91%***
Median one-day abnormal return	-2.08%***	-1.67%***	-1.23%***	-1.13%***	-7.80%***
Extraneous cases (unassociated with a Section 13(b) violation)					
Cases in the database (from Table 2)	2,031	8,170	2,770	417	–
Cases with CRSP data	1,785	3,359	2,251	100	–
Mean one-day abnormal return	-1.38%***	-0.66%***	-0.90%***	-3.07%*	–
Median one-day abnormal return	-0.42%***	-0.27%***	-0.25%***	-0.33%*	–
*Panel F: Effects of extraneous non-fraud cases using the **first event** per case for all unique **cases** in each database*					
Fraud cases (associated with both a section 13(b) violation and a Section 17(a) financial fraud charge)					
Cases in the database	246	155	300	729	821
Cases with CRSP data	222	108	258	478	692
Mean one-day abnormal return	-7.42%***	-5.08%***	-6.03%***	-4.60%***	-17.34%***
Median one-day abnormal return	-2.28%***	-1.73%***	-1.25%***	-1.19%***	-10.39%***
Extraneous cases (unassociated with both a Section 13(b) violation and a Section 17(a) financial fraud charge)					
Cases in the database	2,075	8,203	2,816	627	278
Cases with CRSP data	1,827	3,388	2,294	274	252
Mean one-day abnormal return	-1.45%***	-0.69%***	-0.92%***	-2.56%***	-8.23%***
Median one-day abnormal return	-0.43%***	-0.28%***	-0.27%***	-0.72%***	-3.13%***

Table 7: Feature #4 – Complete and partial data omissions

This table reports the percentage of omitted cases and events associated with each database and firm characteristic differences between included versus omitted cases and events. Panel A reports the number of omitted cases of financial misconduct that occur during the sample period the database covers with a type of event the database purports to capture. Panel B reports the number of omitted same-type events within the cases the database accurately identifies but only provides partial coverage. Panel C compares select firm characteristics between the included versus omitted cases from Panel A, with ** indicating the values are significantly different between the groups at the 5% level. The omission rates are documented using the subset of all events and cases in each database associated with an instance of misconduct that prompted enforcement action by the SEC and/or Department of Justice for financial misrepresentation under section 13(b) of the Securities Exchange Act of 1934. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs). The hand-collected (HC) data include all cases from 1978–2011 for which the SEC and/or Department of Justice brought action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. To be considered outside of the range captured by the relevant sample period we use the following criteria: for GAO if the regulatory enforcement action was completed prior to the beginning of the sample period or if the violation end date is after the end of the sample period (Regulatory End Date is less than January 1, 1997 or Violation End Date is greater than June 30, 2006); for SCAC if the regulatory enforcement action was completed prior to the beginning of the sample period (Regulatory End Date is less than January 1, 1996); and for AA if the regulatory enforcement action was completed before the beginning of the sample period (Regulatory End Date is less than January 1, 2000).

	GAO		AA		SCAC		AAER	
	N	%	N	%	N	%	N	%
Panel A: Total number of omitted events								
Total relevant (same-type) events within the database's sample period	1,124		1,187		468		5,705	
Number identified by the database	427	38.0%	239	20.1%	389	83.1%	2,865	50.2%
Number omitted by the database	697	62.0%	948	79.9%	79	16.9%	2,840	49.8%
	1,124		1,187		468		5,705	
Omitted because the misconduct case is completely missed	219		553		43		649	
Omitted because the misconduct case is partially missed	478		395		36		2,191	
	697		948		79		2,840	
Panel B: Completely omitted cases								
Total cases with one or more same-type event within the sample period	417		408		382		1,099	
Cases at least partially identified by the database (from Table 2)	290	69.5%	188	46.1%	346	90.6%	939	85.4%
Cases completely omitted even though there are same-type events	127	30.5%	220	53.9%	36	9.4%	160	14.6%
	417		408		382		1,099	

Table 7: Feature #4 – Complete and partial data omissions (continued)

	GAO		AA		SCAC		AAER	
	Included cases (n=290)	Omitted cases (n=127)	Included cases (n=188)	Omitted cases (n=220)	Included cases (n=346)	Omitted cases (n=34)	Included cases (n=939)	Omitted cases (n=160)
Panel C: Firm characteristic differences: included vs. omitted cases								
Ln (market capitalization in $M)	6.31	4.18**	6.1	6.24	6.35	6.07	4.95	3.9**
Ln (assets in $M)	6.24	4.51**	6.2	6.16	6.16	6.24	4.94	4.23**
Market-to-book ratio	1.61	1.69**	1.46	1.69	1.72	1.45	1.67	1.48
Firm is in CRSP database	97%	68%**	89%	85%	95%	86%**	88%	75%**
Firm is in I/B/E/S database	93%	62%**	85%	82%	92%	81%**	71%	59%**
"Going concern" language in most recent audit	16%	31%**	18%	16%	13%	11%	22%	27%
Panel D: Violation characteristic differences: included vs. omitted cases								
Initial revelation day abnormal return	–9.3%	–6.1%**	–7.1%	–7.6%	–11.3%	–3.5%**	–8.7%	–2.5%**
% with at least one fraud charge	85%	79%**	82%	83%	86%	81%	78%	59%**
Number of U.S. code violations	11	11	12	11.5	12	11	11	9**
Number of regulatory proceedings	4	3**	4	4	4	3**	3	2**
Ln (regulatory fine), $M	12.6	11.5**	12.6	12.6	13	11.9	11.8	12.4
Ln (class action award), $M	16.3	15.1**	16.1	16.2	16.11	13.91**	15.9	14.3**

Table 8: Summary of how each database performs for each of the four features examined

This table summarizes the performance of four popular databases along each of four database features analyzed in this paper. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs). Late initial revelation dates refers to the time between the initial public revelation of a case of financial misconduct and the initial event in the database related to that case. Scope limitations refer to the fact that each database, by design, excludes value-relevant announcements that affect how the case of misconduct is classified or used. Potentially extraneous events refer to the fact that, for many research designs, the database include events that must be culled to select an appropriate sample. Omission rates refer to omitted same-type events (i.e., restatement announcements for the GAO and AA databases, securities class action lawsuit filings for the SCAC database, and SEC enforcement releases for the AAER database) during the database's sampling period for cases of financial misconduct that were serious enough to prompt SEC enforcement penalties for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934.

	#1		#2	#3	#4
Feature	Late initial revelation dates		Scope limitations	Potentially extraneous events	Omission rates
Description	Number of days the first event identified by the database lags the initial public revelation		Percentage of missed value-relevant events	Percentage of cases to be culled if seeking financial fraud	Percentage of missed events within the database's sampling period
Relative Performance	Mean	Median			
Best	SCAC (150 days)	GAO (14 days)	AAER (64%)	AAER (46%)	SCAC (17%)
	GAO (187 days)	SCAC (23 days)	GAO (90%)	GAO (89%)	AAER (50%)
	AA (242 days)	AA (66 days)	AA (91%)	SCAC (90%)	GAO (62%)
Worst	AAER (1,017 days)	AAER (991 days)	SCAC (94%)	AA (98%)	AA (80%)

Figure 1. Time periods covered by each of the databases

The timeline depicts the time periods covered by the databases that are commonly used to identify or examine financial misconduct or restatements. The databases are the Government Accountability Office (GAO) database of financial restatement announcements, the Audit Analytics (AA) database of financial restatement and non-reliance filings, the Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, the Securities and Exchange Commission's Accounting and Auditing Enforcement Releases (AAERs). The hand-collected (HC) data include all cases from 1978-2011 for which the SEC and/or Department of Justice brought action for financial misrepresentation under Section 13(b) of the Securities Exchange Act of 1934. The AA and SCAC databases are truncated at December 31, 2010 and the AAER and HC databases are truncated at December 31, 2011 for purposes of our analyses. *The dashed segment of the AAER bracket reflects the fact that AAER-1 retroactively reported on releases that would have received a secondary AAER designation from September 24, 1971 through April 15, 1982 if the designation had been in place during that time.

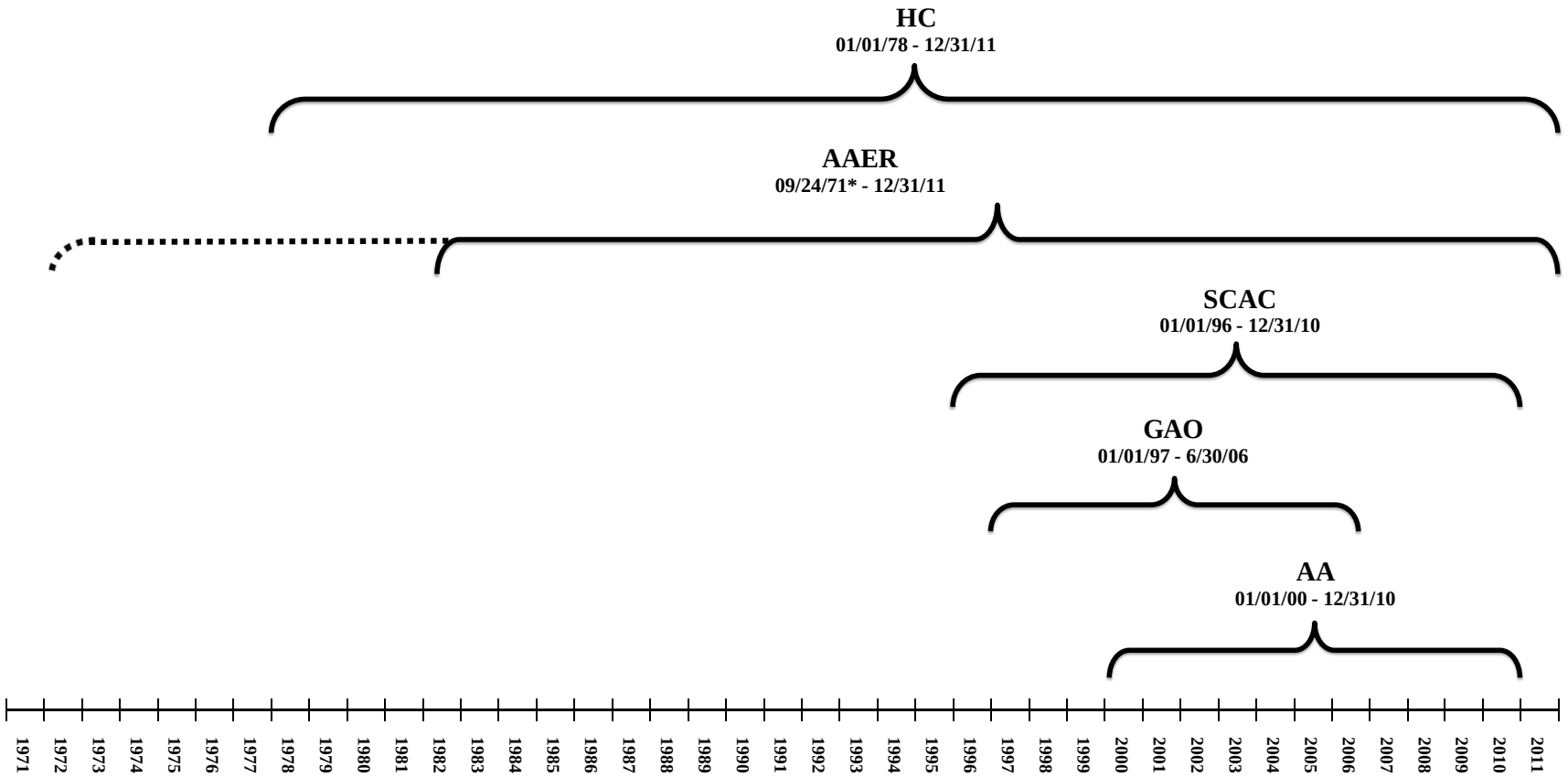


Figure 2. Timeline of information events in the Computer Associates, Inc. case of financial misrepresentation

This figure provides the timeline of events related to Computer Associate's (CA) financial misrepresentation. The events identified by the four commonly used databases (GAO, AA, SCAC, and AAER) are denoted in the bottom half of the timeline. The events identified with the letter 'G' denote restatement announcements included in the GAO database and the events identified with the letters 'AA' denote the restatement announcement included in the AA database. The events identified with the letter 'S' denote securities class action events included in the SCAC database, and the events identified with the letter 'A' denote SEC releases included in the SEC's AAER dataset. The numbers following each letter indicate the sequence of events identified by each database related to the Computer Associates case (i.e., 'G1' is the first event identified by the GAO related to the Computer Associates case of misconduct). The events identified by the hand-collected data are denoted in the top half of the timeline. These events represent the 49 unique days on which new information about Computer Associates' misconduct, or its consequences, were revealed to the public. The events identified with the letter 'R' denote restatement announcements included in the hand-collected data.

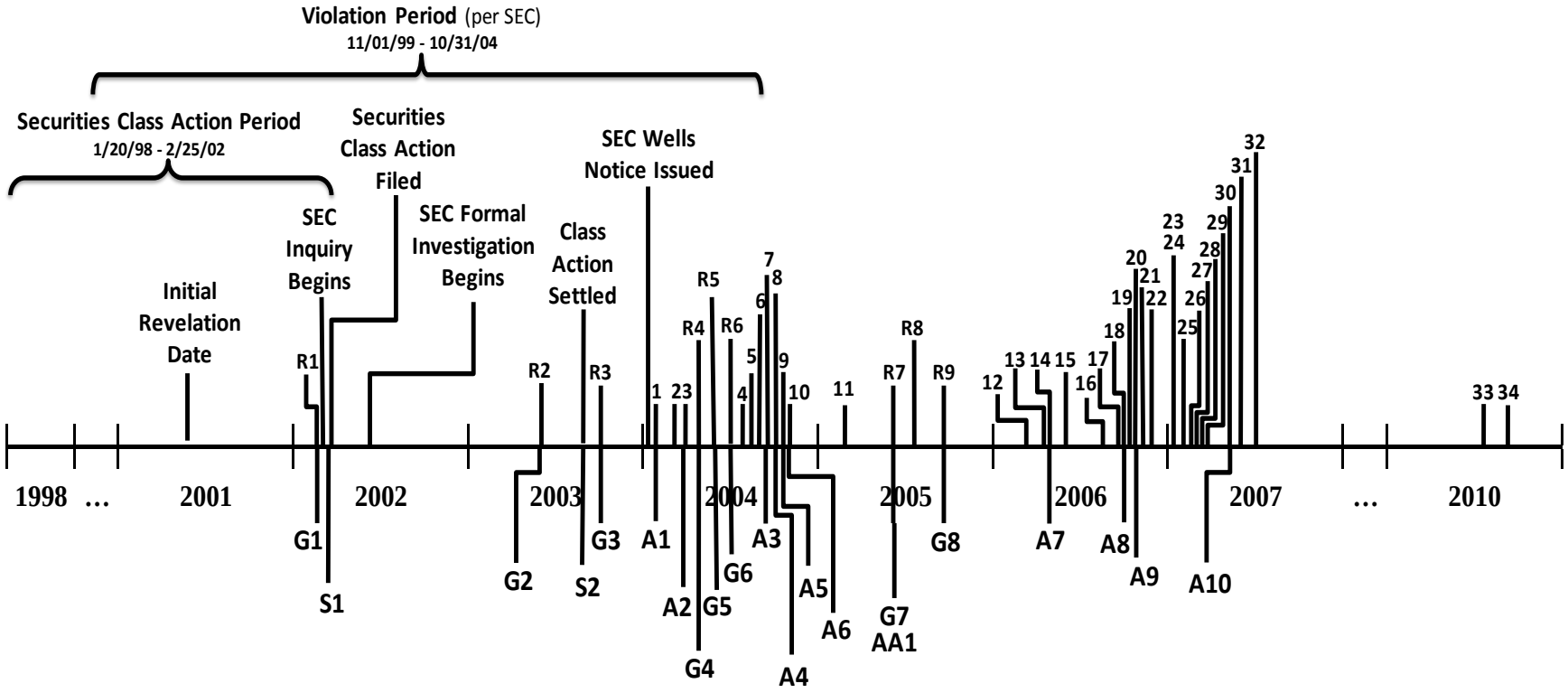
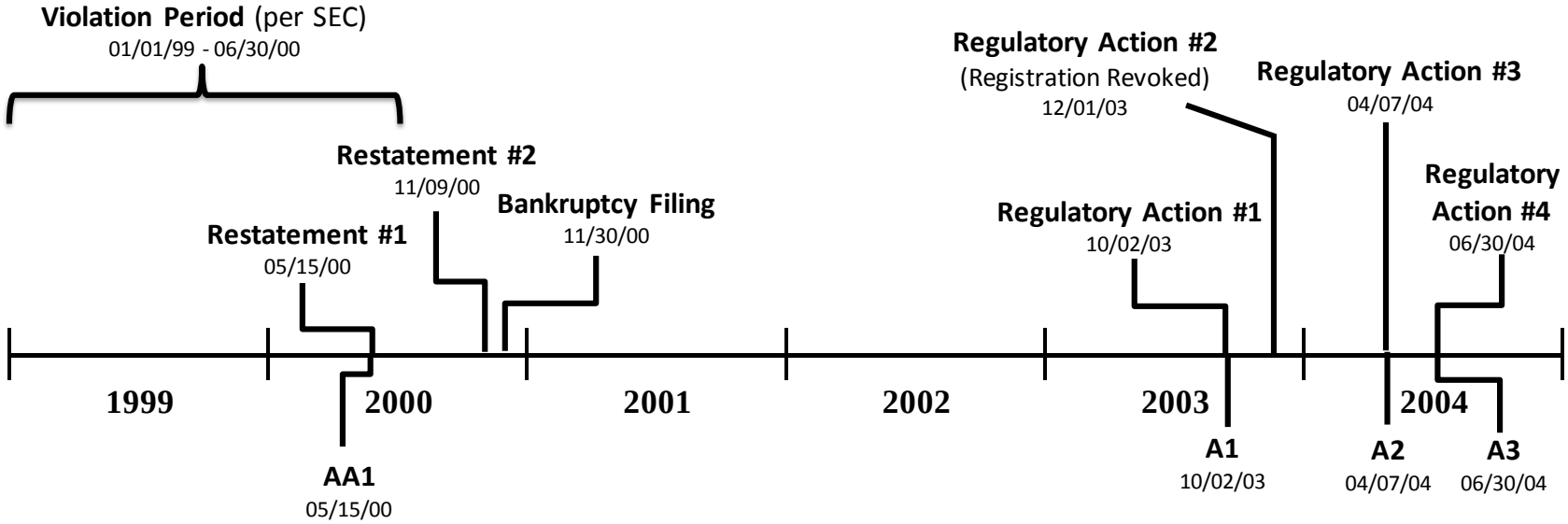


Figure 3. Timeline of information events in the Professional Transportation Group Ltd., Inc. case of financial misrepresentation

This figure provides the timeline of events related to Professional Transportation's financial misrepresentation. The events identified by the four commonly used databases (GAO, AA, SCAC, and AAER) are denoted in the bottom half of the timeline. The events identified with the letters 'AA' denote the restatement announcement included in the AA database and the events identified with the letter 'A' denote SEC releases included in the SEC's AAER dataset. The three AAERs in chronological order are #1887, #1993, and #2048. The events identified by the hand-collected data are denoted in the top half of the timeline. These events represent the seven unique days on which new information about Professional Transportation's misconduct, or its consequences, were revealed to the public



Appendix A: Papers that use data from the four major financial misconduct databases

GAO Database (43)

1. Arthaud-Day, M.L., Certo, S.T., Dalton, C.M., D.R. Dalton, 2006. A changing of the guard: executive and director turnover following corporate financial restatements. *Academy of Management Journal* 49, 1119–1136.
2. Baber, W., Lihong, L, Z. Zinan, 2012. Associations between internal and external corporate governance characteristics: implications for investigating financial accounting restatements. *Accounting Horizons* 26(2): 219-237.
3. Badertscher, B., J. Burks, 2011. Accounting restatements and the timeliness of disclosures. *Accounting Horizons* 25 (4): 609–629. [Also in AA list]
4. Badertscher, B., Hribar, P., N. Jenkins, 2011. Informed trading and the market reaction to accounting restatements. *The Accounting Review* 86(5): 1519-1547. [Also in AA list]
5. Bardos, K., Golec, J., J. Harding, 2011. Do investors see through mistakes in reported earnings? *Journal of Financial and Quantitative Analysis* 46(6): 1917-1946.
6. Burks, J., 2011. Are investors confused by restatements after Sarbanes-Oxley? *The Accounting Review* 86(2): 507-539.
7. Burks, J., 2010. Disciplinary measures in response to restatements after Sarbanes-Oxley. *Journal of Accounting and Public Policy* 29: 195-225.
8. Burns, N., S. Kedia, 2006. The impact of performance-based compensation on misreporting. *Journal of Financial Economics* 79: 35-67.
9. Burns, N., S. Kedia, 2008. Executive option exercises and financial misreporting. *Journal of Banking and Finance* 32: 845-857.
10. Burns, N., Kedia, S., Lipson, M., 2010. Institutional ownership and monitoring: Evidence from financial misreporting. *Journal of Corporate Finance* 16 (4), 443-455
11. Campbell, J., E. Yeung, 2011. Accounting comparability, investor sophistication, and contagion effects. Available at SSRN: http://ssrn.com/abstract=1966715.
12. Carter, M.E., Lynch, L., S. Zechman, 2009. Changes in bonus contracts in the post-Sarbanes-Oxley era. *Review of Accounting Studies* 14: 480-506.
13. Chen, C., B. Goh, Y. Lin, G. Wu. 2012. Spillover effects of restatements on the financial reporting behavior of board-interlocked firms. Available at http://ink.library.smu.edu.sg/soa_research/990/.
14. Chen, X., Q. Cheng, A. Lo. 2013. Accounting restatements and external financing choices. *Contemporary Accounting Research* 30(2): 750-779.
15. Cheng, Q., D. Farber, 2008. Earnings restatements, changes in CEO compensation, and firm performance. *The Accounting Review* 83(5): 1217-1250.
16. Cheng, X., L. Gao, J. Lawrence, D. Smith. 2011. CEO power and SEC prompted restatements. Available at SSRN: http://ssrn.com/abstract=1790128 [Also in AA list]
17. Collins, D., A. Masli, A. L. Reitenga, J. M. Sanchez. 2009. Earnings restatements, the Sarbanes-Oxley Act, and the disciplining of chief financial officers. *Journal of Accounting, Auditing & Finance* 24 (1): 1–34.
18. Desai, H., Hogan, C., M. Wilkins, 2006. The reputational penalty for aggressive accounting: earnings restatements and management turnover. *The Accounting Review* 81 (1): 83–112.
19. Desai, H., Krishnamurthy, S., K. Venkataraman, 2006. Do short sellers target firms with poor earnings quality? Evidence from earnings restatements. *Review of Accounting Studies* 11: 71-90.
20. Doyle, J., Ge, W., S. McVay, 2007. Accruals quality and internal controls over financial reporting. *The Accounting Review* 82(5): 1141-1170.
21. Efendi, J., Srivastava, A., E. Swanson, 2007. Why do corporate managers misstate financial statements? The role of option compensation and other factors. *Journal of Financial Economics* 85: 667–708.
22. Efendi, J., Kinney, M., E. Swanson, 2004. Can short sellers anticipate accounting restatements? Available at SSRN: http://ssrn.com/abstract=591361.
23. Ettredge, M., Huang, Y., W. Zhang, 2012. Earnings restatements and differential timeliness of accounting conservatism. *Journal of Accounting and Economics* 53(3): 489-503. [Also in AA list]
24. Files, R., 2012. SEC enforcement: Does forthright disclosure and cooperation really matter? *Journal of Accounting and Economics* 53(1-2): 353-374. [Also in AA and HC lists]
25. Files, R., Swanson, E., S. Tse, 2009. Stealth disclosure of accounting restatements. *The Accounting Review* 84(5): 1495-1520. [Also in SCAC list]

26. Gleason, C., Jenkins, N., W. Johnson, 2008. Financial statement credibility: the contagion effects of accounting restatements. *The Accounting Review* 83(1): 83-110.
27. Graham, J., Li, S., J. Qiu, 2008. Corporate misreporting and bank loan contracting. *Journal of Financial Economics* 89(1): 44-61.
28. Harris, J., P. Bromiley, 2007. Incentives to cheat: the influence of executive compensation and firm performance on financial misrepresentation. *Organization Science* 18 (3): 350–367.
29. Hazarika, S., Karpoff, J.M., R. Nahata., 2012. Internal corporate governance, CEO turnover, and earnings management. *Journal of Financial Economics* 104 (1), 44–69. [Also in SCAC and HC lists]
30. Hennes, K., Leone, A., B. Miller, 2008. The importance of distinguishing errors from irregularities in restatement research: the case of restatements and CEO/CFO turnover. *The Accounting Review* 83(6): 1487–1519.
31. Hennes, K., Leone, A., B. Miller, 2014. Determinants and market consequences of auditor dismissals after accounting restatements. *The Accounting Review* (forthcoming). [Also in AA list]
32. Hribar, P., N. Jenkins, 2004. The effect of accounting restatements on earnings revisions and the estimated cost of capital. *Review of Accounting Studies* 9: 337-356.
33. Jia, Y., L. van Lent, Y. Zhang. Testosterone and financial misreporting (May 15, 2013). Available at SSRN: http://ssrn.com/abstract=2265510. [Also in AA and AAER list]
34. Kedia, S., T. Phillipon, 2009. The economics of fraudulent accounting. *Review of Financial Studies* 22(6): 2169-2199.
35. Kravet, T., T. Shevlin, 2010. Accounting restatements and information risk. *Review of Accounting Studies* 15: 264-294.
36. Lee, C. J., L. Y. Li, H. Yue, 2006. Performance, growth, and earnings management. *Review of Accounting Studies* 11(2–3): 305–334.
37. Leone, A., M. Liu, 2010. Accounting irregularities and executive turnover in founder-manager firms. *The Accounting Review* 85(1): 287-314.
38. . [Also in AAER and SCAC lists]
39. Peterson, K., 2012. Accounting complexity, misreporting, and the consequences of misreporting. *Review of Accounting Studies* 17: 72-95. [Also in AAER list]
40. Scholz, S., 2008. The changing nature and consequences of public company financial restatements: 1997-2006. The Department of the Treasury. [Also in GAO and AA lists]
41. Srinivasan, S., 2005. Consequences of financial reporting failure for outside directors: evidence from accounting restatements and audit committee members. *Journal of Accounting Research* 43(2): 291-334.
42. Thevenot, M., 2012. The factors affecting illegal insider trading in firms with violations of GAAP. *Journal of Accounting and Economics* 53: 375-390.
43. Wilson, W., 2008. An empirical analysis of the decline in the information content of earnings following restatements. *The Accounting Review* 83(2): 519-548.

AA Restatements Database (**32**)

1. Badertscher, B., J. Burks, 2011. Accounting restatements and the timeliness of disclosures. *Accounting Horizons* 25 (4): 609–629. [Also in GAO list.]
2. Badertscher, B., Hribar, P., N. Jenkins, 2011. Informed trading and the market reaction to accounting restatements. *The Accounting Review* 86(5): 1519-1547. [Also in GAO list]
3. Bens, D., Goodman, T., M. Neamtiu, 2012. Does investment-related pressure lead to misreporting? An analysis of reporting following M&A transactions. *The Accounting Review* 87 (3): 839-865.
4. Bentley, K., Omer, T., N. Sharp. Business strategy, financial reporting Irregularities, and Audit Effort. *Contemporary Accounting Research* 30(2): 780-817. [Also in AAER list]
5. Cao, Y., Myers, L., T. Omer. 2012. Does company reputation matter for financial reporting. *Contemporary Accounting Research*.29(3): 956-990.
6. Cheng, X., L. Gao, J. Lawrence, D. Smith. 2011. CEO power and SEC prompted restatements. Available at SSRN: http://ssrn.com/abstract=1790128 [Also in GAO list]
7. Costello, A., R. Wittenberg-Moerman, 2011. The impact of financial reporting quality on debt contracting: evidence from internal control weakness reports. *Journal of Accounting Research* 49 (1): 97-136.
8. Dey, R. M. and Robin, A. 2012. The Post-SOX Evolution of the Client Portfolio of the Second Tier: A Focus on Restatement and Internal Control Risk. *International Journal of Auditing* 16: 308-334.
9. Dou, Y., Hope, O.K., Thomas, W., Y. Zou. 2013. Blockholder exit threats and financial reporting quality. Available at SSRN: http://ssrn.com/abstract=2374770.

10. Ettredge, M., Huang, Y., W. Zhang, 2012. Earnings restatements and differential timeliness of accounting conservatism. *Journal of Accounting and Economics* 53(3): 489-503. [Also in GAO list]
11. Ettredge, M., Huang, Y., Zhang, W. 2013. Restatement disclosures and management earnings forecasts. *Accounting Horizons* 27(2): 347-369.
12. Files, R., 2012. SEC enforcement: Does forthright disclosure and cooperation really matter? *Journal of Accounting and Economics* 53(1-2): 353-374. [Also in GAO and HC list]
13. Files, R., Sharp, N., A. Thompson. 2014. Empirical evidence on repeat restatement firms. *Accounting Horizons* 28: 93-123.
14. Hayes, L. Identifying unintentional error in restatement disclosures. Working paper, University of Waterloo. [Also in AA list]
15. Hennes, K., Leone, A., B. Miller. 2014. Determinants and market consequences of auditor dismissals after accounting restatements. *The Accounting Review* (forthcoming). [Also in GAO list]
16. Hirschey, M., K. Smith, W. Wilson 2010. Financial reporting credibility after SOX: evidence from earnings restatements. Available at SSRN: http://ssrn.com/abstract=1652982.
17. Hobson, J., Maydew, W., M. Venkatachalam, 2012. Analyzing speech to detect financial misreporting. *Journal of Accounting Research* 50(2): 349-392. [Also in AAER list]
18. Huang, Y., S. Scholz. 2012. Evidence on the association between financial restatements and auditor resignations. *Accounting Horizons* 26(3): 439-464.
19. Irani, A., S. Tate, L. Xu. 2012. Restatements: do they affect auditor reputation for quality? Working paper, Washington and Lee University, University of Massachusetts-Lowell, and University of New Hampshire.
20. Jia, Y., L. van Lent, Y. Zhang. Testosterone and financial misreporting (May 15, 2013). Available at SSRN: http://ssrn.com/abstract=2265510. [Also in GAO and AAER list]
21. Johnstone, K., Li, C., K. Rupley. 2011. Changes in corporate governance associated with the revelation of internal control material weaknesses and their subsequent remediation. *Contemporary Accounting Research* 28(1): 331-383.
22. Kravet, T., Myers, L., J.M. Sanchez, S. Scholz. Do financial statement misstatements facilitate corporate acquisitions? (April 24, 2014). Available at SSRN: http://ssrn.com/abstract=2029953.
23. Lin, S., Pizzini, M., Vargus, M., I. Bardhan. 2011. The role of the internal audit function in the disclosure of material weaknesses. *The Accounting Review* 86(1), 287-323.
24. Lisic, L., Neal, T., Y. Zhang. CEO power, internal control quality, and audit committee effectiveness in substance vs. in form (December 2013). Available at SSRN: http://ssrn.com/abstract=2054966.
25. Lobo, G., Y. Zhao. 2013. Relation between auditor effort and financial reporting misstatements: Evidence from quarterly and annual restatements. *The Accounting Review* 88(4): 1385-1412.
26. McGuire, S., Omer, T., N. Sharp, 2012. The impact of religion on financial reporting irregularities. *The Accounting Review* 87 (2): 645-673.
27. Myers, L., Scholz, S., N. Sharp. Restating under the radar? Determinants of restatement disclosure choices and the related market reactions (April 1, 2013). Available at SSRN: http://ssrn.com/abstract=1309786.
28. Rice, S., D. Weber, 2012. How effective is internal control reporting under SOX 404? Determinants of the (non-)disclosure of existing material weaknesses. *Journal of Accounting Research* 50(3): 811-843.
29. Schmidt, J., 2012. Perceived auditor independence and audit litigation: the role of non-audit services fees. *The Accounting Review* 87 (3): 1033-1065. [Also in SCAC list.]
30. Scholz, S., 2008. The changing nature and consequences of public company financial restatements: 1997-2006. The Department of the Treasury. [Also in GAO and AAER lists.]
31. Srinivasan, S., Wahid, A.S., G. Yu. Admitting mistakes: home country effect on the reliability of restatement reporting (May 23, 2012). Available at SSRN: http://ssrn.com/abstract=2065892.
32. Zhang, J. Securitization and accounting restatements (May 6, 2014). Available at SSRN: http://ssrn.com/abstract=2433758.

SCAC Database (50)

1. Autore, D.M., Hutton, I., Peterson, D., Smith, A. The effect of securities litigation on external financing (May 13, 2014). Available at SSRN: http://ssrn.com/abstract=2290830.
2. Ball, R., L. Shivakumar, 2008. Earnings quality at initial public offerings. *Journal of Accounting and Economics* 45: 324-349.
3. Barabanov S., O. Ozocak, H. Turtle, T. Walker, 2008. Institutional investors and shareholder litigation. *Financial Management* 37(2): 227–250.

4. Bollen, N., V. Pool, 2012. Suspicious patterns in hedge fund returns and the risk of fraud. *Review of Financial Studies* 25(9): 2673-2702.
5. Bowen, R.M., A.C. Call, S. Rajgopal, 2010. Whistle-blowing: Target firm characteristics and economic consequences. *The Accounting Review* 85 (4): 1239-1271.
6. Cao, Z., G. Narayanamoorthy, 2011. The effect of litigation risk on management earnings forecasts. *Contemporary Accounting Research* 28(1): 125-173.
7. Chalmers, K., Naiker, V., F. Navisi. 2012. Earnings quality and Rule 10b- securities class action lawsuits. *Journal of Accounting and Public Policy* 31: 22-43.
8. Chen, S., Matsumoto, D., S. Rajgopal, 2011. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting and Economics* 51: 134-150.
9. Cheng, C.S. A., Huang, H., Li, Y., G. Lobo, 2010. Institutional monitoring through shareholder litigation. *Journal of Financial Economics* 95(3): 356-383.
10. Denis, D. J., P. Hanouna, A. Sarin, 2006. Is there a dark side to incentive compensation? *Journal of Corporate Finance* 12: 467-488.
11. Dyck, A., Morse, A., L. Zingales, 2010. Who blows the whistle on corporate fraud? *Journal of Finance* 65 (6): 2213-2253.
12. Dyck, A., Morse, A., and L. Zingales. 2013. How pervasive is corporate fraud? (February 22, 2013). Available at SSRN: http://ssrn.com/abstract=2222608.
13. Fernandes, N., Lel, U., D. Miller, 2010. Escape from New York: the market impact of loosening disclosure requirements. *Journal of Financial Economics* 95: 129-147.
14. Ferris, S., A. Pritchard, 2001. Stock price reactions to securities fraud class actions under the Private Securities Litigation Reform Act. Michigan Law and Economics Research Paper No. 01-009. Available at SSRN: http://ssrn.com/abstract=288216
15. Fich, M., A. Shivdasani, 2007. Financial fraud, director reputation, and shareholder wealth. *Journal of Financial Economics* 86(2): 306–336.
16. Field, L.; M. Lowry; S. Shu, 2005. Does disclosure deter or trigger litigation*? Journal of Accounting and Economics* 39: 487–507.
17. Files, R., Swanson, E., Tse, S., 2009. Stealth disclosure of accounting restatements. *The Accounting Review* 84(5): 1495-1520. [Also in GAO list]
18. Gande, A., C. Lewis, 2009. Shareholder-initiated class action lawsuits: shareholder wealth effects and industry spillovers. *Journal of Financial and Quantitative Analysis* 44: 823-850.
19. Gong, G., Louis, H., A. Sun, 2008. Earnings management, lawsuits, and stock-for-stock acquirers' market performance. *Journal of Accounting and Economics* 46(1): 62-77.
20. Hab, L.H., M.A. Muller. Capital market consequences of corporate fraud: from infringement to settlement. Working paper, University of Lancaster and WHU.
21. Hanley, K.W., G. Hoberg, 2012. Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics* 103 (2): 235–254.
22. Hao, G., 2011. Securities litigation, withdrawal risk and initial public offerings. *Journal of Corporate Finance* 17(3): 438-456.
23. Hazarika, S., Karpoff, J.M., R. Nahata, 2012. Internal corporate governance, CEO turnover, and earnings management, *Journal of Financial Economics* 104 (1), 44–69. [Also in GAO and HC lists]
24. Hege, S., Malone, C., J. Finnerty. Fraud and firm performance: evidence from fraud on the market and securities class action lawsuits (May 1, 2010). Available at SSRN: http://ssrn.com/abstract=1598783.
25. Helland, E., 2006. Reputational penalties and the merits of class-action securities litigation. *Journal of Law and Economics* 49(2): 365-395.
26. Hochberg, Y. V. Sapienza, P., A. Vissing-Jorgensen, 2009. A lobbying approach to evaluating the Sarbanes-Oxley Act of 2002. *Journal of Accounting Research* 47 (2): 519-583.
27. Houston, J., Lev, B., J. Tucker, 2010. To guide or not to guide? Causes and consequences of stopping quarterly earnings guidance. *Contemporary Accounting Research* 27(1): 143-185.
28. Huddart, S., Ke, B., C. Shi, 2007. Jeopardy, non-public information, and insider trading around SEC 10-K and 10-Q filings. *Journal of Accounting and Economics* 43: 3-36.
29. Johnson, M., Kasznik, R., K. Nelson, 2001. The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research* 39: 297–327.
30. Johnson, M. F., K. K. Nelson, A. C. Pritchard, 2007. Do the merits matter more? The impact of the Private Securities Litigation Reform Act. *Journal of Law, Economics & Organization* 23(3): 627-652.

31. Kim, I., D. Skinner, 2012. Measuring securities litigation risk. *Journal of Accounting and Economics 53:* 290–310.
32. Li, F., 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45: 221-247.
33. Lin, C., Song, F., Z. Sun. The financial implications of corporate fraud. Working paper, Chinese University of Hong Kong and University of Hong Kong.
34. Lowry, M., S. Shu, 2002. Litigation risk and IPO underpricing. *Journal of Financial Economics* 65: 309-335.
35. McNichols, M., S. Stubben, 2008. Does earnings management affect firms' investment decisions? *The Accounting Review* 83(6): 1571-1603. [Also in AAER and GAO lists.]
36. McTier, B.C. and J.K. Wald, 2011. The causes and consequences of securities class action litigation. *Journal of Corporate Finance* 17(3): 649-665.
37. Peng, L., A. Roell, 2008. Executive pay and shareholder litigation. *Review of Finance* 12(1): 141-184.
38. Rogers, J., P. Stocken, 2005. Credibility of management forecasts. *The Accounting Review* 80: 1233–1260.
39. Rogers, J., A. Van Buskirk, 2009. Shareholder litigation and changes in disclosure behavior. *Journal of Accounting and Economics* 47(1-2): 136-156.
40. Schmidt, J., 2012. Perceived auditor independence and audit litigation: the role of non-audit services fees. *The Accounting Review 87* (3): 1033-1065. [Also in AA list.]
41. Schumann, K. Cross-listed firms and shareholder-initiated lawsuits: the market penalties of securities class action lawsuits against foreign firms. Working paper, University of Tennessee.
42. Shivdasani, A., W.L. Song, 2011. Breaking down the barriers: competition, syndicate structure, and underwriting incentives. *Journal of Financial Economics* 99: 581-600.
43. Shu, S., 2000. Auditor resignations: clientele effects and legal liability. *Journal of Accounting and Economics* 29: 173-205.
44. Tian, X., Udell, G., X. Yu. Disciplining delegated monitors: the consequences of failing to prevent fraud (September 26, 2012). Available at SSRN: http://ssrn.com/abstract=1782238. [Also in AAER list.]
45. Wang, T., A. Winton. Competition and corporate fraud waves (April 2012)..Available at SSRN: http://ssrn.com/abstract=1783752. [Also in AAER list.]
46. Wang, T. Y., A. Winton, X. Yu, 2010. Corporate fraud and business conditions: Evidence from IPOs. *Journal of Finance* 65(6): 2255-2292. [Also in AAER list.]
47. Xu, W., 2010. Do management earnings forecasts incorporate information in accruals? *Journal of Accounting and Economics* 49: 227-246.
48. Yu, F., 2008. Analyst coverage and earnings management. *Journal of Financial Economics* 88: 245-271.
49. Yu, F., X. Yu, 2011. Corporate lobbying and fraud detection. *Journal of Financial and Quantitative Analysis* 46(6): 1865-1891.
50. Zhu, Y., 2009. The relation between IPO underpricing and litigation risk revisited: changes between 1990 and 2002. *Financial Management* 38(2): 323-355.

AAER Database (56)

1. Armstrong, C., Jagolinzer, A., D. Larcker, 2010. Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research* 48, 225–271.
2. Beasley, M., 1996. An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review* 71(4), 443-465.
3. Beatty, A., Liao, S., Yu, J. 2013. The spillover effect of fraudulent financial reporting on peer firms' investments. *Journal of Accounting and Economics* 55: 183-205.
4. Beneish, M., 1997. Detecting GAAP violations: implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16(3), 271-309.
5. Beneish, M., 1999. Incentives and penalties related to earnings overstatements that violate GAAP. *The Accounting Review* 74, 425-457.
6. Bentley, K., Omer, T., N. Sharp. Business strategy, financial reporting irregularities, and audit effort. *Contemporary Accounting Research* 30(2): 780-817. [Also in AA list]
7. Bonner, S., Palmrose, Z., S. Young, 1998. Fraud type and auditor litigation: an analysis of SEC accounting and auditing enforcement releases. *The Accounting Review* 73(4): 503-532.
8. Brazel, J., Jones, K., M. Zimbelman, 2009. Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research* 47(5): 1135-1166.

9. Cao, Z., Leng., F., Feroz, E.H., and S. Davalos. 2013. Corporate governance and default risk of firms cited in the SEC's Accounting and Auditing Enforcement Releases. *Review of Quantitative Finance and Accounting* (forthcoming).
10. Carcello, J., Neal., T., Palmrose, Z., S. Scholz, 2011. CEO involvement in selecting board members, audit committee effectiveness, and restatements. *Contemporary Accounting Research* 28(2): 396-430.
11. Caskey, J., M. Hanlon. 2013. Dividend policy at firms accused of accounting fraud. *Contemporary Accounting Research* 30(2): 818-850.
12. Cecchini, M., Aytug, H., Koehler, G., P. Pathak, 2010. Detecting management fraud in public companies. *Management Science* 56(7): 1146–1160.
13. Chen, K., Y. Zhao, 2008. Staggered boards and earnings management. *The Accounting Review* 83(5): 1347-1381.
14. Davidson, R., 2013. Accounting fraud: booms, busts, and incentives to perform. Working paper, Georgetown University.
15. Davidson, R., Dey, A., and S. Smith. Executives' "off-the-job" behavior, corporate culture, and financial reporting risk. *Journal of Financial Economics* (forthcoming).
16. Dechow, P., Hutton, A., Kim, J.H., R. Sloan, 2012. Detecting earnings management: a new approach. *Journal of Accounting Research* 50(2): 275-334.
17. Dechow, P., Sloan, R., A. Sweeney, 1995. Detecting earnings management. *The Accounting Review* 70(2): 193-225.
18. Dechow, P., Sloan, R., A. Sweeney, 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13: 1-36.
19. Dechow, P., Ge, W., Larson, C., R. Sloan, 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28(1): 17-82.
20. DeFond, M., D. Smith, 1991. Discussion of the financial and market effects of the SEC Accounting and Auditing Enforcement Releases. *Journal of Accounting Research* 29: 143-148.
21. Donelson, D., J. McInnis, R. Mergenthaler. 2012. The effect of corporate governance reform on financial reporting fraud (August 29, 2012). Available at SSRN: http://ssrn.com/abstract=2138348.
22. Dunn, P., 2004. The impact of insider power on fraudulent financial reporting. *Journal of Management* 30(3): 397-412.
23. Erickson, M., Hanlon, M., E. Maydew, 2004. How much will firms pay for earnings that do not exist? Evidence of taxes paid on allegedly fraudulent earnings. *The Accounting Review* 79(2): 387-408.
24. Erickson, M., Hanlon, M., E. Maydew, 2006. Is there a link between executive equity incentives and accounting fraud? *Journal of Accounting Research* 44: 113-143.
25. Ettredge, M., Sun, L., Lee, P., A. Anandarajan., 2008. Is earnings fraud associated with high-deferred tax and/or book minus tax levels? *Auditing: A Journal of Practice & Theory* (May): 1-34.
26. Ettredge, M., Scholz, S., Smith, K., L. Sun., 2010. How do restatements begin? Evidence of earnings management preceding restated financial reports. *Journal of Business Finance & Accounting* 37(3-4): 332–355.
27. Farber, D., 2005. Restoring trust after fraud: Does corporate governance matter? *The Accounting Review* 80: 539-561.
28. Feng, M., Ge, W., Luo, S., T. Shevlin, 2011. Why do CFOs become involved in material accounting manipulations? *Journal of Accounting and Economics* 51: 21-36.
29. Feng, M. Ge, W., Li, C., Nagarajan, N. Forecast Strategies of Earnings Manipulators and Associated Consequences (April 2014). Available at SSRN: http://ssrn.com/abstract=1942076.
30. Feroz, E., Park, K., V. Pastena, 1991. The financial and market effects of the SEC's accounting and auditing enforcement releases. *Journal of Accounting Research* 29: 107-142.
31. Fulmer, S., A. Knill. Political contributions and the severity of SEC enforcement (March 1, 2012). Available at SSRN: http://ssrn.com/abstract=2024069.
32. Ge, W., Matsumoto, D., J. Zhang, 2011. Do CFOs have style? An empirical investigation of the effect of individual CFOs on accounting practices. *Contemporary Accounting Research* 28(4): 1141-1179.
33. Geiger, M., Lennox, C., D. North, 2008. The hiring of accounting and finance officers from audit firms: How did the market react? *Review of Accounting Studies* 13: 55-86.
34. Gerakos, J., A. Kovrijinykh. 2013. Performance shocks and misreporting. *Journal of Accounting and Economics* 56: 57-72.
35. Guan, L., Kaminski, K., T. Wetzel, 2007. Can investors detect fraud using financial statements: an exploratory study. *Advances in Public Interest Accounting* 13: 17-34.

36. Hayes, L. Identifying unintentional error in restatement disclosures. Working paper, University of Waterloo. [Also in AA list]
37. Hobson, J., Mayew, W., M. Venkatachalam, 2012. Analyzing speech to detect financial misreporting. *Journal of Accounting Research* 50(2): 349-392. [Also in AA list]
38. Jia, Y., L. van Lent, Y. Zhang. Testosterone and financial misreporting (May 15, 2013). Available at SSRN: http://ssrn.com/abstract=2265510. [Also in GAO and AA list]
39. Johnson, S., Ryan, H., Y. Tian, 2009. Managerial incentives and corporate fraud: The sources of incentives matter. *Review of Finance* 13: 115-145.
40. Larcker, D., A. Zakolyukina, 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50(2): 495-540.
41. Lennox, X., J. Pittman, 2010. Big Five audits and accounting fraud. *Contemporary Accounting Research* 27: 208-247.
42. Markelevich, A., and R. Rosner. 2013. Auditor fees and fraud firms. *Contemporary Accounting Research* 30(4): 1590-1625.
43. McNichols, M., S. Stubben, 2008. Does earnings management affect firms' investment decisions? *The Accounting Review* 83(6): 1571-1603. [Also in GAO and SCAC lists.]
44. Miller, G.S., 2006. The press as a watchdog for accounting fraud. *Journal of Accounting Research* 44: 1-33.
45. Ozbas, O., 2008. Corporate fraud and real investment (February 2008). Available at SSRN: http://ssrn.com/abstract=891450.
46. Palmrose, Z.V. S. Scholz, 2004. The circumstances and legal consequences of non-GAAP reporting: evidence from restatements. *Contemporary Accounting Research* 21(1): 139-80.
47. Palmrose, Z.V., Richardson, V., S. Scholz, 2004. Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics* 37: 59–89.
48. Peterson, K., 2012. Accounting complexity, misreporting, and the consequences of misreporting. *Review of Accounting Studies* 17: 72-95. [Also in GAO list.]
49. Pincus, K., Holder, W.H., T.J. Mock, 1988. Reducing the incidence of fraudulent financial reporting: the role of the Securities and Exchange Commission. Los Angeles, CA: SEC Financial Reporting Institute of the University of California.
50. Scholz, S., 2008. The changing nature and consequences of public company financial restatements: 1997-2006. The Department of the Treasury. [Also in GAO and AA lists.]
51. Schrand, C., S. Zechman, 2012. Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics* 53(1-2): 311-329.
52. Skousen, C., B. Twedt, 2009. Fraud in emerging markets: a cross-country analysis. *Cross Cultural Management: An International Journal* (16): 301-316.
53. Stubben, S. 2010. Discretionary revenues as a measure of earnings management. *The Accounting Review* 85(2): 695-717.
54. Tian, X., Udell, G., X. Yu. Disciplining delegated monitors: the consequences of failing to prevent fraud. (September 26, 2012). Available at SSRN: http://ssrn.com/abstract=1782238. [Also in SCAC list.]
55. Wang, T., A. Winton. Competition and corporate fraud waves (April 2012). Available at SSRN: http://ssrn.com/abstract=1783752. [Also in SCAC list.]
56. Wang, T. Y., A. Winton, X. Yu, 2010. Corporate fraud and business conditions: Evidence from IPOs. *Journal of Finance* 65(6): 2255-2292. [Also in SCAC list]

Appendix B: Database descriptions and coverage periods

This appendix describes four databases commonly used in financial misconduct research (GAO, AA, SCAC, and AAER) and the hand-collected data used in this paper. Table 1 briefly summarizes this information, but this appendix provides a more elaborate delineation of each database's period of coverage, and how the GAO, AA, SCAC, and AAER databases relate to the hand-collected data.

Government Accountability Office (GAO) financial statement restatements database

Between 2002 and 2006, the U.S. Government Accountability Office (GAO) issued three reports that identified a large sample of financial restatements that academic researchers have used extensively. The reports were conducted at the request of Senator Paul Sarbanes, and strive to: "(1) determine the number of, reasons for, and other trends in financial statement restatements since 1997; (2) analyze the impact of restatement announcements on the restating companies' stock market capitalization; (3) research available data to determine the impact of financial statement restatements on investors' confidence in the existing U.S. system of financial reporting and capital markets; (4) analyze SEC enforcement actions involving accounting and auditing irregularities; and (5) describe the major limitations of the existing oversight structure and steps that have been and are being taken to ensure the integrity of corporate financial disclosures and ongoing challenges" (GAO 2002, p. 1-2).

The first report identified 919 restatements by 845 firms (689 publicly traded) that "involved accounting irregularities resulting in material misstatements of financial results" (GAO 2002, p. 2). The restatements were identified by Lexis-Nexis keyword searches for variants of "restate," then screened with the intent of removing restatements unrelated to correcting accounting standards application mistakes. Additional information about the 919 restatement announcements (including the date of the announcement) was issued in GAO Report 03-395R.

GAO Reports 06-678 and 06-1053R extended the original report through June 30, 2006. The restatement announcements added through these two reports were identified by Lexis-Nexis "US Newspapers and Wires" database keyword searches for variants of "restate," "adjust," "amend," and "revise" within 50 words of "financial statement" or "earning" (GAO 2006a, p.52). In aggregate, the

three reports include a total of 2,705 restatement announcements spanning 9½ years (January 1, 1997 through June 30, 2006).

Hennes, Leone, and Miller (2008) cull the 2,705 GAO restatements classifying 715 as "irregularities" based on their search for (1) the presence of the word "fraud" or "irregularity" in the restatement announcement; (2) whether the SEC or DOJ conducted an investigation related to the restatement; or (3) whether other investigations into the accounting matter (i.e., the hiring of a forensic accounting firm) were conducted (p. 1489). They recognize the need to adjust for multiple restatements per firm (per enforcement action) and pointedly avoid referring to these 715 announcements as "fraud" events, using the more inclusive term "irregularities." They note that the legal distinction between the two terms is blurred in practice and the literature by auditing guidelines (e.g., SAS No. 82, AICPA 1997) (that) use the term "fraud" to refer to all intentional misstatements (footnote 1).

Audit Analytics (AA) financial statement restatements database

As of August 29, 2011, the restatement database compiled and maintained by Audit Analytics (the research arm of the Ives Group, Inc.) includes 11,001 financial restatements and/or non-reliance filings made by any SEC registrant since January 1, 2000.[22] Audit Analytics (AA) defines a restatement as "an adjustment to previously issued financial statements as a result of an error, fraud, or GAAP misapplication [and] does not include restatements caused by adoption of new accounting principles or revisions for comparative purposes as a result of mergers and acquisitions."[23]

AA extracts its data principally from SEC Form 8-K or required amended periodic reports (Forms 10-K/A, 10-Q/A, 10KSB/A, 20-F/A, and 40-F/A). AA claims to analyzed all 8-K and 8-K/A filings that contain "Item 4.02 - Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review" (an item required by the SEC since August 2004).[24] In addition, all

[22] Beginning August 15, 2004, the SEC requires firms to file Form 8-K Item 4.02 whenever previously issued annual and interim financial statements should no longer be relied upon.
[23] This definition was provided in an email from an Audit Analytics analyst on November 15, 2011.
[24] http://auditanalytics.com/doc/dd-restatements.pdf

amended Forms 10-K/A, 10-Q/A, 10KSB/A, 20-F/A, and 40-F/A are reviewed to determine if the amendment is due to a restatement, and all audit opinions are searched for derivatives of the word "restate" with the intent of detecting the so-called "stealth" restatements contained in periodic reports rather than event filings.

Audit Analytics describes its data collection methodology as follows:

> "After beginning a record that identifies a restatement cause or issue, we subsequently attach filings that address or add information to that original record, in essence creating a history for it. Generally, we consider such a history of filings to be one restatement. In certain circumstances, however, a company clearly identifies a completely new issue in a subsequent filing, and therefore this new issue is treated as a new restatement. For example, if a company files an 8K indicating a revenue recognition problem, but then files an ensuing 10K/A that discloses not only a revenue recognition issue, but also a Cash Flow Statement (FAS 95) issue, then a separate and second record is created to track that newly disclosed restatement issue as a distinct restatement. We do not, however, identify the revenue recognition issue in the second restatement so as not to double count the restatement issues in this process. Generally, the intent is to err on the side of combining new disclosures (such as a change in period or amounts) in restatements unless it is clear that the issues are different."[25]

The AA database identifies the firm and the date on which Form 8-K, Item 4.02 (Non-Reliance on Previously Issued Financial Statements), or a Related Audit Report, or a Completed Interim Review was filed with the SEC. Also reported are the beginning and ending dates of the period to be restated and the AA analyst's opinion of whether the restatement will have a positive or negative impact on the firm's financial statements and whether the restatement disclosure identified: accounting rule application failures, financial fraud / irregularities / misrepresentations, accounting and clerical application errors, or other issues as the reason for the restatement.

Securities Class Action Clearinghouse (SCAC) database

The SCAC database provides public information about federal class action securities fraud litigation. Coverage begins in 1996, following the passage of the Private Securities Litigation Reform Act (PSLRA) of 1995. Each SCAC record:

> "…[I]dentifies a defendant or defendants that are being sued in Federal Court by shareholders of the same issuer, or issuers, for the first time in violation of Federal Securities Laws. In other

[25] http://www.alacra.com/acm/2033_sample.pdf, page 3.

> words, if two or more securities class action complaints against the same defendant, or defendants, have the same underlying allegations, there is only one record in the hand collected database. Accordingly, when a case arises out of the same subject matter as one of the actions first filed and is brought by shareholders of the same issuer, or issuers, it is hereafter part of that record ... as a general rule, we select the first complaint we have identified. If multiple complaints are filed at one time, we choose the complaint that appeared to contain the most detailed allegations...*we [do not] track SEC enforcement proceedings when there is no parallel federal civil class action*" (emphasis added).[26]

The last phrase is emphasized because it explains the relatively small overlap between the SCAC database and the hand-collected data (i.e., many SEC enforcement proceedings prompt no federal civil class actions. As of December 31, 2010, the SCAC database contained information relating to 3,227 issuers named in federal class action securities fraud lawsuits.

Accounting and Auditing Enforcement Releases (AAER) database

In 1982, the SEC began assigning a new secondary designation to some of its enforcement releases if the proceeding involved accountants. The first AAER (AAER-1) states that:

> "Future Commission releases announcing enforcement actions involving accountants will be issued as Accounting and Auditing Enforcement Releases (AAER). Henceforth, interested persons will be able to easily distinguish enforcement releases involving accountants from releases in which the Commission announces the adoption or revision of rules related to financial reporting or discusses its interpretive views on financial reporting matters."[27]

Labeling an SEC enforcement action as an AAER is a discretionary SEC staff decision; no standard protocol exists for this determination. As of December 31, 2010, the SEC had issued 3,610 administrative proceedings and litigation releases with a secondary AAER designation. AAER-3222 is the 3,610th AAER because releases sometimes involve two or more actions distinguished by letter suffixes (i.e. AAER-1A, AAER-1B, etc.) and several AAER numbers were assigned erroneously to separate enforcement actions.[28]

[26] Securities Class Action Clearinghouse Database: Definition of the SCAC Sample Data (September 27, 2004) http://securities.stanford.edu/info.html
[27] Accounting and Auditing Enforcement Release No. AAER-1, 1982 SEC LEXIS 2565, May 17, page 1.
[28] See *http://www.sec.gov/divisions/enforce/friactions.shtml* for additional information.