Did Accrual Earnings Management Decline and Real Earnings Management Increase Post-SOX? A Re-examination and Replication

By: Morton Pincus, Shijia Wu and Jasper Hwang

ABSTRACT: A widely cited paper, Cohen, Dey, and Lys (2008, hereinafter CDL), examines accrual (AEM) and real earnings management (REM) pre- and post-Sarbanes-Oxley. It has been a dozen years since CDL's publication and almost 20 years since SOX became law. Our re-examination analyses investigate whether CDL's findings are robust to a battery of robustness checks regarding sample construction, model specification, and variable definitions, and whether CDL's conclusions hold after 2005, the end of their sample period. We find support for many of CDL's conclusions, but also evidence suggesting the need to adjust our understanding of the use of AEM and REM post-SOX. CDL's time trends evidence of decreasing AEM and increasing REM is sensitive to research design choices; and the substitution between AEM and REM that CDL find is attenuated post-SOX, thus cautioning against automatically assuming AEM and REM are substitutes. We also replicate CDL with mixed success.

Key words: accrual-based and real earnings management, SOX, re-examinations, replications

I. INTRODUCTION

In July 2002, the Sarbanes-Oxley Act became law with the aim to restore investor confidence in the integrity of financial reporting and capital markets by enhancing corporate governance. The act requires public companies to strengthen internal controls, makes senior managers personally liable for misreporting, and establishes stricter criminal penalties for corporate accounting fraud such as at Enron and WorldCom. In a highly cited paper, Cohen, Dey, and Lys (2008, hereafter CDL) argue that, “after the passage of SOX, accrual manipulations were more likely to draw auditors’ or regulators’ scrutiny than real earnings management.”¹ Thus, ostensibly, SOX increased the costs of accrual-based earnings management (AEM) relative to real earnings management (REM).² Consistent with this prediction, CDL find accrual-based earnings management (AEM) decreases and real earnings management (REM) increases post-SOX (2002-2005). They also find AEM and REM behave as substitutes in their multivariate analyses. Researchers have widely cited CDL and examined AEM and/or REM in other settings, although most studies have focused on relatively short time-periods post-SOX.

It has been almost two decades since the passage of SOX, and many changes in regulatory scrutiny of corporate disclosures have taken place. For example, the 2012 Jumpstart Our Business Startups (i.e., JOBS) Act limits the regulatory requirements for SOX compliance to help reduce the burden that smaller companies face in raising funds. Additionally, Cunningham, Johnson, Johnson, and Lisic (2020) find the

¹ Graham et al. (2005, 36) “acknowledge that the aftermath of accounting scandals at Enron and WorldCom and the certification requirements imposed by the Sarbanes-Oxley Act may have changed managers’ preferences for the mix between taking accounting versus real actions to manage earnings. Alternatively, it could simply be that managers are more willing to admit to taking real decisions than to accounting decisions.”

² The legal literature when SOX became law reflected disagreement as to whether SOX represented new and substantive reform. Brickey (2003) argued SOX expanded prohibitions against fraud and obstruction of justice, increased criminal penalties, and strengthened sentencing guidelines. However, other legal writers (Cunningham 2003; Perino 2002; Ribstein 2002) argued existing statutes already effectively embodied SOX’s criminal provisions and that SOX was a political response to the high-profile cases of fraudulent financial reporting and would not significantly affect firms in general. Empirical results from a SOX event study focusing on earnings management (Li, Pincus, and Rego 2008) suggest stock market participants anticipated SOX would constrain AEM more the more firms had managed earnings in the past, and result in higher earnings quality and/or mitigate agency problems between shareholders and managers. Also, see Coates and Srinivasan (2014).

number of SEC comment letters decreased by 51% from 2011 to 2016, suggesting less scrutiny over financial reporting.³ In this paper, we are interested in whether managers began using AEM again as scrutiny declined and whether it changes the cited substitution between AEM and REM. In addition, we seek to investigate whether the conclusions in CDL are robust to empirical design changes, including some methodological developments appearing in the literature subsequent to CDL.

We begin with a re-examination of CDL using the same sample period as CDL used in their multivariate tests (1992-2005). Our re-examination analyses include several robustness checks regarding sample construction, model specification, and variable definitions (see Exhibit 1). This procedure allows us to assess whether the findings in CDL are sensitive to research design choices. We find that, (1) consistent with CDL, AEM decreases during 1992-2005, and the results are robust to various empirical design changes except for one case of a different sample construction; (2) REM either increases or remains unchanged during 1992-2005, suggesting sensitivity to the research design; and (3) AEM and REM behave as substitutes over the pre-SOX sample period (1992-2001). However, the substitute relation attenuated following SOX (2002-2005), and this finding is generally robust to different research designs. Overall, our first set of re-examination analyses shed new light into both the research design choice and the interpretation regarding the examination of AEM and REM substitution.

Next, we further re-examine CDL by extending the sample period to span 1992-2017. As we allude to above, it is unclear whether the conclusions from CDL can describe managers' accrual and real activities management decisions nearly 20 years after the passage of SOX. To understand better how the relation between AEM and REM may have evolved over time, we investigate both the entire extended

³ Katz and McIntosh (2019) also find the number of comment letters issued by the SEC has dramatically decreased from January 2010 to June 2018, but the percentage of letters addressing non-GAAP measures grew. They argue there was an increase in SEC scrutiny and enforcement of non-GAAP reporting. Moreover, Black, Christensen, and Schmardebeck (2017) document substitution between non-GAAP reporting and both REM and AEM. Viewed collectively, the increased scrutiny on non-GAAP reporting can be seen as inducing more AEM as well as REM, confounding the substitution relation between AEM and REM.

post-SOX period (2002-2017) and, within it, four post-SOX sub-periods. We find that, (1) AEM decreases during 2002-2005 and 2010-2017 but remains unchanged 2006-2009; (2) REM increases during 2002-2005 and 2008-2017 but significantly decreases during 2006-2007; and (3) the attenuated AEM and REM substitution relation continues during 2006-2017. Viewed collectively, the second set of our re-examination results suggest the decreasing AEM trend, increasing REM trend, and the AEM and REM substitution relation do not consistently hold across different time-periods post-SOX. We believe these results should be of interest to future research in studying the relation between AEM and REM. We also call for caution in automatically assuming an AEM and REM substitution relation in future studies.

In the remainder of the paper: Section II discusses the importance of re-examinations and replications, and why we choose to re-examine and replicate CDL. Section III reports the sample selection and results of our re-examination, which is our main focus, and considers the sensitivity that empirical design changes and an extended sample have on the robustness and generalizability of CDL's results. Section IV reports on our replication of CDL, and section V concludes with a summary and discussion.

II. Background and Summary of Results

Why are Re-examination and Replication Important?

Researchers in several disciplines have raised concerns about the ability to reproduce the results of empirical research. A startling example is from psychology. The Reproducibility Project in Psychology could only replicate the examined parts of 39 out of 100 experimental and correlational studies (Open Science Collaboration, 2015).⁴ Goodman et al. (2016, 1) state that “‘reproducibility’ refers to the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator.” Goodman et al. (2016, 4) also discuss ‘robustness’ and ‘generalizability,’ terms that have

⁴ Also see, for example, Camerer et al. (2016); Goodman, Fanelli, and Ioannidis (2016); and Banks, Boyle Jr., Pollack, White, Batchelor, Whelpley, Abston, Bennett, and Adkins (2016).

sometimes been used in lieu of reproducibility. They define robustness as “the stability of experimental conclusions to variations in either baseline assumptions or experimental procedures. [And robustness] is related to the concept of ‘generalizability’.” Goodman et al. (p. 4) add, “Whether a study design is similar enough to the original to be considered a replication, a ‘robustness test,’ or some of many variations of pure replication that have been identified, particularly in the social sciences...is an unsettled question.”

In accounting, Ball and Brown (2019, 429) replicate Ball and Brown (1968) and say, “...replicability possibly is the most important criterion for scientific work.” Recently, Hail, Lang, and Leuz (2020) published survey results on reproducibility of accounting research based on responses by attendees at the 2019 *Journal of Accounting Research (JAR)* Conference, thereby adding to the discussion of research reproducibility across disciplines. Hail et al. (2020, 521) state “a study is replicated when it is repeated exactly as published, using the same underlying data sets and methods.” They refer to this as replication in a “narrow sense” (p. 523). They define reproducibility as “the ability of a researcher to confirm the findings of a peer-reviewed and published study in a similar setting that may include slight but reasonable variations of method, sample, and/or time period,” and view this as replication in the “wide sense” (p. 523). Hail et al. (p. 526) expect fewer replications than reproducibility studies since replicability problems can result “from coding errors or difficulties in simply following what the authors have done in their study.” Their survey results indicate respondents believe the “inability to reproduce results significantly detracts from the usefulness of research findings, but [respondents] seldom attempt to publish irreproducible results when they detect them” (p. 521), and “reproducibility is perceived as not receiving sufficient attention” in accounting research (p. 523).

The *Journal of Financial Reporting’s (JFR)* welcomes replications and its editorial policy states “the term ‘Replication’ [is used] to describe an archival empirical analysis that *primarily* performs the same analysis as an existing study but adds, for example, another control variable or additional sensitivity

analysis, or uses a slightly different sample” (emphasis in the original). *JFR* also calls for re-examinations, three elements of which are reproductions (applying the same analyses to a different data set), re-analyses (applying new analyses to the same or similar data set), and extensions (applying the same or new analyses to data drawn from a different target population or setting). Our re-examination reflects these elements.⁵

Why CDL?

SOX, the Sarbanes-Oxley Act of 2002, is arguably the most important legislation related to accounting and auditing enacted in the United States since the securities acts of the 1930s and, among other things, made external auditing of public companies a regulated industry. Given the importance of both SOX and earnings management in the financial accounting literature, it is not surprising CDL is widely cited. As of mid-October 2019, there were 668 cites of CDL on the *Web of Science* and 2,458 cites on *Google Scholar*. As a benchmark, in the same issue of *The Accounting Review* as CDL there are eight other main articles. CDL has more citations than the *total* citations of the other eight articles combined: 668 versus 590 on *Web of Science* and 2,458 versus 2,119 on *Google Scholar*. We thus believe it is fair to claim CDL is an influential paper that made important contributions to the literature and that re-examining and replicating it are useful and important exercises.

CDL (p. 758) state they “investigate the prevalence of both accrual-based and real earnings management activities in the period leading to the passage of SOX and in the period following the passage of SOX. [Their] primary motivation for conducting this analysis [was] to investigate whether the period leading to the passage of SOX was characterized by a widespread increase in earnings management rather than by a few highly publicized events, and whether the passage of SOX resulted in a reduction in earnings management,” by which they meant AEM.

⁵ Our study was underway before *JFR*’s call for re-examination proposals, but is consistent with *JFR*’s notion of re-examinations. See *JFR*’s re-examination document: <https://aaahq.org/Portals/0/documents/calls/2019/PRex%20Call%20for%20Proposals%202019-01-05.pdf>.

Our primary focus is a re-examination of CDL’s analyses, and we seek to investigate whether the results CDL report continue to hold a decade after its publication. We do this by considering the robustness and generalizability of their earnings management results in light of a number of empirical design changes, including some methodological developments appearing in the literature subsequent to CDL. We consider different sample constructions, model specifications, and definitions of earnings management variables; in addition, we extend CDL’s sample period. Distinct from the re-examination, we also seek to replicate CDL by closely duplicating their study.

To facilitate the reading and understanding of this paper, we provide Exhibit 2, which displays a table of contents, as well as Exhibit 1 that lists the re-examination robustness and generalization checks (hereinafter robustness checks) we consider, and their rationales.

Show Exhibit 1 About Here

Show Exhibit 2 About Here

Summary of Results

CDL report time trends analyses for accrual-based (AEM) and real earnings management (REM), and provide evidence of AEM and REM increases in the accounting scandal period (SCA, 2000-01), followed by AEM decreases and REM increases post-SOX (2002-05). They also find AEM and REM behave as substitutes in their sample period (1987-2005).⁶

Summary of re-examination results

Show Table 1 About Here

Our re-examination results are as follows, and we briefly summarize the key results in Table 1 by comparing the similarities and differences between CDL’s and our main findings. First, using CDL’s empirical approach, we can confirm their time trends results based on our ExecuComp sample. We also

⁶ We use of the phrase “post-SOX” to refer to the period following the passage of SOX. In CDL, that period begins in 2002 and goes through 2005. Later we extend the post-SOX period through 2017.

adapt CDL's time trends model not only to examine the sign of AEM during the scandal and post-SOX periods, but importantly to also assess any change (i.e., trend) in AEM and REM by including interaction terms that reflect changes in the accounting scandal and post-SOX periods. For the scandal period (2000-01), the time trends evidence across our robustness checks generally suggest higher levels of AEM during scandal period, but the results also suggest AEM peaks in 2000 and then exhibits a decline in 2001 (i.e., before the passage of SOX), which is not what CDL seem to expect. Second, for the post-SOX period (2002-05), the evidence suggests a declining AEM trend; this is consistent with CDL's results. Also, there is evidence indicating REM increases, but only when we drop foreign firms or use industry average R&D to replace missing R&D observations in computing discretionary expenses. However, when we compute the performance-matched version of our aggregate REM proxy, the results for the robustness checks indicate REM increases post-SOX. Hence, CDL's results for REM increases post-SOX are somewhat dependent on how we compute the aggregate REM proxy.

Third, there is evidence of a negative relation between AEM and REM, suggestive of substitution over the entire sample period (1992-2005), which is consistent with CDL. However, to test more directly for a relation between AEM and REM post-SOX (2002-2005), we augment CDL's multivariate models with interaction terms between AEM and the post-SOX period indicator (and between REM and the post-SOX indicator). Inconsistent with CDL, we find there is a decline in substitution between AEM and REM post-SOX, and even some suggestion of a complementary relation.

Fourth, our SUSPECT firms' analyses on the change from the pre- to post-SOX periods in meeting/just beating three earnings targets yield results generally inconsistent with CDL across the set of robustness checks, especially for signed discretionary accruals (*DA*). *DA* declines post-SOX in only 14 percent of the cases we examine across the three earnings targets, with half of the remaining cases showing

DA increases. Hence, SUSPECT firms' *DA* results are sensitive to the robustness checks; on the other hand, as in CDL, the REM results reflect increases and thus are mostly insensitive to the robustness checks.

Fifth, we extend the sample period to span 1992-2017 and separately investigate the entire extended post-SOX period (2002-2017) and, within it, four post-SOX sub-periods. The time trends results generally indicate AEM declines, consistent with what an extrapolation of CDL's results over the extended post-SOX period would suggest, and REM is positive but reflects both increases and declines. However, when we use multivariate analyses with interactions to test more directly the impact of SOX in the extended post-SOX period (2002-2017), we find evidence suggesting REM and AEM behave as substitutes pre-SOX (1992-2001), but substitution declines during the extended post-SOX period (2002-2017) and within each of four post-SOX sub-periods.

Summary of replication results

As for our replication of CDL, it illustrates difficulties in following what CDL did in their study and in dealing with changes in publicly available databases. First, we cannot exactly replicate CDL's Compustat sample. It has been more than a dozen years since the publication of CDL and there have been changes and updates to Compustat in the interim that likely altered the original Compustat data CDL used. Using annual Compustat and following CDL's sample selection procedures, we obtain a sample 3 percent larger than CDL's, which, following Anderson and Hopkins (2016), means statistical power differences should not be an issue in interpreting our replication results of CDL's time trends, graphical, and SUSPECT firms' analyses. However, surprisingly, our Compustat sample includes 54 percent more unique firms than CDL's Compustat sample.

Second, consistent with CDL, we find significant increases in both AEM and in CDL's aggregate REM proxy during the accounting scandal period. However, our findings post-SOX (i.e., 2002-2005) differ from CDL's results: We do not find AEM decreases and REM increases post-SOX; in fact, our

significant coefficient signs are opposite to those that CDL report. We find evidence that the opposite REM results are due to a problematic definition in computing CDL's aggregate REM proxy variable.

Third, our replication of CDL's graphical analysis suggests the declines in AEM that CDL find in their time trends analysis post-SOX mostly occur in 2002 and 2005. With regard to the relation between AEM and REM, our graphical analysis closely corresponds to CDL's plots. We combine CDL's plots of AEM and their aggregate REM proxy; along with our corresponding plots, we find these plot pairs suggest AEM and REM tend to reflect less substitution and even some tendency for AEM and REM to move together over the scandal period and somewhat so post-SOX. Finally, we are unable to replicate CDL's ExecuComp sample, and since our ExecuComp sample is substantially smaller than CDL's, our replication of their multivariate analyses would face issues of statistical power (Anderson and Hopkins 2016), so we end our attempt to replicate CDL's results at that point.

III. RE-EXAMINATION: EMPIRICAL DESIGN AND RESULTS

Re-Examination Sample Selection and Descriptive Statistics

In our re-examination of CDL's findings, we consider the empirical design changes we list in Exhibit 1; we categorize them as different sample constructions, different model specifications, different variable definitions, and an extension of CDL's sample period. As in CDL, we restrict our sample to nonfinancial firms with available data and at least eight observations in each two-digit SIC grouping per year. Using ExecuComp data, we focus initially on 1992-2005, so that we can assess the sensitivity of their results to our robustness checks. We obtain a sample of 17,011 firm-year observations for 2,110 unique firms.⁷ We winsorize the top and bottom 1 percent of sample distributions for the continuous

⁷ CDL use ExecuComp data for their multivariate analyses and they use Compustat over 1987-2005 for their time trends, graphical, and meet/just beat SUSPECT firms analysis. We use Compustat data for our replication of those analyses in section IV and use ExecuComp data for all our re-examination analyses here in section III. CDL find that ownership and management compensation variables are significant determinants of earnings management and these variables are included in ExecuComp (but not Compustat). Thus, ExecuComp data enable more properly specified multivariate models and yield a consistent sample for our re-examination analyses. ExecuComp data are available starting in 1992 and include the S&P 1500 plus firms that once

variables. Table 2 presents sample descriptive statistics and we separate observations between pre-SOX (1992-2001) and post-SOX (2002-2005) periods.⁸ In comparing the two sub-samples, we note there generally are no differences in AEM pre- versus post-SOX while REM is significantly higher post-SOX.

Show Table 2 About Here

Re-Examination: Time Trends Analysis and Results

CDL (p. 760) state their “main objective is to examine whether the degree of earnings management increased over time and reached a zenith in the period surrounding the corporate accounting scandals, and declined after the passage of SOX.” Their results are consistent with those expectations.

CDL base their AEM and REM time trends analysis on the following model:

$$Dep_{jq} = a + b \times Time + c \times SCA + d \times SOX + \varepsilon. \quad (1a)$$

The dependent variable, *Dep*, is AEM or REM, and *Dep* is regressed on a time trends variable (*Time*) measuring the difference between a current year and the first year in the dataset, and on two 0/1 event variables: *SCA* (accounting scandal period, 2000-2001), and *SOX* (post-SOX period, 2002-2005).

CDL primarily focus their analyses and discussion on (a) *ABS_DA*, the absolute value of discretionary accruals, as the AEM *Dep*, and (b) their aggregate REM proxy, denoted *RM_PROXY*, as the REM *Dep*. Note that *ABS_DA* allows both upward and downward earnings management and accrual reversals (see computational details in the Appendix of this paper). CDL (p. 766) compute *RM_PROXY* “as the sum of the standardized variables, *R_CFO*, *R_PROD*, and *R_DISX*,” that is, the three real earnings management variables, abnormal operating cash flows, abnormal production, and abnormal discretionary expenditures (defined in the Appendix). In contrast, we use Zang’s (2012) approach and compute the

were in the S&P 1500 but later were dropped, and some ExecuComp client requests. Firms on ExecuComp tend to be larger than Compustat firms, although CDL state their Compustat sample firms are larger than the average Compustat firm. In addition, ExecuComp firms tend to be more widely held and scrutinized by analysts, institutional investors, BigN auditors, and regulators relative to smaller firms that may be more likely to engage in AEM. Note, Figure 2, Panels A and B plot AEM and REM data over time using ExecuComp and Compustat, respectively; we leave for future research an in-depth examination of small firms’ AEM and REM practices.

⁸ We compare our ExecuComp sample with CDL’s ExecuComp sample in section IV.

aggregate REM proxy by multiplying abnormal discretionary expenditures by -1 and adding the result to abnormal production expenditures, since this approach seems more sensible economically.⁹ We denote Zang's 2-variable aggregate REM proxy as *RM_PROXY2*; we also adapt it by performance matching the REM variables (Cohen et al. 2020), and denote this variable as *RM_PROXY2_PM*.

Using model (1a) and Compustat data, CDL find a positive and significant *c* coefficient on *SCA* when *Dep*=*ABS_DA*; they find the same result when *Dep* is their aggregate REM proxy, *RM_PROXY*. CDL interpret their results as indicating AEM as well as REM increase in the scandal period. For the post-SOX period, aligned with the prediction that SOX would be associated with AEM decreases, CDL find a negative and significant *d* coefficient on *SOX* when *Dep*=*ABS_DA*, and a positive and significant coefficient on *SOX* when *Dep*=*RM_PROXY*.

We present a summary of the regression results using CDL's time trends model (1a) and ExecuComp data in the Online Appendix OL Table 1. The results suggest AEM declines post-SOX (i.e., when *Dep*=*ABS_DA*) for all of the robustness checks except when pooling data to estimate *ABS_DA*. When *Dep*=*RM_PROXY2*, the results confirm REM increases except when there is a change in sample construction: when we drop foreign issuers, drop delisted firms, or drop missing R&D (all three of which reduce sample size). However, if we compute the performance-matched version of Zang's aggregate REM proxy (*RM_PROXY2_PM*), the results for all the robustness checks indicate REM increases post-SOX.

Note, however, that showing the coefficient, respectively, on *SCA* is positive and on *SOX* is negative, tests for the sign, that is, the level of AEM or REM during a certain period, but not the trend of

⁹ CDL (p. 766) acknowledge the three REM variables can have different implications for reported earnings. Roychowdhury (2006) examines a setting of avoiding a loss and, in his hypothesis and empirical analysis, effectively multiplies discretionary expenditures by -1. In an early working paper, Zang (2007, Appendix I) multiplies REM expenditures by -1. Moreover, Zang (2012) examines the manipulation of two real activities that increase earnings: cutting discretionary costs (*R_DISX*) and overproducing inventory; the latter spreads fixed production costs over a larger quantity of goods and increases costs in ending inventory and thus reduces cost of sales (*R_PROD*). When constructing her aggregate REM measure, Zang uses $(-1) \times R_DISX$ since discretionary expenses reduce income. Zang (2012) notes the directional effects of *R_CFO* can be ambiguous and excludes it in estimating her aggregate REM proxy.

any change in AEM (or REM) in the scandal period or post-SOX. To more directly test for trends, we adapt model (1a) to be the following:¹⁰

$$Dep_{jq} = a + b \times Time + c1 \times SCA + c2 \times SCA \times SCA_Time + \\ d1 \times SOX + d2 \times SOX \times SOX_Time + \varepsilon \quad (1b)$$

where *SCA_Time* is the difference between a current year and 1999, and *SOX_Time* is the difference between a current year and 2001. The interaction terms allow for possible non-linear trends in the scandal and post-SOX periods.¹¹

Table 3 summarizes our re-examination results of the time trends analyses using model (1b). *Dep* is either *ABS_DA* (Panel A) or *RM_PROXY2* (Panel B); the variables of interest are both *SOX* and *SOX*×*SOX_Time*. We consider each of the robustness checks separately.¹² Row 0 displays CDL's results (which are based on model (1a) and Compustat data) for AEM (under the *Dep*=*ABS_DA* column in Panel A) and for CDL's aggregate REM proxy (under the *Dep*=*RM_PROXY2* column in Panel B). Our results based on model (1b) and ExecuComp data are in rows 1-6. In row 1 we report the benchmark case of estimating model (1b); in row 2a we replace *ABS_DA* as *Dep* with absolute value of performance-adjusted discretionary accruals (*ABS_DA_PM*); and in row 2b we replace *ABS_DA* with pooled data discretionary accruals (*ABS_DA_Pooled*). Again using *Dep*=*ABS_DA*, in row 3 we drop foreign issuers; in row 4 we drop delisted firms. In row 5a we use industry average R&D for missing R&D observations to compute

¹⁰ We thank Terry Shevlin for pointing out the limitations of using model (1a) and suggesting a classic Cook and Campbell (1979) pre-post design on a single sample. Model (1b) is consistent with that.

¹¹ *Time* controls for the overall time trends of AEM and REM in the sample periods. For model (1b), we find there is no collinearity problem with *Time*, *SCA*, *SCA_Time*, *SOX* and *SOX_Time* over the 1992-2005 sample period. Since the scandal period is a different event than the post-SOX period, in addition to separate 0/1 event variables *SCA* and *SOX*, we also need separate time trend variables for each event period. *SCA_Time* and *SOX_Time* represent the trends within their respective event periods (scandal period and post-SOX period) in addition to the 0/1 event variables, which only reflect the sign or level, for example, of AEM (but not the trend). Note, the scandal period only includes two years; while only two data points are less than desirable, it nevertheless allows for a trend estimate for the scandal period. Also note, using a single time trend variable interacted with both *SCA* and *SOX* could create multicollinearity for models (1b)-(1d).

¹² There are three exceptions: we use ExecuComp data, Zang's approach for computing the aggregate REM proxy, and clustering of standard errors by firm and year throughout all of the re-examination analyses, unless otherwise noted.

discretionary expenditures that we need to compute the aggregate REM proxy (denoted as *RM_PROXY2_SICRD*); in row 5b we drop firms with missing R&D values in computing *RM_PROXY2*. Lastly, in row 6 we include firm fixed effects. Table 3's footnote c discloses whether the performance-matched version of Zang's aggregate REM proxy (*RM_PROXY2_PM*) yields different results than when using *RM_PROXY2*.¹³

Our Table 3 results in Panel A, when *Dep=ABS_DA*, indicate the coefficient on *SOX* for the benchmark case is -0.010 and significant, and it is negative and significant for all of the robustness checks except when estimating discretionary accruals by pooling data. Moreover, the coefficient on *SOX*×*SOX_Time* is always negative and significant; for example, it is -0.005 for the benchmark case. These results indicate a lower level of AEM, as in CDL, and declines post-SOX; these results generalize CDL's findings to the typically larger ExecuComp firm.¹⁴ In Panel B, where *Dep=RM_PROXY2*, the coefficient on *SOX* for the benchmark case is 0.023 and significant, but the coefficient is significant only for two robustness checks: when using industry average R&D for missing R&D observations, and when including firm fixed effects. The *SOX*×*SOX_Time* coefficient is positive significant for the benchmark case (0.008) and when dropping foreign issuers (0.017); otherwise it is insignificant. However, when we use the performance-matched version of Zang's aggregate REM proxy (*RM_PROXY2_PM*), the coefficients on *SOX* for all but one of the robustness checks are positive and significant; only the coefficient on *SOX*×*SOX_TIME* in the robustness checks (dropping observations with missing R&D) is positive and significant. Overall, these results suggest the average level of REM is higher post-SOX but

¹³ Cohen et al. (2020) suggest reporting both performance-matched and non-performance-matched REM variables. Also, see Collins, Pungaliya, and Vijh (2020).

¹⁴ CDL (2008, 777) suggest "several simultaneous occurrences could have contributed to a decrease in earnings management activities after passage of SOX, including the increased vigilance of investors, auditors and regulators, and greater care taken by managers in financial reporting after the adverse publicity caused by the scandals." CDL are "cautious in attributing the decrease in the level of earnings management solely to the passage of SOX from [their] analysis."

there is only weak evidence about an increasing trend of REM post-SOX. See the panels of Online Appendix OL Table 2 for the detailed regression results.¹⁵

Show Table 3 About Here

For brevity, we do not table scandal period results.¹⁶ Regarding time trends, when $Dep=ABS_DA$ the coefficients on SCA across the robustness checks are generally positive and significant, suggestive of higher AEM levels during the scandal period. However, the coefficients on $SCA \times SCA_Time$ are generally negative and significant, indicating a lower level of AEM and a declining trend from the initially positive AEM. That is, although the scandal period only includes data for two years, the time trends results do not reflect a steadily increasing AEM trend in the scandal period just prior to SOX, as CDL seem to expect (p. 757, p. 760). The results when $Dep=RM_PROXY2$ generally indicate insignificant coefficients on $SCA \times SCA_Time$, but when $Dep=RM_PROXY2_PM$, the coefficients on $SCA \times SCA_Time$ are uniformly positive and significant, suggesting a higher REM level and a consistent increasing REM trend.

Using ExecuComp data over 1992-2005, Figure 1 plots Panel A plots ABS_DA and ABS_DA_PM . ABS_DA increases sharply in 2000, the first year of the scandal period, followed by a partial reversal in 2001. Post-SOX, ABS_DA declines initially in 2002 and somewhat again in 2005. The plot for ABS_DA_PM reflects smaller changes. Panel B displays plots for RM_PROXY2 and RM_PROXY2_PM . RM_PROXY2 is relatively flat whereas RM_PROXY2_PM exhibits a sharp rise in the scandal period and a sharper increase post-SOX. Compared to model (1a), model (1b) appears to better capture AEM and REM changes during the scandal and post-SOX periods.

Insert Figure 1 about Here

Re-Examination: Multivariate Analyses and Results

¹⁵ All tables and some footnotes and text include references to Online Appendix tables showing estimated regression details.

¹⁶ Scandal period time trends regression results are included in Online Appendix OL Table 2's panels.

We re-examine CDL's multivariate analyses regarding the relation between AEM and REM by considering our set of robustness checks. We base our regression models on CDL and control for auditor type (*BIG*), change in GDP (ΔGDP), market value of equity (*MKTVAL*), as well as *SCA* and *SOX*. Multivariate model (2) is:

$$\begin{aligned}
Dep_j = & \alpha_0 + \alpha_1 \times BIG_j + \alpha_2 \times \Delta GDP_j + \alpha_3 \times MKTVAL_j + \alpha_4 \times SCA_j + \alpha_5 \times SOX_j \\
& + \alpha_6 \times RM_PROXY2_j + \alpha_7 \times BONUS_j + \alpha_8 BONUS_j \times SCA_j \\
& + \alpha_9 \times BONUS_j \times SOX_j + \alpha_{10} UN_OPTION_j + \alpha_{11} UN_OPTION_j \times SCA_j \\
& + \alpha_{12} \times UN_OPTION_j \times SOX_j \\
& + \alpha_{13} GRANT_OPTION_j + \alpha_{14} GRANT_OPTION_j \times SCA_j \\
& + \alpha_{15} \times GRANT_OPTION_j \times SOX_j \\
& + \alpha_{16} EX_OPTION_j + \alpha_{17} EX_OPTION_j \times SCA_j \\
& + \alpha_{18} \times EX_OPTION_j \times SOX_j + \alpha_{19} OWNER_j + \alpha_{20} OWNER_j \times SCA_j \\
& + \alpha_{21} \times OWNER_j \times SOX_j + \varepsilon_j.
\end{aligned} \tag{2}$$

As noted above, variable definitions are in the Appendix of this paper.¹⁷

The dependent variable (*Dep*) in model (2) is *ABS_DA*, and the key independent variable is *RM_PROXY2*. We also estimate model (3), which is the same as model (2) except *Dep*=*RM_PROXY2*, and the key independent variable is *ABS_DA*. CDL find negative and significant coefficients on their aggregate REM variable and on *ABS_DA* when, respectively, each is an independent variable in model (2) and model (3).

Table 4 shows the results of our estimation of models (2) and (3). Regarding the relation between AEM and REM, we find the coefficient on *RM_PROXY2* for the benchmark case is -0.013 and significant when *Dep*=*ABS_DA*, and the coefficient on *ABS_DA* is -0.405 and significant for the benchmark case when *Dep*= *RM_PROXY2*. Moreover, the coefficients are significant for all but one of the robustness

¹⁷ We drop *Time* from model (2) since variance inflation factors (VIFs) exceed 10 for *Time*, *SOX*, and *SCA*, which indicate the presence of multicollinearity; after dropping *Time*, VIFs for *SOX* and *SCA* are less than seven.

checks. Hence, Table 4's results generally indicate a negative relation between AEM and REM, suggestive of substitution during the 1992-2005 sample period.¹⁸

Show Table 4 About Here

Direct Analysis of the Relation between AEM and REM Post-SOX

CDL predict a substitution between AEM and REM (CDL pp. 761-2). They find a negative and significant relation between *ABS_DA* and their aggregate REM proxy (*RM_PROXY*) in estimating both models (2) and (3). CDL comment on substitution results in connection with model (3), and say (p. 781) the finding “indicates that firms substitute between real and accrual-based earnings management activities, a result consistent with Graham et al. (2005 [p. 36, 66]) and Zang (2006).”

Assuming greater scrutiny of public companies' accrual practices by auditors and regulators post-SOX, it is reasonable to expect that the cost of AEM increases relative to REM. REM may be harder to detect and, as Graham et al. (2005, 36) suggest, what appear to be REM activities instead may reflect managers “taking real decisions [rather than admitting to] accounting decisions.”

The results from estimating models (2) and (3) suggest a substitution between REM and AEM. However, these results for the relation between *ABS_DA* and the aggregate REM proxy are for the entire sample period (1992-2005), not just for the post-SOX period (2002-2005). Thus, it appears CDL have not directly tested for a substitution between REM and AEM in the post-SOX period in their multivariate analyses. We believe a more direct test augments model (2) (where *Dep=ABS_DA*) by adding an interaction variable, *RM_PROXY2*×*SOX*, as well as a corresponding interaction variable for the scandal period (*RM_PROXY2*×*SCA*). The coefficient on the interaction variable provides additional evidence on the relation between REM and AEM post-SOX. Similarly, for model (3), where *Dep=RM_PROXY2*, we

¹⁸ In a robustness test, we find that the inclusion of two interactions terms *SCA*×*SCA_TIME* and *SOX*×*SOX_TIME* that we use in model (1b) do not change the coefficient results on *RM_PROXY2* when *Dep=ABS_DA*, or the coefficient results on *ABS_DA* when *Dep=RM_PROXY2*.

add the interaction variable $ABS_DA \times SOX$ (and $ABS_DA \times SCA$). These augmented models should shed additional light on the relation between REM and AEM during the *SCA* and post-SOX event periods.¹⁹

Table 5 summarizes the results for the augmented models (2) and (3) and displays the estimated coefficients on AEM and REM and the corresponding interaction variables. In Panel A for augmented model (2), the coefficient on RM_PROXY2 for the benchmark case is -0.015 and significant, indicating AEM and REM are negatively related pre-SOX, while the coefficient on $SOX \times RM_PROXY2$ is 0.015 and significant, which suggests the negative relation between AEM and REM declines post-SOX. We obtain highly similar results across the robustness checks, except for insignificant results for the two alternative measures of AEM (i.e., for ABS_DA_PM and ABS_Pooled).

The results for augmented model (3) in Panel B are similar and uniformly significant. The benchmark case shows a coefficient on ABS_DA of -0.562 and a coefficient on $SOX \times ABS_DA$ of 0.512. We find a negative relation between AEM and REM, but the negative relation declines over the post-SOX period. The results for all of the robustness checks tell the same story. That is, AEM and REM are substitutes but that substitution declines post-SOX. Further, in three cases in Panel A (dropping foreign firms, and both ways of addressing missing R&D data) and in two cases in Panel B (dropping foreign firms and using industry average data for R&D) the results suggest the decline in substitution between AEM and REM becomes a complementary relation.²⁰

Insert Table 5 About Here

SUSPECT Firms Analysis

¹⁹ Given the presence of multicollinearity, we also drop *Time* from augmented model (3). Untabulated results that include *Time* indicate VIF values in excess of 10 for both *Time* and *SOX* for all of our robustness checks. After dropping *Time*, VIFs are below 10 across the robustness checks.

²⁰ In a robustness test, including the interactions terms $SCA \times SCA_TIME$ and $SOX \times SOX_TIME$ that we use in model (1b) do not change the coefficient results on RM_PROXY2 and $RM_PROXY2 \times SOX$ when $Dep=ABS_DA$, or the coefficient results on ABS_DA and $ABS_DA \times SOX$ when $Dep=RM_PROXY2$.

Similar to CDL, we examine firms that meet/just beat three separate earnings targets to validate further the earnings management results. Using unbalanced samples CDL find, from pre- to post-SOX, significant AEM declines, specifically, declines in discretionary accruals (*DA*),²¹ and significant REM increases for each of three earnings targets: avoiding a loss, avoiding a decline from the previous year's earnings, and meeting/just beating analyst forecasts. We follow CDL's sample selection process, but use a balanced firm sample design, thereby comparing the same firms pre- and post-SOX (although the number of firm-years can differ). Using ExecuComp data, we focus on observations that significantly meet/just beat the respective earnings targets.²²

Table 6 displays results for each of the three earnings target cases across our benchmark and robustness checks. The benchmark results are as follows: for avoiding a loss, *DA* significantly increases from pre- to post-SOX by 0.019 and REM significantly increases by 0.048; for avoiding an earnings decline, *DA* insignificantly declines while REM significantly increases by 0.032; and for meeting/just beating analyst forecasts, *DA* insignificantly increases while REM significantly increases by 0.088 (see Online Appendix OL Table 7). Across the benchmark and all of the robustness checks, column A displays the results for avoiding a loss, where one robustness check indicates an insignificant change in *DA*, but six robustness checks indicate *DA* increases. REM increases in four cases but is insignificant in three cases. For column B, avoiding an earnings decline, two of the robustness checks show *DA* declines in meeting the target, four are insignificant, and one shows a *DA* increase. Five of the seven robustness checks indicate REM increases. Lastly, column C reveals that for meeting/just beating analysts' forecasts,

²¹ CDL's SUSPECT firms meet/just beat analysis assumes income-increasing activity to meet/just beat each earnings target; thus, the focus is on discretionary accruals (*DA*) rather than *ABS_DA*, since the sign of the accruals matters in this analysis.

²² As in CDL, we use observations as follows as SUSPECT firms: (i) in the interval [0, 0.005) of net income before extraordinary items (deflated by total assets) that just avoid a loss; (ii) in the interval [0, 0.005) of the change in net income before extraordinary items (deflated by total assets) that just avoid an earnings decline; and (iii) with I/B/E/S forecast errors from zero to one cent per share meeting/just beating I/B/E/S consensus analysts' earnings forecasts (Roychowdhury 2006). We note (a) we calculate forecast errors as the actual EPS minus the mean of the most recent forecasts made by individual analysts; (b) only 9 percent of our ExecuComp sample observations do not have analyst coverage, and (c) defining the I/B/E/S sample to be the interval [0, 0.005) would reduce the sample size by half.

one robustness check shows a *DA* decline, four are insignificant, and two show *DA* increases; all seven robustness checks indicate REM increases.

In aggregate, the SUSPECT firms' analyses yield mixed results, especially for *DA*. In three of 21 (or 14% of) cases examined across the three earnings targets, *DA* falls, nine cases (43%) show an insignificant change in *DA*, and nine (43%) show *DA* increases. Regarding *RM_PROXY2*, in 16 (76%) of the 21 cases, *RM_PROXY2* increases while the remaining cases are insignificant. Thus, CDL's *DA* results are mostly sensitive to robustness checks, whereas the results for *RM_PROXY2* mostly are not.

Insert Table 6 About Here

Extended Sample Period: Empirical Design and Results

We extend our ExecuComp sample to span 1992-2017 and explore whether the results for 1992-2005 generalize to the extended period. SOX became law on July 30, 2002, so less than half of 2002 is effectively in CDL's 2002-2005 post-SOX period. Moreover, DeFond (2010, 405-6) notes there was "anecdotal evidence that managers and auditors had difficulties interpreting SOX in the initial years after its adoption." For example, it was not until 2004 that the PCAOB issued Auditing Standard No. 2 to provide guidance regarding external auditors' assessments of client firms' internal controls over financial reporting (ICFR) and the relation between clients' ICFR effectiveness and external auditors' audits of clients' financial statements. Given CDL's relatively short post-SOX period and the extent to which the interpretation, guidance, and implementation of SOX changed in the initial years under SOX, our consideration of an extended sample period arguably enables us to more reliably test for a post-SOX link to earnings management and for a relation between REM and AEM. Furthermore, the extended period allows for an exploration of AEM and REM in differing macroeconomic environments (e.g., the financial crisis and slow growth economy that followed) and evolving regulatory scrutiny over financial reporting.

Using the same selection procedures as above, the extended ExecuComp sample comprises 33,315 firm-years (2,908 unique firms). We present results for the entire extended post-SOX period (2002-2017) and for four sub-periods within the extended post-SOX period. Note, of course, there are no CDL results to compare to for the extended sample analyses.

Time Trends Analysis in the Extended Sample Period

Our time trends model for the entire extended period (1992-2017) is:

$$Dep_{jq} = a + b1 \times SCA + b2 \times SCA \times SCA_Time + \\ c1 \times SOXext + c2 \times SOXext \times SOXext_Time + \varepsilon \quad (1c)$$

where *SOXext* is a 0/1 variable that equals 1 if a year is in the range 2002-2017; *SOXext_Time* is the difference between a current year and 2001; *SCA_Time* is the difference between a current year and 1999; and as before *SCA* is the scandal period, 2000-2001. *Dep* is *ABS_DA* (or *ABS_DA_PM*). The key variable of interest is *SOXext*×*SOXext_Time*. If there is greater scrutiny of AEM post-SOX, then extrapolating from CDL suggests an expectation of a negative coefficient on the interaction variable *SOXext*×*SOXext_Time* when *Dep*=*ABS_DA*. Further, if REM receives relatively less scrutiny than AEM, then when *Dep*=*RM_PROXY2* we expect a positive coefficient on *SOXext*×*SOXext_Time* post-SOX.

The four post-SOX sub-periods in the extended sample are as follows. *SOX1* equals 1 if the year is in the range 2002-2005 (as in CDL). *SOX2* equals 1 if the year is 2006 or 2007. The financial crisis (*FC*), 2008-2009, follows, and we denote it as *SOX3*, or equivalently *FC*, since *SOX3* and *FC* overlap. *SOX4* equals 1 if the year is in the range 2010-2017. Model (1d) includes the four SOX sub-periods:

$$Dep_{jq} = a + b1 \times SCA + b2 \times SCA \times SCA_Time \\ + c1 \times SOX1 + c2 \times SOX1 \times SOX1_Time \\ + d1 \times SOX2 + d2 \times SOX2 \times SOX2_Time \\ + e1 \times SOX3 + e2 \times SOX3 \times SOX3_Time$$

$$+ f1 \times SOX4 + f2 \times SOX4 \times SOX4_Time + \varepsilon \quad (1d)$$

The interaction term, $SOXn \times SOXn_Time$, is the key variable of interest.²³ When $Dep=ABS_DA$, we expect AEM declines and thus a negative coefficient on $SOX1 \times SOX1_Time$. By extension, we also expect $SOX2 \times SOX2_Time$ and $SOX4 \times SOX4_Time$ to have negative coefficients. (We consider $SOX3$, i.e., FC , below.) When $Dep=RM_PROXY2$, we expect positive coefficients on the related interaction variables, $SOX1 \times SOX1_Time$, $SOX2 \times SOX2_Time$, and $SOX4 \times SOX4_Time$, indicating REM increases.

As for the financial crisis, the resulting recession could change earnings management incentives. For example, managers may face incentives to cut expenditures and firms wishing to manage earnings upwards might then favor AEM over REM. Alternatively, some firms may not seek to manage earnings when peer firms likely have declining earnings; still other firms may opt to take a big bath. Hence, it is unclear what to expect regarding AEM and REM and thus we make no predictions regarding the trends for or the relations between AEM and REM during the financial crisis.

Table 7 presents the time trend results for the extended period.²⁴ The results for model (1c) are in Panel A. When $Dep=ABS_DA_PM$, row 1 shows the coefficient on $SOXext$ is -0.0003 and insignificant, while row 2 indicates the coefficient on $SOXext \times SOXext_Time$ is -0.0011 and significant. These results suggest AEM trends downward over the extended post-SOX period (2002-2017). When $Dep=RM_PROXY2$, while the coefficient on $SOXext$ in row 1 is 0.202 and significant, indicating a higher REM level, the interaction variable coefficient of 0.001 is insignificant, suggesting the absence of an increasing REM trend post-SOX. The results for model (1d), which divides the post-SOX period into four post-SOX sub-periods (see Panel B), show the coefficients on $SOX1-SOX4$ and the related interaction

²³ $SOXn$ represents $SOXext$ for the entire extended post-SOX period (2002-2017) or $SOX1-SOX4$ for the four post-SOX sub-periods within the extended post-SOX period.

²⁴ Before reporting the results based on our models (1c) and (1d), we apply model (1a), CDL's model, with firm fixed effects to our extended period. Online Appendix OL Table 8 summarizes the results, which suggest AEM decreases and REM increases over the extended post-SOX period.

variables. The coefficient on *SOX4* is -0.0105 and significant, indicating the AEM level during *SOX4* is lower than in the *SOX1-SOX3* sub-periods, and the significant coefficient on *SOX4*×*SOX4_Time* (-0.0009) suggests a decreasing AEM trend in the slow growth period following the financial crisis. While the coefficient on *SOX1* is insignificant, the coefficient on *SOX1*×*SOX1_Time* is -0.001 and significant; these results provide some evidence of decreasing AEM in the initial post-SOX sub-period. Surprisingly, if we replace *ABS_DA_PM* with *ABS_DA*, the findings appear to be somewhat mixed (see Online Appendix OL Table 9). They indicate an insignificant coefficient on *SOXext* for the entire extended post-SOX period; however, the sub-period results indicate positive and significant coefficients on *SOX2* and *SOX3*, implying higher levels of AEM, while the coefficient on *SOX4* is negative and significant and implying a lower level of AEM. The coefficients on *SOX2*×*SOX2_Time* and *SOX3*×*SOX3_Time* are each negative and significant, suggesting a negative AEM trend, while the coefficient on *SOX4*×*SOX4_Time* is positive and significant, suggesting a positive AEM trend. Hence, the results for AEM in the extended post-SOX sub-periods suggest somewhat offsetting trends and are sensitive to the measurement of AEM. Regarding REM, it increases over the entire extended post-SOX period and in each of the four sub-periods.

Insert Table 7 about Here

Figure 2, Panel A displays *ABS_DA_PM* and *RM_PROXY2* plots over the extended period based on ExecuComp data. *ABS_DA_PM* peaks and reverses in the scandal period, then is flat or mostly slightly decreasing post-SOX. *RM_PROXY2* increases in the scandal period and during *SOX1* and exhibits both an increase and then a decrease in *SOX2*. Panel B graphs the same variables using Compustat data, which generally reflects smaller firms vis-à-vis ExecuComp. *ABS_DA_PM* spikes then reverses in the scandal period, and gradually declines post-SOX. *RM_PROXY2* declines in the period before the accounting scandals, and drops farther before reversing in the scandal period; it also increases in *SOX3* (*FC*) and is

again volatile in *SOX4*. The Compustat data (in Panel B) appear to reflect more substitution between AEM and REM than the ExecuComp data (in Panel A).

Insert Figure 2 about Here

Multivariate Analyses in the Extended Sample Period

We base the extended sample period multivariate analysis on adapting augmented model (2) to include *SOX_n* for the entire extended period (model 2a) or for the four sub-period post-SOX variables (model 2b). As noted, there are no CDL results for the extended period. Hence, we are not assessing the sensitivity of CDL results; rather, we define a basic model and test for the significance of the predicted coefficients. The model defines *Dep*=*ABS_DA_PM* (or *ABS_DA*). It controls for *RM_PROXY2* based on Zang's 2-variable aggregate REM proxy; interacts *SOX_n* variable(s) with *RM_PROXY2*; controls for firm fixed effects and for ownership and managerial compensation; adds a 0/1 control variable (*MWEAK*) indicating the occurrence of a material weakness over internal controls in a given year;²⁵ interacts *MWEAK* with the post-SOX sub-period variables;²⁶ replaces missing R&D values with industry average R&D and an indicator that equals 1 for missing R&D observations; and clusters standard errors by firm and year. We detect the presence of multicollinearity and again drop *Time* from the model.

Model (2a) is as follows:

²⁵ *MWEAK* is linked to and affects accrual quality (Doyle, Ge, McVay 2007; Ashbaugh-Skaife, Collins, Kinney, LaFond 2008).

²⁶ Since auditing standards regarding material weaknesses were evolving in the early post-SOX years, we interact *MWEAK* only with *SOX2*, *SOX3*, and *SOX4*.

$$\begin{aligned}
Dep_j = & \alpha_0 + \alpha_1 \times BIG_j + \alpha_2 \times \Delta GDP_j + \alpha_3 \times MKTVAL_j + \alpha_4 \times SCA_j + \alpha_5 \times SOX \\
& + \alpha_6 \times RM_PROXY2_j \\
& + \alpha_7 \times RM_PROXY2_j \times SCA_j + \alpha_8 \times RM_PROXY2_j \times SOX_{ext_j} + \alpha_9 \\
& \times BONUS_j + \alpha_{10} \times BONUS_j \times SCA_j + \alpha_{11} \times BONUS_j \times SOX_j \\
& + \alpha_{12} \times UN_OPTION_j + \alpha_{13} \times UN_OPTION_j \times SCA_j \\
& + \alpha_{14} \times UN_OPTION_j \times SOX_j \\
& + \alpha_{15} \times GRANT_OPTION_j + \alpha_{16} \times GRANT_OPTION_j \times SCA_j \\
& + \alpha_{17} \times GRANT_OPTION_j \times SOX_j \\
& + \alpha_{18} \times EX_OPTION_j + \alpha_{19} \times EX_OPTION_j \times SCA_j + \alpha_{20} \times EX_OPTION_j \times SOX_j \\
& + \alpha_{21} \times OWNER_j + \alpha_{22} \times OWNER_j \times SCA_j + \alpha_{23} \times OWNER_j \times SOX_j \\
& + \alpha_{24} \times MWEAK_j + \varepsilon_j.
\end{aligned} \tag{2a}$$

Model (2b) replaces *SOXext* with the four separate extended period post-SOX sub-periods, *SOX1-SOX4*.

Focusing on the AEM results for model (2a) in Table 8, Panel A, row 1 indicates the coefficient on *RM_PROXY2* is -0.018 and significant, consistent with AEM and REM behaving as substitutes over the pre-SOX period (1992-2001). With regard to the coefficient on *SOXext*×*RM_PROXY2*, it is 0.014 and significant; this indicates that during the extended post-SOX period (2002-2017) AEM and REM reflect declining substitution. For model (2b), while the coefficient on *RM_PROXY2* is -0.019 and significant, which suggests substitution over the pre-SOX period, the coefficients on the interaction of *RM_PROXY2* and each of the four post-SOX sub-period variables range from 0.010 to 0.015 and are all significant, providing evidence of declining substitution in the post-SOX sub-periods.

Similar to models (2a) and (2b), we adapt augmented model (3), with *Dep*=*RM_PROXY2* (or *RM_PROXY2_PM*), include *ABS_DA_PM* as the AEM control, interact *ABS_DA_PM* with the post-SOX variable(s) – *SOXext* in model (3a) and *SOX1-SOX4* in model (3b) – and include the other controls.

In Table 8, Panel B, the coefficient on *ABS_DA_PM* is -1.086 and significant, suggesting REM and AEM are substitutes pre-SOX. However, the coefficient on *SOXext*×*ABS_DA_PM* is 1.028 and significant, again suggesting REM and AEM exhibit declining substitution in the extended post-SOX period (2002-2017). As for model (3b), the results follow: the coefficient on *ABS_DA_PM* is -1.083 and significant, indicating substitution between REM and AEM pre-SOX; and the coefficients on the

interaction variables range from 0.506 to 1.297 and all but the coefficient on $SOX3 \times ABS_DA_PM$ are significant, which again indicates declining substitution and even complementarity between REM and AEM in $SOX1$ of the extended post-SOX period. All regression details are in Online Appendix Table 10.

Insert Table 8 about Here

Figure 2, Panel C plots regression coefficients on ABS_DA and RM_PROXY2 over time based on ExecuComp data.²⁷ The coefficients on ABS_DA and RM_PROXY2 appear to move together over most of the entire period. Both coefficients increase in the scandal period. Post-SOX they exhibit considerable volatility, declining then increasing in $SOX1$, increasing then declining in $SOX2$ and again in $SOX3$, and are very volatile again in $SOX4$. In addition, we observe the coefficients on ABS_DA and RM_PROXY2 as independent variables being greater than zero in several years, suggesting a complementary relation.

Additional Analysis of ABS_DA

Hribar and Nichols (2007) raise concerns about using absolute value of discretionary accruals to test for earnings management. They show mean ABS_DA varies with the standard deviation of signed DA from a first-stage estimation, and state that tests based on ABS_DA are “exposed to a class of correlated omitted variables” (p. 1049). The likely impact is over-rejection of the null of no earnings management.

We investigate this in our multivariate analyses by adopting Hribar and Nichol’s suggestions to mitigate this potential bias.²⁸ Specifically, we include operating volatility variables – standard deviation of cash flows from year t-5 to t (σ_{CFO}), standard deviation of cash-based revenue from year t-5 to t (σ_{REV}), and log of total assets ($LogTA$) – as additional controls in the multivariate analyses. We note $LogTA$ is highly correlated with market capitalization (MVE), which is already a control variable in our multivariate regressions, and thus we drop MVE from the regressions. Relative to our previously reported results, we

²⁷ We estimate the following regression annually for $Dep=ABS_DA$: $Dep_j = \alpha_0 + \alpha_1 \times BIG_j + \alpha_2 \times \Delta GDP_j + \alpha_3 \times MKTVAL_j + \alpha_4 \times RM_PROXY2_j + \alpha_5 \times BONUS_j + \alpha_6 \times UN_OPTION_j + \alpha_7 \times GRANT_OPTION_j + \alpha_8 \times EX_OPTION_j + \alpha_9 \times OWNER_j + \alpha_{10} \times MWEAK_j + \varepsilon_j$. When $Dep=RM_PROXY2$, we substitute ABS_DA for RM_PROXY2 as a RHS variable.

²⁸ This does not affect our later replication of CDL where we seek to duplicate as closely as possible what CDL did.

find the declining substitution between AEM and REM post-SOX in Table 5, Panel A and in Table 8, Panel A generally continue to hold. See the Online Appendix OL Table 11 for the tabulated results.²⁹

IV. REPLICATION

Replication – Time Trends Analyses

Note, first, that CDL obtain their time trends analysis sample from the Compustat annual industrial and research files for 1987-2005. However, it has been more than a dozen years since CDL's study was published, and given subsequent restatements and Compustat's update feature (e.g., regarding restated numbers) as well as mergers, acquisitions, and bankruptcies, it is doubtful a researcher could download a dataset that is exactly the same as CDL used by simply using the Compustat annual files. We investigated using data from Compustat's Snapshot Point-in-Time database, which would facilitate re-creating the data CDL used; unfortunately, Snapshot data only became effective around 2007, which is after CDL's sample period ended.³⁰ Thus, we use the Compustat annual database (as of October 2018). Following CDL's sample selection criteria, we obtain 89,806 firm-years having the necessary data from Compustat for 1987-2005 (see Online Appendix OL Table 13); our sample is 3 percent larger than CDL's 87,217 firm-years. Surprisingly, CDL's sample consists of 8,157 unique firms, whereas our sample has 12,566 firms.³¹

After double-checking our procedures and coding, we find substantial differences between descriptive statistics for our Compustat sample and for CDL's Compustat sample, especially for *Growth of Sales*, operating cycle in days (*OC*), *Leverage*, and *Market-to-Book*.³² We contacted the CDL co-

²⁹ We summarize multivariate ownership and management compensation results in Online Appendix OL Table 12.

³⁰ On WRDS, one can see Compustat Snapshot data before 2006. However, based on further investigation by WRDS, Snapshot data were loaded onto the WRDS platform after 2006 and data before that on Snapshot now reflect changes due to M&A, restatements, etc., and thus we cannot assume the data would be what was in CDL's actual sample. A Compustat representative suggested we use Compustat-Legacy (an old Compustat dataset that was discontinued and stopped reflecting restatement changes made after 2007) since this could be the database CDL may have used. Using Compustat-Legacy we obtain a sample of size and number of unique firms similar to what we report using our Compustat sample.

³¹ It was after approaching WRDS to try to understand why our sample differed from the number of unique firms in CDL's sample (to no avail) that we learned Snapshot data are not available for CDL's sample period.

³² See Online Appendix OL Table 13, which lists some major differences in descriptive statistics with and without winsorizing relative to CDL's descriptive statistics. Note, CDL (2008, 769) mention winsorizing when noting they check for outliers by

authors and described some of the major differences. We heard back that they are unable to answer the specific questions we had raised, which is understandable given the length of time since they had worked on the paper. Unfortunately, their data and programs are somewhere in computer storage at a former university of one of the co-authors and cannot be accessed. Thus, we use our Compustat sample data, and it is clear it is not possible to replicate exactly CDL's Compustat-based analyses. Nevertheless, since our Compustat sample size is slightly larger than CDL's, our sample should be sufficient to allow us to determine whether we can confirm CDL's main findings (Anderson and Hopkins 2016).³³

Time Trends Results

Following CDL, we use our Compustat sample and model (1a) and separately regress the AEM and REM variables on a time trends variable (*Time*), which measures the difference between a current year in the replication period and 1987 for each observation, and on two event dummy variables.

Table 9 reports our time trends results based on our Compustat sample. Our replication of CDL's analysis yields results consistent with CDL's findings of significantly higher levels of *ABS_DA* and *RM_PROXY* in the scandal period. However, post-SOX (2002-2005) our results differ; CDL find negative and significant AEM and positive and significant REM; we find insignificant effects for both.

CDL cite an early working paper by Zang (2006). As we note in footnote 9, we found a 2007 version of the same working paper on SSRN that indicates Zang (2007, Appendix I) multiplied discretionary expenditures by -1. Adopting Zang's (2012) approach for the aggregate REM proxy, we find a significant coefficient on *RM_PROXY2* of 0.028, consistent with CDL's results of positive REM post-SOX. Perhaps CDL adopted Zang's approach in estimating the aggregate REM proxy in their time trends analysis but failed to indicate that in the paper.

winsorizing the top and bottom 1 percent of the distribution, and state (p. 769) their "cross-sectional results and inferences are not affected by repeating all of [their] analyses using the winzorized sample (results not tabulated)."

³³ Anderson and Hopkins (2016, 139) argue that a replication's sample size "should be at least as large as the original to ensure statistical power differences are not the reason for disparate original and replicated findings."

Graphical Analysis

Like CDL, we graphically illustrate AEM and REM time trends. In Figure 3, both CDL and our plot of *ABS_DA* overlap considerably, and show *ABS_DA* increases in 2000 and declines in 2001 (i.e., the scandal period). Post-SOX, CDL's plot shows *ABS_DA* is flat for 2002-2004 and declines in 2005; our plot shows *ABS_DA* declines in 2002, increases in 2003 and 2004, and declines somewhat in 2005. Thus, the AEM declines CDL find in their time trends plot post-SOX appear to occur mostly in 2002 and 2005. Our Online Appendix OL Figure 1 shows our and CDL's plots of *Positive_DA* and *Negative_DA*. Post-SOX, CDL's plot of *Positive_DA* shows a slight increase over 2002-2004 and a sharp decline in 2005; we find a sharper increase in 2002-2004 and a slight decline in 2005. CDL find *Negative_DA* is flat for 2002-2004 and increases slightly in 2005, while we show a decline in 2003 and a slight increase in 2005.

Figure 4 shows a relatively close correspondence of CDL's and our plots of the relation between AEM and REM over time. Both sets of plots suggest little substitution between REM and AEM post-SOX; the plot pairs only weakly suggest substitution between REM and AEM over some intervals prior to the scandal period.

Meet/Just Beat SUSPECT Firms Analysis

CDL's SUSPECT firms' analysis uses unbalanced samples and indicates both *DA* declines and *RM_PROXY* increases for avoiding a loss, for avoiding an earnings decline, and for meeting/just beating analysts' forecasts. CDL do not say they use Compustat data nor disclose their SUSPECT firms sample sizes; we assume they use Compustat since its database is larger than ExecuComp's. Using unbalanced samples, our results (in Online Appendix OL Table 14) indicate the following: for avoiding a loss, *DA* increases significantly from pre- to post-SOX, opposite to CDL's finding; for both avoiding an earnings decline and meeting/just beating analysts' forecasts, *DA* reflects insignificant changes. Regarding the aggregate REM proxy, using CDL's *RM_PROXY* variable, we find no significant changes for avoiding a

loss and avoiding an earnings decline but significant REM declines for meeting/just beating analysts' forecasts, also opposite CDL's results. Thus, we cannot replicate CDL's results. However, if we estimate the aggregate REM proxy by adapting Zang's approach for estimating the aggregate REM proxy and use balanced samples, the results for avoiding a loss indicate significant REM increases post-SOX, although the other two earnings targets are insignificant.

Multivariate Analyses

CDL use ExecuComp data for their multivariate analyses. However, CDL and our ExecuComp samples differ. Online Appendix OL Table 15 indicates almost all of the variables have means differing significantly between CDL's sample and our sample. Hence, we cannot replicate CDL's ExecuComp sample.³⁴ Moreover, our sample (n=17,011) is substantially smaller than CDL's. Given Anderson and Hopkins's (2016) recommendation, which *JFR* has adopted, that a replication sample should at least be as large as the original sample, we ceased our replication of the CDL's multivariate analyses at this point.³⁵

V. CONCLUSION: SUMMARY AND DISCUSSION

Summary

We re-examine and replicate the highly cited study by Cohen, Dey, and Lys (2008) regarding AEM and REM pre- and post-SOX. Our re-examination seeks to determine the extent to which CDL's results continue to hold by assessing the sensitivity of their results to empirical design changes (i.e., robustness checks) spanning different sample constructions, model specifications, and variable definitions, and an

³⁴ CDL report an ExecuComp sample of 31,668 firm-years for 2,018 firms (CDL p. 762 and their Table 1, Panel B). Note, 14 years of data (1992-2005) for 2,018 firms implies a maximum possible sample of 28,252 firm-years. CDL do not report sample sizes in their Tables 3 and 4 where they report their multivariate results, and those tables state the sample period is 1987-2005 rather than 1992-2005; but recall that ExecuComp data start in 1992.

³⁵ The ExecuComp database, for the version we downloaded (August 2018), contains 3,462 firms, both active and liquidated. ExecuComp's collection of S&P 1500 firms' data began in 1994. Some ExecuComp data go back to 1992 but, apparently, that is mostly S&P 500 firms. ExecuComp has a monthly update feature and no equivalent to Compustat's Snapshot (point-in-time) database. Overall, there may be differences between our version of ExecuComp and CDL's (likely 2005) version.

extended sample period. It is important to state that most studies likely would not stand up to such a large set of robustness checks.

Our main re-examination findings follow. First, partially consistent with CDL, our time trends results suggest AEM increases then declines in the scandal period (2000-01), and REM generally increases. Post-SOX (2002-05), AEM declines while REM increases, consistent with CDL. We base our time trends analysis on ExecuComp data, which on average reflect larger firms, and thus mostly generalize CDL's Compustat-based results. The findings are somewhat mixed when we extend our sample to 2017.

Second, with regard to our multivariate analyses of the relation between AEM and REM, the results suggest they behave as substitutes pre-SOX (1992-2001). We augment CDL's multivariate models by including interactions between AEM and *SOX* (and REM and *SOX*) to test more directly the relation between AEM and REM post-SOX (2002-2005). We find consistent evidence of a negative relation between AEM and REM pre-SOX period (1992-2001), which suggests REM and AEM behave as substitutes; whereas post-SOX (2002-2005), the evidence indicates declining substitution and even complementarity, which also holds for the extended post-SOX period (2002-2017).

Third, our SUSPECT firms meet/just beat analysis compares earnings management strategies pre-versus post-SOX (2002-2005) for three earnings targets. We find mixed results across the robustness checks, especially for discretionary accruals. *DA* declines post-SOX in 14 percent of cases examined across the three earnings targets, indicating the SUSPECT firms' *DA* results are sensitive to the robustness checks. On the other hand, the aggregate REM proxy generally increases post-SOX, and the REM results are mostly unaffected when considering our robustness checks.

Finally, regarding our replication, we are unable to replicate exactly CDL's Compustat sample. In the dozen years since CDL was published Compustat made changes and updates, likely altering the original data CDL used. We obtain a Compustat sample that is slightly larger than CDL's, which means

statistical power differences should not be an issue in interpreting our replication results (Anderson and Hopkins 2016). We note, however, our Compustat sample includes substantially more unique firms than CDL's Compustat sample.

Consistent with CDL, our replication finds time trends results showing significant increases in AEM (*ABS_DA*) and in CDL's aggregate REM proxy (*RM_PROXY*) in the scandal period. However, post-SOX (i.e., 2002-2005) our results differ. CDL report significant AEM decreases and significant REM increases. We do not find those changes; rather, our significant coefficient signs are opposite to those CDL report. Our replication of CDL's graphical evidence suggests the post-SOX AEM declines that CDL find in their time trends analysis mostly occur in 2002 and 2005. When we combine CDL's plots of *ABS_DA* and *RM_PROXY* and compare them to our corresponding plots, there is a relatively close correspondence. AEM and REM appear to move together in the scandal period, and somewhat so post-SOX; there is only a weak suggestion of any substitution between REM and AEM, and it is mostly prior to the scandal period. Finally, we cannot replicate CDL's ExecuComp sample. Moreover, since our ExecuComp sample is substantially smaller than CDL's, there are concerns about statistical power (Anderson and Hopkins 2016); hence, we ended our effort to replicate CDL's multivariate analyses.

Discussion

We have learned a great deal from the re-examination and replication of CDL. First, there is evidence AEM increases sharply but then declines in the accounting scandal period, suggestive of widespread AEM during at least the early part of the scandal period.

Second, our evidence that AEM generally declines post-SOX, but less consistently in the extended sample period, should be replicated or re-examined. Prior research mostly finds AEM declines post-SOX, but like CDL, most studies focus on a relatively short time period post-SOX. For example,³⁶ Lobo and

³⁶ Reviewer #3 identified all of the prior studies we discuss here, and similarly below where we discuss substitution between AEM and REM.

Zhou (2006) consider less than two years (through 2003) post-SOX. Koh, Matsumoto, and Rajgopal's (2008) analyses end in the second quarter of 2006, and they note their "results provide only early evidence on Enron's legacy" (p. 1093). Bartov and Cohen's (2009) sample ends in the fourth quarter of 2006; Baber, Kang, and Li's (2011) sample includes 2004-2007 as the post-SOX period; and Krishnan, Win, and Zhao's (2011) pre- versus post-SOX comparison includes only 2003 as post-SOX. While Zang's (2012) post-SOX sample ends in 2008, only 14.3 percent of her sample observations are post-SOX, as her study does not focus on a pre- versus post-SOX comparison. We believe it is fair and appropriate to ask whether the AEM declines that CDL and other studies report as occurring post-SOX, continue to hold in the many years beyond what prior research has examined. Our extended sample period analysis enables us to begin further examining that question.

Third, our results document considerable evidence REM increases post-SOX (2002-05), although there is mixed evidence in our extended post-SOX sample period (2002-2017). Pre-SOX, the commonly held view was that REM was more costly than AEM, since REM implies managing real activities whereas AEM simply requires making adjusting entries in the accounting system. However, at the margin, SOX was expected to make AEM more costly (risky) for managers, likely due to the expected greater scrutiny of firms and their managers by auditors and regulators, and to the criminal penalties associated with failing to meet SOX's requirement (Section 302) of CEO/CFO certifications as to the fair presentation of their firms' financial statements. Future research might further investigate REM increases post-SOX. It is also important to examine the occurrence and extent of legal action related to the CEO/CFO certifications.

Fourth, our results indicate considerable evidence that substitution between AEM and REM occurs mostly pre-SOX, while the results also suggest AEM and REM behave less as substitutes and sometimes more like complements post-SOX. That is, there appears to have been a change in firms' use of earnings management tools post-SOX such that substitution declines and even in some cases complementarity is a

better characterization of the relation between AEM and REM. We note CDL's results regarding substitution between AEM and REM are widely cited.³⁷ Zang (2012) finds substitution over her entire sample period (1987-2008), but, as noted, her post-SOX sample represents less than 15 percent of her observations and her evidence of substitution mostly occurs pre-SOX. Kama and Melumad (2019) examine substitution over 1989-2008, and again most of the evidence of substitution is pre-SOX. Vorst (2016) investigates whether abnormal cuts in discretionary expenditures (e.g., R&D and SG&A) that reverse within a year are predictive of the extent of REM. His analyses include an examination of whether proxies for incentives to engage in REM and/or that reflect its costs and benefits, such as meeting earnings targets or external monitoring of accruals, are useful in predicting abnormal cuts in real activities that reverse. Vorst posits (p. 1236) "the willingness to engage in REM should increase if accruals are more closely monitored," citing CDL for that prediction. Vorst's sample period ends in 2012; post-SOX he finds evidence supporting his prediction when the monitoring variable is a BigN auditor, but he finds only limited support when the monitoring variable is regulators or analysts monitoring accruals. Thus, there is only limited evidence of AEM and REM substitution since post-SOX there is not a strong link between the extent of REM increases and AEM declines.

We believe it is also reasonable and appropriate to ask if the findings of substitution between REM and AEM continue to hold as prior research has reported. We find substitution between REM and AEM pre-SOX, but it attenuates post-SOX. Furthermore, as part of our robustness checks, we augment CDL's multivariate models by adding interaction variables ($RM_PROXY2 \times SOX$, and $ABS_DA \times SOX$). The coefficients on the interaction variables provide additional evidence post-SOX on the relation between REM and AEM. The results using this approach again indicate substitution between REM and AEM pre-SOX, but declining substitution and even evidence of complementarity post-SOX. Our results should

³⁷ A sampling of literature reviews citing this result include Beyer, Cohen, Lys, and Walther 2010; Dechow, Ge, and Schrand 2010; DeFond and Zhang 2014; Coates and Srinivasan 2014; Keune, Keune, and Quick 2017.

caution researchers against automatically assuming AEM and REM substitution is the expected relation in general. Future research that replicates or re-examines our findings of declining substitution between AEM and REM post-SOX would be useful.

Recall, our re-examination analyses investigate the sensitivity of CDL's results to a battery of robustness checks; these are exploratory analyses since with one exception, we, like CDL, examine changes in AEM and REM and their relation in a general setting, not in contexts in which we expect firms to have greater incentives to manage earnings. As a result, no consideration is given as to which individual robustness check is (or which checks are) most relevant regarding our assessments of earnings management using AEM and/or REM.

The exception we referred to is the SUSPECT firms' analysis. With regard to Zang's (2012) aggregate REM proxy, in the case of meeting/just beating analysts' earnings forecasts, the results are generally insensitive to the robustness checks; however, that is not the case for avoiding a loss or for avoiding an earnings decline. With regard to AEM (specifically, to signed discretionary accruals), the expected result of AEM declines appears sensitive to multiple robustness checks for all three earnings targets, and especially for avoiding a loss. Understanding why this is so is worthy of further investigation.

While we have considered a large set of robustness checks in our re-examination of CDL, other robustness checks remain unexplored, and perhaps could be the basis for future research. These might include Stubben's (2010) approach for estimating AEM based on abnormal revenues, or making REM and AEM decisions sequentially, as in Pincus and Rajgopal (2002) and Zang (2012). More generally, future research might explore earnings management measures beyond AEM and REM and/or interactions with them, including major restatements, GAAP-based lawsuits, fraud and enforcement cases, and recessionary cash crunches.

Re-Examinations and Replications

“When scientists cannot confirm the results from a published study, to some it is an indication of a problem, and to others, it is a natural part of the scientific process that can lead to new discoveries” (National Academies of Sciences, 2019, p. 1, quoted in Hail et al. 2020).

Replications and re-examinations are important and positively contribute to a discipline. We believe that, similar to other disciplines, accounting will benefit from a serious and continuing dose of the production and publication of replications and re-examinations. An academic culture in which replications and re-examinations are routinely published means researchers are continually shedding light on the state of the art in various research areas; this fosters the research conversation within the discipline about methodological and underlying substantive issues, as well as about the implications of current research and possible future research directions.³⁸ Timely execution of replications and re-examinations also reduces the likelihood publicly available databases will change – due to updating, format changes, and database withdrawals, thereby raising the risk that exact replications become impossible.

Hail et al. (2020) have contributed to the accounting literature as well as to the discussion of research reproducibility across disciplines by providing survey evidence on the reproducibility of research results. By re-examining and replicating a widely cited publication that deals with SOX and earnings management, two important topics in financial accounting, we have contributed to the accounting literature by providing new evidence on the robustness and generalizability of CDL’s results. We hope accounting journals will follow *JFR*’s lead and publish re-examinations and replications, and, furthermore, that departmental and university promotion and tenure reviews of accounting (and other business school) scholars move to seriously encourage replications and re-examinations by significantly increasing the weight such publications receive in faculty reviews. We believe many other well-cited articles could fruitfully be re-examined and/or replicated.

³⁸ “Scholarly work is rooted in the lively exchange of ideas – conversation at its best...Written work is the most enduring and often the most influential contribution a scholar makes to academic conversation” (Huff, 1999, p. 3).

REFERENCES

- Anderson, S., and P. E. Hopkins. 2016. Commentary on: Experienced financial managers' views of the relationships among self-serving attribution bias, overconfidence, and the issuance of managerial forecasts: A replication. *Journal of Financial Reporting* 1(1): 137-142.
- Ashbaugh-Skaife, H., D. Collins, W. Kinney, R. LaFond, 2008. The effect of SOX internal control deficiencies and their remediation on accrual quality. *The Accounting Review* 83: 217-250.
- Baber, W.R., S.H. Kang, and Y. Li. 2011. Modeling discretionary accrual reversal and the balance sheet as an earnings management constraint. *The Accounting Review* 86(4): 1189-1212.
- Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6: 159-178.
- Ball, R., and P. Brown. 2019. Ball and Brown (1968) after fifty years. *Pacific-Basin Finance Journal* 53: 410-431.
- Banks, G., E. Boyle Jr., J. Pollack, C. White, J. Batchelor, C. Whelpley, K. Abston, A. Bennett, and C. Adkins. 2016. Questions about questionable research practices in the field of Management: A guest Commentary. *Journal of Management* 42 (Jan.): 5-20.
- Bartov, E., and D. Cohen. 2009. The “numbers game” in the pre-and post-Sarbanes-Oxley eras. *Journal of Accounting, Auditing & Finance* 24(4): 505-534.
- Beyer, A., D. Cohen, T. Lys, and B. Walther. 2010. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* 50(2-3): 296-343.
- Black, E.L., Christensen, T.E., Taylor Joo, T. and Schmardebeck, R., 2017. The relation between earnings management and non-GAAP reporting. *Contemporary Accounting Research*, 34(2), pp.750-782.
- Brickey, K. 2003. From Enron to WorldCom and beyond: Life and crime after Sarbanes-Oxley. *Washington University Law Quarterly* 81: 357–401.
- Camerer, C., et al. 2018. Evaluating replicability of laboratory experiments in economics. *Science* 351: 1433-1436.
- Chen, W., P. Hribar, and S. Melessa. 2018. Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research* 56(3): 751-796.
- Coates, J.C., and S. Srinivasan. 2014. Sox after ten years. *Accounting Horizons* 29(3): 627-671.
- Cohen, D., A. Dey, and T. Lys. 2008. Real and accrual-based earnings management in the pre- and post-Sarbanes-Oxley periods. *The Accounting Review* 83(3): 757-787.
- Cohen, D., S. Pandit, C. Wasley, and T. Zach. 2020. Measuring real activity management. *Contemporary Accounting Research* 37(2): 1172-1198.

- Collins, D., R. Pungaliya, and A. Vijh. 2020. Analytics behind type I error rates and power in earnings management research. Working paper, University of Iowa.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Cunningham, L. 2003. The Sarbanes-Oxley yawn: Heavy rhetoric, light reform (and it might just work). *Connecticut Law Review* 35: 915–89.
- Cunningham, L.M., Johnson, B.A., Johnson, E.S. and Lisic, L.L., 2020. The Switch - Up: An Examination of Changes in Earnings Management after Receiving SEC Comment Letters. *Contemporary Accounting Research*, 37(2), pp.917-944.
- Dechow, P., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50(2-3): 344-401.
- Dechow, P., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70(2): 193-225.
- Dechow, P., A. Hutton, J. Kim, and R. Sloan. 2012. Detecting earnings management: A new approach. *Journal of Accounting Research* 50(2): 275-334.
- DeFond, M. 2010. Earnings quality research: Advances, challenge and future research. *Journal of Accounting and Economics* 5(1-2): 402-409.
- DeFond, M., and J. Zhang. 2014. A review of archival auditing research. *Journal of Accounting Research* 58(2-3): 275-326.
- Doyle, J., W. Ge, and S. McVay. 2007. Accruals quality and internal control over financial reporting, *The Accounting Review* 82(5): 1141-1170.
- Goodman, S., D. Fanelli, and J. Ioannidis. 2016. What does research reproducibility mean? *Science Translational Medicine* 8(341): 1-6.
- Gow, I.D., G. Ormazabal, and D.J. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85(2): 483-512.
- Graham, J., C. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40(1-3): 3-73.
- Hail, L., M. Lang, and C. Leuz. 2020. Reproducibility in Accounting Research: Views of the Research Community. *Journal of Accounting Research* 58(2): 519-543.
- Hribar, P., and C. Nichols. 2007. The use of unsigned earnings quality measures in tests of Earnings management. *Journal of Accounting Research* 45(5): 1017-1053.

- Huff, A.S. 1999. *Writing for scholarly publication*. Thousand Oaks, Sage Publications.
- Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29: 193-228.
- Kama, I., and Melumad, N., 2019. Camouflaged indicators of earnings management. *European Accounting Review* 29(2): 361-382.
- Katz, D., and McIntosh, L., 2019. SEC Scrutiny of Non-GAAP Financial Measures. *Harvard Law School Forum on Corporate Governance*.
- Keune, M., T. Keune, and L. Quick. 2017. Voluntary changes in accounting principles: Literature review, descriptive data and opportunities for future research. *Journal of Accounting Literature* 39: 52-81.
- Koh, K., D.A. Matsumoto, and S. Rajgopal. 2008. Meeting or beating analyst expectations in the post-scandals world: Changes in stock market rewards and managerial actions. *Contemporary Accounting Research* 25(4): 1067-1098.
- Koh, P-S., and D. Reeb. 2015. Missing R&D. *Journal of Accounting and Economics* 60(1): 73-94.
- Kothari, S.P., A. Leone, and C. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39(1): 163-197.
- Krishnan, J., Y. Wen, and W. Zhao. 2011. Legal expertise on corporate audit committees and financial reporting quality. *The Accounting Review* 86(6): 2099-2130.
- Li, H., M. Pincus, and S. Rego. (2008). Market reaction to events surrounding the Sarbanes-Oxley Act of 2002 and earnings management. *Journal of Law & Economics* 51(February): 111-134.
- Lobo, G.J., and J. Zhou. 2006. Did conservatism in financial reporting increase after the Sarbanes-Oxley Act? Initial evidence. *Accounting Horizons* 20(1): 57-73.
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349 (6251): 2-8.
- Perino, M. 2002. Symposium: Enron and its aftermath. Enron's legislative aftermath: Some reflections on the deterrence aspects of the Sarbanes-Oxley Act of 2002. *St. John's Law Review* 76: 671-98.
- Petersen, M. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22(1): 435-480.
- Pincus, M., and S. Rajgopal. 2002. The interaction between accrual management and hedging: Evidence from oil and gas firms. *The Accounting Review* 77(1): 127-160.

Ribstein, L. 2002. Market vs. regulatory responses to corporate fraud: A critique of the Sarbanes-Oxley Act of 2002. *Journal of Corporation Law* 28: 1–67.

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

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

Vorst, P., 2016. Real earnings management and long-term operating performance: The role of reversals in discretionary investment cuts. *The Accounting Review* 91(4): 1219-1256.

Zang, A. 2007. Evidence on the tradeoff between real manipulation and accrual manipulation. Working paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=961293.

Zang, A. 2012. Evidence on the trade-off between real activities and accrual-based earnings management. *The Accounting Review* 87(2): 675-703.

APPENDIX

Variable Definitions

<i>Total Assets</i>	= total assets (Compustat data item <i>AT</i>) at the fiscal year-end
<i>LogTA</i>	= the natural logarithm of total assets
<i>Market Capitalization</i>	= the firm's market value of common equity [the price per share (Compustat data item <i>PRCC_F</i>) times the number of shares outstanding (Compustat data item <i>CSHO</i>)] at the fiscal year-end t
<i>MKTVAL</i>	= natural logarithm of <i>Market Capitalization</i>
<i>Sales</i>	= sales (Compustat data item <i>SALE</i>) at the fiscal year-end t
<i>Growth of Sales</i>	= the difference between sales at the fiscal year-end t and sales at the fiscal year-end t-1, then divided by sales at the fiscal year-end t-1
<i>OC (Days)</i>	= the operating cycle (in days), calculated as $\frac{(AR_t - AR_{t-1})/2}{\left(\frac{Sales}{360}\right)} + \frac{(INV_t - INV_{t-1})/2}{\left(\frac{COGS}{360}\right)}$, where <i>AR</i> is accounts receivable (Compustat data item <i>RECT</i>) and <i>INV</i> is inventory (Compustat data item <i>INVT</i>)
<i>Leverage</i>	= total liabilities (Compustat data item <i>DLTT</i> + <i>LCT</i>) at the fiscal year end t divided by total assets at fiscal year-end t
<i>Market-to-Book</i>	= market value of common equity at the fiscal year end t divided by the book value of common equity (Compustat data item <i>CEQ</i>) at fiscal year-end t
<i>Book-to-Market</i>	= the book value of common equity (Compustat data item <i>CEQ</i>) at fiscal year-end t divided by the market value of common equity at the fiscal yearend t
σ_{CFO}	= Standard deviation of cash flows divided by lagged total assets, computed over the period t-5 to t
σ_{REV}	= Standard deviation of cash-based revenues (revenues + Δ accounts receivable) divided by lagged total assets, computed over the period t-5 to t
<i>Total Accruals</i>	= the difference between accrual earnings and cash earnings, calculated as the difference between earnings before extraordinary items and discontinued operations and operating cash flow, scaled by beginning total assets: $[IBC_t - (OANCF_t - XIDOC_t)]/AT_{t-1}$
<i>DA</i>	= discretionary accruals computed using the Modified Jones Model: $\frac{TA_{it}}{Assets_{i,t-1}} = k_{1t} \frac{1}{Assets_{i,t-1}} + k_{2t} \frac{\Delta Sales_{it}}{Assets_{i,t-1}} + k_{3t} \frac{PPE_{it}}{Assets_{i,t-1}} + \varepsilon_{it},$ where $\Delta Sales$ is change in sales from the preceding year and <i>PPE</i> is the gross value of property, plant and equipment (Compustat data item <i>PPEGT</i>)
<i>Positive_DA</i>	= <i>DA</i> with positive value, and missing with negative value
<i>Negative_DA</i>	= <i>DA</i> with negative value, and missing with positive value
<i>ABS_DA</i>	= the absolute value of <i>DA</i> ³⁹
<i>ABS_DA_PM</i>	= the absolute value of performance-matched discretionary accruals
<i>R_CFO</i>	= the level of abnormal cash flow from operations estimated from: $\frac{CFO_{it}}{Assets_{i,t-1}} = k_{1t} \frac{1}{Assets_{i,t-1}} + k_{2t} \frac{Sales_{it}}{Assets_{i,t-1}} + k_{3t} \frac{\Delta Sales_{it}}{Assets_{i,t-1}} + \varepsilon_{it},$ where <i>CFO</i> (Compustat data item <i>OANCF</i> - <i>XIDOC</i>) is cash flow from operation in period t

³⁹ Fewer than 20 observations of *DA* equal 0, and they are dropped.

<i>R_PROD</i>	= the level of abnormal production costs, where production costs are defined as the sum of cost of goods sold and the change in inventories, estimated from: $\frac{PROD_{it}}{Assets_{i,t-1}} = k_{1t} \frac{1}{Assets_{i,t-1}} + k_{2t} \frac{Sales_{it}}{Assets_{i,t-1}} + k_{3t} \frac{\Delta Sales_{it}}{Assets_{i,t-1}} + k_{4t} \frac{\Delta Sales_{it-1}}{Assets_{i,t-1}} + \varepsilon_{it}$, where <i>PROD</i> is defined as the sum of cost of goods sold (Compustat data item <i>COGS</i>) and the changes in inventories
<i>R_DISX</i>	= the level of abnormal discretionary expenses, where discretionary expenses are the sum of advertising expenses, R&D expenses, and SG&A expenses, estimated from: $\frac{DISX_{it}}{Assets_{i,t-1}} = k_{1t} \frac{1}{Assets_{i,t-1}} + k_{2t} \frac{Sales_{it-1}}{Assets_{i,t-1}} + \varepsilon_{it}$, where <i>DISX</i> is defined as the sum of advertising expenses (Compustat data item <i>XAD</i>), R&D expenses (Compustat data item <i>XRD</i>), and SG&A expenses (Compustat data item <i>XSGA</i>)
<i>RM_PROXY</i>	= the sum of the standardized three REM proxies, i.e., <i>R_CFO</i> , <i>R_PROD</i> and <i>R_DISX</i>
<i>RM_PROXY2</i>	= the sum of the standardized <i>R_PROD</i> and (-1)* <i>R_DISX</i>
<i>Time</i>	= fiscal year minus 1987
<i>SCA</i>	= 1 if the year is either 2000 or 2001
<i>SOX</i>	= 1 for years 2002-2005, and 0 otherwise
<i>SOXext</i>	= 1 for years 2002-2017, and 0 otherwise
<i>SOX1</i>	= a dummy variable equals 1 if the year of observation is in 2002-2005
<i>SOX2</i>	= a dummy variable equals 1 if the year of observation 2006-2007
<i>SOX3 (FC)</i>	= a dummy variable equals 1 if the year of observation is in 2008 or 2009
<i>SOX4</i>	= a dummy variable equals 1 if the year of observation is in 2010-2017
<i>BONUS (%)</i>	= the average bonus compensation as a proportion of total compensation received by the CEO and CFO of the firm
<i>BONUS (\$)</i>	= the <i>BONUS</i> variable in ExecuComp in thousands of dollars
<i>EX_OPTION (%)</i>	= exercisable options, which is the number of unexercised options that the executive held at year-end were vested scaled by total outstanding shares of the firm
<i>UN_OPTIONS (%)</i>	= unexercisable options defined as the number of unexercised options that the executive held at year-end that had not vested scaled by total outstanding shares of the firm
<i>GRNT_OPTION (%)</i>	= number of new option grants made during the current period scaled by total outstanding shares of the firm
<i>OWNER</i>	= the sum of restricted stock grants in the current period and the aggregated number of shares held by the executive at year-end (excluding stock options) scaled by total outstanding shares of the firm
<i>BIG</i>	= a dummy variable equals to 1 if the company is audited by a Big 4/5 auditor
<i>ΔGDP</i>	= the percent change in the real gross domestic product from the previous year
<i>MWEAK</i>	= 1 if there is at least one internal control (SOX 302 or SOX 404) material weakness, and 0 otherwise

FIGURE 1

Panel A: Re-Examination: Absolute Value of Discretionary Accruals over Time, 1992-2005 (ExecuComp Data)

Panel B: Re-Examination: Real Earnings Management Proxies over Time, 1992-2005 (ExecuComp Data)

FIGURE 2

Panel A: Re-Examination: Comparison of Absolute Discretionary Accruals and Real Earnings Management Proxies over the Extended Sample Period, 1992-2017 (ExecuComp Data)

Panel B: Re-Examination: Comparison of Absolute Discretionary Accruals and Real Earnings Management Proxies over the Extended Sample Period, 1987-2017 (Compustat Data)

Panel C: Re-Examination: Regression Coefficients on ABS_DA and RM_PROXY2 Over 1992-2017 (ExecuComp Data)

Figure 3

Replication: Absolute Value of Discretionary Accruals over Time, 1987-2005 (Compustat Data)
(Our plot and CDL's plot from their Figure 2)

FIGURE 4

Replication: Comparison of Absolute Discretionary Accruals and Real Earnings Management Proxies over Time, 1987-2005 (Compustat Data)

(Our plots and CDL's plots from their Figure 2 and from their Figure 4)

Panel A: CDL Plots

Panel B: Our plots

Exhibit 1

Robustness and Generalizability Checks

For our re-examination analyses, we investigate the following robustness checks to assess the sensitivity of CDL's results (our related table numbers are listed).

Changes to sample construction:

- (i) We use ExecuComp data for time trends (Table 3), multivariate (Tables 4 & 5), SUSPECT firms meet/just beat (Table 6), and extended sample analyses (Tables 7 & 8). ExecuComp data provide a consistent database for our re-examination analyses (see footnote 6) and enable us to determine whether CDL's time trends results, which are based on CDL's Compustat data, hold for firms that on average are larger than Compustat firms.
- (ii) We drop foreign issuers traded in the U.S. (Tables 3, 4, 5 & 6). Such firms are subject to different accounting standards and operate in different institutional environments.
- (iii) We drop delisted firms (Tables 3, 4, 5 & 6), thus eliminating firms due to bankruptcy, liquidation, or acquisition.
- (iv) We use balanced samples in our analyses of SUSPECT firms (Table 6) and thus examine the same firms pre- and post-SOX in testing whether firms meet/just beat earnings targets.

Changes to model specifications:^a

- (v) We alter CDL's time trends model by including interaction terms to test trends as well as the sign of AEM and REM post-SOX (Table 3).
- (vi) We include firm fixed effects (Tables 3, 4, 5 & 7).^b Sample firms can appear in a varying number of years and firm fixed effects control for unobserved and time-invariant characteristics that can vary across firms.
- (vii) In assessing statistical significance, we cluster standard errors by firm and year to adjust for time-series and cross-sectional dependence (Petersen 2009; Gow, Ormazabal, and Taylor 2010) (Tables 3, 4, 5 & 7).
- (viii) We augment CDL's multivariate models by including interaction terms to test more directly earnings management in the post-SOX period (Table 5). Thus, for example, we interact *SOX* with an AEM (or an REM) variable to capture the relation between REM and AEM post-SOX.
- (ix) Also in the multivariate analyses, we drop the *Time* variable due to the presence of multicollinearity involving *SOX* and *Time* (Tables 4 & 5).

^a We do not employ the suggestion of Chen, Hribar, and Melessa (2018) to use a single-stage (rather than a two-step) regression approach when estimating discretionary earnings management variables. For much of CDL's and our analyses, the focus is on (a) the absolute value of abnormal accruals and (b) an aggregate REM proxy. Consider accruals. To get positive and negative discretionary accruals we have to use the first-stage regression (a Jones-type model) and then, given the residuals from the first-stage regression, separate the residuals by sign into positive and negative groups. Thus, without doing the first-stage regression we would not be able to identify firms' sets of positive and negative discretionary accruals in a given year. While the single-stage regression approach that Chen et al. propose has the advantage of yielding less biased coefficients, it assumes that the first-stage regression residual is the same as the dependent variable in the second stage of the two-step regressions *without* any transformation (e.g., taking absolute values, or subsets of the signed abnormal accruals). A single-stage regression cannot apply to our setting since our second-stage variable is the absolute value of discretionary accruals. Furthermore, while the Chen et al. approach can apply to individual real earnings management variables, it cannot apply to the aggregate REM proxy for the same reason since the aggregate REM proxy is the sum of residuals from different models.

^b Alternatively, we use industry fixed effects (untabulated) and the results are qualitatively the same.

Changes in the definitions of accruals-based and real earnings management variables:

- (x) We use performance matching to alternatively estimate discretionary accruals (Kothari, Leone, and Wasley 2005)^c and abnormal REM variables (Cohen, Pandit, Wasley, and Zach 2020) (Tables 3, 4, 5 & 6).
- (xi) We also estimate discretionary accruals using pooled data (Dechow, Hutton, Kim, and Sloan 2012) (Tables 3, 4, 5 & 6).
- (xii) We use the approach proposed by Zang (2012), of multiplying real discretionary expenditures by -1 and adding the result to real production expenditures, to estimate an aggregate REM proxy, which we denote as *RM_PROXY2* (and *RM_PROXY2_PM* for the performance-matched version) (Tables 3, 4, 5 & 6).
- (xiii) Regarding missing R&D values, we: (a) follow Koh and Reeb (2015) and substitute missing R&D observations with a firm's 2-digit SIC industry average R&D, coupled with an indicator variable equaling 1 for missing R&D observations (and 0 otherwise), to estimate real discretionary expenditures;^d (b) alternatively, we drop missing R&D observations (Tables 3, 4, 5 & 6).

Extended sample period:

- (xiv) We extend the sample period to 2017 to determine whether the results found in CDL's sample period, which ends in 2005, generalize to the extended post-SOX period, 2002-2017 (Tables 7 & 8).
- (xv) For the extended sample period's time trends and multivariate analyses (Tables 7 & 8), we drop the *Time* variable due to the presence of multicollinearity.^e
- (xvi) For the extended sample period, we include an indicator control variable in the multivariate model (Table 8) that equals 1 if a firm receives a material weakness in its internal controls over financial reporting in a given year.

^c CDL use Dechow, Sloan, and Sweeney's (1995) modified Jones (1991) model discretionary accruals in their main tests. In two untabulated robustness tests, CDL include the change in receivables as a deduction from revenues in the first-stage accrual estimation, and alternatively use performance-matched accruals.

^d CDL state (p. 762) they "require that each firm-year observation has the data necessary to calculate the discretionary accruals metrics and real earnings management proxies..." This description is somewhat incomplete for cases of missing R&D data, at least with regard to the estimation of discretionary expenditures, which include R&D. A firm may incur no R&D expenditures, in which case we can replace missing values with zero. Or, a firm's R&D can be non-zero but the database does not separately report it since the firm may include R&D in SG&A expenses. Based on CDL's sample size, we believe they replaced missing values for R&D expense with zero, since if they had dropped firms with missing R&D data their sample size would have sharply declined. We note that Koh and Reeb (2015) find that for firms with missing or zero R&D on Compustat, they could link the absence of R&D to the presence of patent activity, indicating R&D activity. Koh and Reeb (2015) suggest using a firm's industry average R&D and an indicator variable if R&D data are missing.

^e One can argue there is no reason to assume that earnings management always grows. For example, in the time trends model, the time trends variable (*Time*) yields an estimate of the average change in earnings management over the entire sample period. CDL's (and our) primary focus is on the post-SOX period, so controlling for the average trend in earnings management over the entire sample period allows us, in principle, to estimate any change in the post-SOX period beyond the average trend over the entire sample period. However, the time trends variable and *SOX* period indicator variable are highly positively correlated in our sample ($r \approx 0.79$); hence, we perform checks for multicollinearity for our analyses and find it is generally present.

Exhibit 2
Table of Contents

0.	Abstract
I.	Introduction: Motivation and Brief Overview
II.	Background and Summary of Results
	A. Why Re-examination and Replication are Important:
	B. Re-examination: Robustness Checks List and Rationales (see Exhibit 1)
	1. Changes in sample construction
	2. Changes in model specification
	3. Changes in variable definitions
	4. Extended sample period
	C. Why CDL?
	D. Summary of Re-examination and Replication Results
III.	Re-examination analyses – ExecuComp Sample
	1. Time trends analysis: Model and results
	a. More direct trend analysis
	2. Multivariate analyses: Models and results
	a. More direct analyses with post-SOX interaction terms
	3. Meet/Just beat SUSPECT firms' analyses and results
	4. Extended sample period
	a. Time trends analysis: Models and results
	b. Multivariate analysis with post-SOX interaction terms: Models and results
	4. Additional analysis regarding absolute value of discretionary accruals
III.	Replication Analyses: Compustat Sample
	1. Time trends models and results
	2. Graphical analyses
	3. Meet/just beat SUSPECT firms' analysis and results
	4. ExecuComp sample (Multivariate)
IV.	Conclusions
	A. Summary
	B. Discussion
	C. Re-examinations and Replications
V.	References
VI.	Appendix: Variable Definitions
VII.	Online Appendix of Tables and Figure

Table 1
Re-Examination Findings Comparison

#	CDL's Findings	Our Re-examination Findings	Consistent?
Time Trend of AEM and REM			
1	AEM increases in the accounting scandal period (2000-01)	AEM initially increases but <i>then declines</i> in the accounting scandal period (2000-01)	Partially
2	AEM declines post-SOX (2002-05)	AEM declines post-SOX (2002-05) but <i>mixed evidence</i> in the extended period (2002-17)	Yes/Mixed
3	REM increases post-SOX (2002-05)	REM increases post-SOX (2002-05) but <i>mixed evidence</i> in the extended period (2002-17)	Yes/Mixed
4	AEM decreases and REM increases in the SUSPECT firms meet/just beat analysis	<i>Weak evidence</i> AEM decreases; stronger evidence REM increases	Weak/Yes
Relation between AEM and REM			
5	REM and AEM are substitutes during the entire sample period (1992-2005) including post-SOX (2002-2005)	REM and AEM are substitutes pre-SOX (1992-2001), but <i>substitution declines and even some evidence of complementarity</i> post-SOX (2002-05 and 2002-17)	Weak/No

Table 2
Descriptive Statistics
1992-2005 (n = 17,011; 2,110 unique firms)

Variable	1992-2001, n = 11,934			2002-2005, n = 5,077			(4)-(1) Diff in Means
	(1) Mean	(2) Median	(3) Std Dev	(4) Mean	(5) Median	(6) Std Dev	
<i>Total Assets</i> (millions)	2845.53	731.09	5654.98	4248.28	1063.48	8638.85	10.64***
<i>LogTA</i>	6.76	6.59	1.57	7.13	6.97	1.59	13.94***
<i>Market Capitalization</i> (millions)	3453.69	809.43	7722.87	4711.59	1163.67	9936.36	8.046***
<i>Sales</i> (millions)	2799.73	796.52	5141.15	3975.59	1066.88	7935.68	9.73***
<i>Growth of Sales</i>	0.19	0.10	0.45	0.12	0.09	0.30	-11.09***
<i>OC (Days)</i>	131.88	112.52	350.31	119.67	103.90	83.21	-3.58***
<i>Leverage</i>	0.19	0.17	0.18	0.18	0.16	0.19	-3.76***
<i>Market to Book</i>	4.16	2.45	64.78	3.81	2.25	80.95	-0.27
<i>Book to Market</i>	0.13	0.40	28.34	-1.12	0.42	113.61	-0.78
σ_{CFO}	0.08	0.05	0.09	0.08	0.05	0.08	-0.72
σ_{REV}	0.37	0.23	0.44	0.35	0.23	0.39	-2.94***
<i>Total Accruals</i>	-0.06	-0.05	0.11	-0.07	-0.06	0.11	-6.53***
<i>DA</i>	0.01	0.01	0.11	0.01	0.02	0.10	1.93*
<i>Positive_DA</i>	0.07	0.05	0.07	0.06	0.05	0.07	-3.59***
<i>Negative_DA</i>	-0.07	-0.04	0.10	-0.07	-0.04	0.12	-0.64
<i>ABS_DA</i>	0.07	0.04	0.08	0.07	0.04	0.09	-1.55
<i>ABS_DA_PM</i>	0.06	0.04	0.07	0.06	0.04	0.07	-2.30**
<i>ABS_DA_POOL</i>	0.07	0.04	0.09	0.07	0.04	0.09	-0.93
<i>R_CFO</i>	0.05	0.05	0.12	0.06	0.06	0.12	4.14***
<i>R_PROD</i>	-0.10	-0.12	0.26	-0.10	-0.13	0.25	0.62
<i>R_DISX</i>	0.05	0.00	0.27	0.02	0.00	0.25	-6.40***
<i>RM_PROXY</i>	0.00	-0.04	0.19	-0.02	-0.05	0.18	-5.23***
<i>RM_PROXY2</i>	0.00	0.04	0.30	0.04	0.05	0.28	7.39***
<i>RM_PROXY2_PM</i>	0.08	0.01	0.65	0.30	0.08	0.94	15.31***
<i>BONUS (%)</i>	545.39	271.32	822.52	818.20	422.00	1110.18	15.77***
<i>BONUS(\$)</i>	0.21	0.19	0.18	0.22	0.19	0.18	1.32
<i>EX_OPTIONS (%)</i>	1.49	0.93	1.73	2.07	1.50	1.93	18.40***
<i>UN_OPTIONS (%)</i>	1.33	0.89	1.39	1.23	0.89	1.23	-4.66***
<i>GRNT_OPTIONS (%)</i>	0.84	0.44	1.12	0.65	0.39	0.87	-11.74***
<i>OWNER (%)</i>	2.24	0.46	5.53	2.44	0.61	5.86	2.03**

Table 2 is based on our ExecuComp sample.

See the Appendix for variable definitions.

***, **, * indicate, respectively, a two-tail significance level of 1%, 5%, or 10%.

Table 3
Summary of Time Trends
Results for Test Variable SOX
(SOX Period 2002-2005)

Panel A

Dep=ABS_DA

Independent variables:	<u>n</u>	<i>SOX</i>	<i>SOX</i> × <i>SOX_TIME</i>
(0) CDL report <i>Model (1a)</i>	b	-0.015***	n/a
(1) Benchmark <i>Model (1b)</i>	17,011	-0.010***	-0.005***
(2a) Using <i>ABS_DA_PM</i>	17,011	-0.011***	-0.003***
(2b) Using <i>ABS_DA_Pooled</i>	17,011	-0.004	-0.009***
(3) Drop foreign issuers	16,560	-0.010***	-0.004***
(4) Drop delisted firms	12,777	-0.009**	-0.004***
(5a) Industry avg. for missing R&D	17,011	n/a	n/a
(5b) Drop missing R&D	11,066	-0.016***	-0.006***
(6) Firm fixed effects	16,959	-0.007**	-0.004***

Panel B

Dep=RM_PROXY2^a

Independent variables:	<u>n</u>	<i>SOX</i>	<i>SOX</i> × <i>SOX_TIME</i> ^c
(0) CDL report <i>Model (1a)</i>	b	0.041***	n/a
(1) Benchmark <i>Model (1b)</i>	17,011	0.023***	0.008*
(2a) Using <i>ABS_DA_PM</i>	17,011	n/a	n/a
(2b) Using <i>ABS_DA_Pooled</i>	17,011	n/a	n/a
(3) Drop foreign issuers	16,560	0.008	0.017***
(4) Drop delisted firms	12,777	0.006	0.008
(5a) Industry avg. for missing R&D	17,011	0.100**	0.034
(5b) Drop missing R&D	11,066	-0.031	0.007
(6) Firm fixed effects	16,959	0.023**	-0.000

See the Appendix for variable definitions.

Table 3's results in rows (1)-(6) are based on our ExecuComp data (spanning 1992-2005) using model (1b). See section IV for a discussion of CDL's ExecuComp sample.

Model (1b) is: $Dep_{jq} = a + b \times Time + c1 \times SCA + c2 \times SCA \times SCA_Time + d1 \times SOX + d2 \times SOX \times SOX_Time$.

Note 1: Details of regression results are in Online Appendix OL Table 2 (Panels A, Ba, Bb, C, D, Ea, Eb, F).

Note 2: There is no evidence of the presence of collinearity in the period 1992-2005 due to a correlation between *Time* and *SOX* (that is, no variance inflation factor, VIF, is in excess of 10).

^a CDL's aggregate REM proxy (*RM_PROXY*) sums the three REM variables. *RM_PROXY2* is Zang's 2-variable aggregate REM proxy, which multiplies *R_DISX* by -1 and sums the result with *R_PROD*. *RM_PROXY2_PM* is the performance-matched version of Zang's 2-variable aggregate REM proxy; and *RM_PROXY2_SICRD* is the aggregate REM proxy in which a firm's industry average R&D replaces missing R&D values.

^b CDL reports results for model (1a) analyzing time trends based on Compustat data for 1987-2005. See Online Appendix OL Table 1-Model (1a) for our results based on model (1a) using ExecuComp data (1992-2005).

^c Using *RM_PROXY2_PM* (instead of *RM_PROXY2*) yields positive and significant coefficients on *SOX* for robustness checks (1), (3), (4), (5b), and (6), and on *SOX*×*SOX_TIME* for robustness check (5b). See Online Appendix OL Table 2 (Panels A, C, D, Ea, Eb).

***, **, * indicate, respectively, a two-tail significance level of 1%, 5%, or 10%. We cluster standard errors by firm and year.

Tables 4
Multivariate Model (2) and (3) Results:
Relations between AEM and REM
(SOX Period 2002-2005)

		Model (2)	Model (3)
<i>Dependent variable:</i>		<i>ABS_DA</i>	<i>RM_PROXY2</i> ^a
Independent variable	n	<i>RM_PROXY2</i>	<i>ABS_DA</i>
(0) CDL report	^b	-0.427***	-0.083***
(1) Benchmark	17,011	-0.013**	-0.405***
(2a) Using <i>ABS_DA_PM</i>	17,011	-0.018***	-0.822***
(2b) Using <i>ABS_DA_Pooled</i>	17,011	-0.025***	-0.720***
(3) Drop foreign issuers	16,560	-0.012**	-0.351***
(4) Drop delisted firms	12,777	-0.011**	-0.365***
(5a) Using industry average for missing R&D (<i>RM_PROXY2_SICRD</i>)	17,011	-0.002	-0.304
(5b) Drop missing R&D	11,066	-0.018***	-0.473***
(6) Firm fixed effects	16,959	-0.031***	-0.205***

See the Appendix for variable definitions.

Note 1: Details of regression results are in Online Appendix OL Table 3 (Panels A, Ba, Bb, C, D, Ea, Eb, F) and Online Appendix OL Table 4 (Panels A, Ba, Bb, C, D, Ea, Eb, F).

Note 2: There is evidence of the presence of multicollinearity for *Time*, *SOX*, and *SCA* (i.e., VIFs for each in excess of 10); we drop *Time* from model (2) and VIFs for *SOX* and *SCA* decline to less than seven.

^a *RM_PROXY2* is Zang's 2-variable aggregate REM proxy. CDL's aggregate REM proxy (*RM_PROXY*) sums the three REM variables.

^b Our Table 4 results are based on our ExecuComp sample data (spanning 1992-2005). Model (2) is:

$$\begin{aligned}
 Dep_j = & \alpha_0 + \alpha_1 \times BIG_j + \alpha_2 \times \Delta GDP_j + \alpha_3 \times MKTVAL_j + \alpha_4 \times SCA_j + \alpha_5 \times SOX_j + \alpha_6 \times RM_PROXY2_j \\
 & + \alpha_7 \times BONUS_j + \alpha_8 \times BONUS_j \times SCA_j + \alpha_9 \times BONUS_j \times SOX_j \\
 & + \alpha_{10} \times UN_OPTION_j + \alpha_{11} \times UN_OPTION_j \times SCA_j + \alpha_{12} \times UN_OPTION_j \times SOX_j \\
 & + \alpha_{13} \times GRANT_OPTION_j + \alpha_{14} \times GRANT_OPTION_j \times SCA_j + \alpha_{15} \times GRANT_OPTION_j \times SOX_j \\
 & + \alpha_{16} \times EX_OPTION_j + \alpha_{17} \times EX_OPTION_j \times SCA_j + \alpha_{18} \times EX_OPTION_j \times SOX_j \\
 & + \alpha_{19} \times OWNER_j + \alpha_{20} \times OWNER_j \times SCA_j + \alpha_{21} \times OWNER_j \times SOX_j + \varepsilon_j.
 \end{aligned}$$

Model (3) is the same as model (2) except *Dep*=*RM_PROXY2*, and the key independent variable is *ABS_DA*.

***, **, * indicate, respectively, a two-tail significance level of 1%, 5%, or 10%. We cluster standard errors by firm and year.

Table 5
Multivariate Analyses
with Interactions for
Augmented Models (2) and (3)
(SOX Period 2002-2005)

Panel A: Augmented Model (2)

		<u>Augmented Model (2)</u> <u>Dep=ABS_DA</u>	
Independent variables:	n	RM_PROXY2	SOX×RM_PROXY2
(1) Benchmark Model (2)	17,011	-0.015***	0.015**
(2a) Using ABS_DA_PM ^{b1}	17,011	-0.018***	0.007
(2b) Using ABS_DA_Pooled	17,011	-0.024***	0.008
(3) Drop foreign issuers	16,560	-0.014***	0.017***
(4) Drop delisted firms	12,777	-0.016***	0.015***
(5a) Industry avg. for missing R&D	17,011	-0.007*	0.008**
(5b) Drop missing R&D	11,066	-0.024***	0.026***
(6) Firm fixed effects	16,959	-0.032***	0.014***

Panel B: Augmented Model (3)

		<u>Augmented Model (3)</u> <u>Dep=RM_PROXY2</u>	
Independent variables:	n	ABS_DA	SOX×ABS_DA
(1) Benchmark Model (3)	17,011	-0.562***	0.512***
(2a) Using ABS_DA_PM ^{b2}	17,011	-0.844***	0.298*
(2b) Using ABS_DA_Pooled	17,011	-0.783***	0.292*
(3) Drop foreign issuers	16,560	-0.548***	0.568***
(4) Drop delisted firms	12,777	-0.669***	0.622***
(5a) Industry avg. for missing R&D	17,011	-0.594***	0.897***
(5b) Drop missing R&D	11,066	-0.750***	0.742***
(6) Firm fixed effects	16,959	-0.340***	0.304***

See the Appendix for variable definitions.

Note 1: Details of regression results are in Online Appendix OL Table 5 (Panels A, Ba, Bb, C, D, Ea, Eb, F), and Online Appendix OL Table 6 (Panels A, Ba, Bb, C, D, Ea, Eb, F).

Note 2: There is evidence of the presence of multicollinearity for *Time*, *SOX*, and *SCA* (i.e., VIFs for each in excess of 10); we drop *Time* from model (3) and VIFs for *SOX* and *SCA* decline to less than eight.

^a Table 5's results are based on our ExecuComp sample data (1992-2005). Augmented model (2) is:

$$\begin{aligned}
 Dep_j = & \alpha_0 + \alpha_1 \times BIG_j + \alpha_2 \times \Delta GDP_j + \alpha_3 \times MKTVAL_j + \alpha_4 \times SCA_j + \alpha_5 \times SOX + \alpha_6 \times RM_PROXY2_j \\
 & + \alpha_7 \times RM_PROXY2_j \times SCA_j + \alpha_8 \times RM_PROXY2_j \times SOX_j + \alpha_9 \times BONUS_j + \alpha_{10} \times BONUS_j \times SCA_j \\
 & + \alpha_{11} \times BONUS_j \times SOX_j + \alpha_{12} \times UN_OPTION_j + \alpha_{13} \times UN_OPTION_j \times SCA_j \\
 & + \alpha_{14} \times UN_OPTION_j \times SOX_j + \alpha_{15} \times GRANT_OPTION_j + \alpha_{16} \times GRANT_OPTION_j \times SCA_j \\
 & + \alpha_{17} \times GRANT_OPTION_j \times SOX_j + \alpha_{18} \times EX_OPTION_j + \alpha_{19} \times EX_OPTION_j \times SCA_j \\
 & + \alpha_{20} \times EX_OPTION_j \times SOX_j + \alpha_{21} \times OWNER_j + \alpha_{22} \times OWNER_j \times SCA_j + \alpha_{23} \times OWNER_j \times SOX_j + \varepsilon_j.
 \end{aligned}$$

Augmented model (3), where *Dep*=*RM_PROXY2*, adds the interaction variable *ABS_DA*×*SOX* (and *ABS_DA*×*SCA*).

^{b1} In Panel A, when *RM_PROXY2_PM* is the independent variable, the coefficients on *RM_PROXY2_PM* and *SOX*×*RM_PROXY2_PM* are -0.005*** and 0.009***, respectively.

^{b2} In Panel B, when *Dep* is *RM_PROXY2_PM*, the coefficients on *ABS_DA_PM*, and *SOX*×*ABS_DA_PM* are -0.483*** and 1.187***, respectively.

***, **, * indicate, respectively, a two-tail significance level of 1%, 5%, or 10%. We cluster standard errors by firm and year.

Table 6
Meet/Just Beat SUSPECT Firms Analysis
Direction of Change in AEM and REM from Pre-SOX to Post-SOX^b
Balanced Firm Samples
Three Earnings Targets
(1992-2005)

	Panel A (Avoid a Loss)		Panel B (Avoid an Earnings Decline)		Panel C (Meet/Just Beat Analysts' Forecasts)	
	<i>DA</i>	<i>RM_PROXY2</i>	<i>DA</i>	<i>RM_PROXY2</i>	<i>DA</i>	<i>RM_PROXY2</i>
(0) CDL report ^{a, c}	-	+	-	+	-	+
(1) Benchmark	+	+	Insig	+	Insig	+
(2a) Performance matching	Insig	+	-	+	Insig	+
(2b) Pooled accruals	+	+	-	+	-	+
(3) Drop foreign issuers	+	Insig	Insig	Insig	+	+
(4) Drop delisted firms	+	Insig	Insig	Insig	+	+
(5a) Industry average for missing R&D	+	+	Insig	+	Insig	+
(5b) Drop missing R&D	+	Insig	+	+	Insig	+

See the Appendix for variable definitions.

Table 6 summarizes the directional change in *DA* and in the aggregate REM proxy for each earnings target.

^aCDL use Compustat data for 1987-2005 and unbalanced samples; we use our ExecuComp data and balanced firm samples.

^b + (-) for *DA* means AEM increases (decreases); + (-) for *RM_PROXY2* means REM increases (decreases).

^cCDL's aggregate REM proxy is *RM_PROXY*; following Zang (2012), we use *RM_PROXY2*.

Note 1: Details of univariate results are in Online Appendix OL Table 7 (Panels A, Ba, Bb, C, D, Ea, Eb).

Note 2: Sample sizes across the robustness checks range: for avoiding a loss from 117 to 214 pre-SOX and from 192 to 340 post-SOX; for avoiding an earnings decline from 422 to 748 pre-SOX and from 651 to 1,088 post-SOX; and for meeting/just beating analysts' forecasts from 849 to 1,187 pre-SOX and from 954 to 1,382 post-SOX.

Insig means the change in *DA* or the aggregate REM proxy coefficient is insignificant at the 0.1 level.

Table 7

**Summary Time Trends Results
(Extended SOX Period 2002-2017)**

<i>Dependent Variable</i>	<i>ABS_DA_PM</i>	<i>RM_PROXY2</i>
<i>Independent Variable</i>		
Panel A: Model (1c)		
(1) <i>SOXext</i>	-0.0003	0.202***
(2) <i>SOXext</i> × <i>SOXext_TIME</i>	-0.0011***	0.001
Panel B: Model (1d)		
(3) <i>SOX1</i>	-0.002	0.095***
(4) <i>SOX1</i> × <i>SOX1_TIME</i>	-0.0013*	0.059***
(5) <i>SOX2</i>	-0.0000	0.450***
(6) <i>SOX2</i> × <i>SOX2_TIME</i>	-0.0038	-0.142***
(7) <i>SOX3 (FC)</i>	-0.0025***	0.103***
(8) <i>SOX3</i> × <i>SOX3_TIME</i>	-0.0029	0.019*
(9) <i>SOX4</i>	-0.0105***	0.104***
(10) <i>SOX4</i> × <i>SOX4_TIME</i>	-0.0009***	0.022***

See the Appendix for variable definitions.

Table 7 is based on ExecuComp data for the period 1992-2017. n = 33,315.

SOXext spans 2002-2017; *SOX1* spans 2002-2005; *SOX2* spans 2006-2007; *SOX3* (=FC) spans 2008-2009; and *SOX4* spans 2010-2017.

Models (1c) and (1d) include firm fixed effects. Model (1c) is: $Dep_{jq} = a + c1 \times SCA + c2 \times SCA \times SCA_Time + d1 \times SOXext + d2 \times SOXext \times SOXext_Time$. Model (1d) is: $Dep_{jq} = a + c1 \times SCA + c2 \times SCA \times SCA_Time + d1 \times SOX1 + d2 \times SOX1 \times SOX1_Time + e1 \times SOX2 + e2 \times SOX2 \times SOX2_Time + f1 \times SOX3 + f2 \times SOX3 \times SOX3_Time + g1 \times SOX4 + g2 \times SOX4 \times SOX4_Time$.

We drop *Time* from both models (1a) and (1b) due to VIF vales in excess of 10 when *Time* is in the model; there is no indication of multicollinearity after *Time* is dropped from the model.

Note: Replacing *ABS_DA_PM* with *ABS_DA* yields similar results for (2), insignificant for (4), negative and significant for (6) and (8), and positive and significant for (10). Replacing *RM_PROXY2_PM* with *RM_PROXY2* yields qualitatively identical results.

Note 1: Details of regression results are in Online Appendix OL Table 9 (Panels A & B). Also, find in the Online Appendix OL Table 9 results based on CDL's model 1 (our model (1a)) adapted to the extended sample period.

***, **, * indicate, respectively, a two-tail significance level of 1%, 5%, or 10%. We cluster standard errors by firm and year.

Table 8

Multivariate Analyses with Interactions for Augmented Models (2) and (3)
(Extended SOX Period 2002-2017)

Panel A: Dependent Variable=ABS_DA_PM

Independent variables:	n	RM_PROXY2	SOXext/SOXn×RM_PROXY2
(1) Model (2a)	33,216	-0.018***	0.014***
Model (2b): ^b	33,216	-0.019***	
(2) SOX1			0.015***
(3) SOX2			0.010**
(4) SOX3 (FC)			0.013***
(5) SOX4			0.015***

Panel B: Dependent Variable=RM_PROXY2

Independent variables:	n	ABS_DA_PM	SOXext/SOXn×ABS_DA_PM
(1) Model (3a) ^c	33,216	-1.086***	1.028***
Model (3b): ^d	33,216	-1.083***	
(2) SOX1			1.297***
(3) SOX2			0.795***
(4) SOX3 (FC)			0.506
(5) SOX4			1.019***

See the Appendix for variable definitions. Table 8 is based on our ExecuComp sample data for 1992-2017.

^a SOX_n is SOXext or SOX1, SOX2, SOX3, or SOX4. SOXext spans 2002-2017; SOX1 spans 2002-2005; SOX2 spans 2006-2007; SOX3 (= FC) spans 2008-2009; and SOX4 spans 2010-2017.

Model (2a) is:

$$Dep_j = \alpha_0 + \alpha_1 \times BIG_j + \alpha_2 \times \Delta GDP_j + \alpha_3 \times MKTVAL_j + \alpha_4 \times SCA_j + \alpha_5 \times SOXext_j + \alpha_6 \times RM_PROXY2_j + \alpha_7 \times SCA_j \times RM_PROXY2_j + \alpha_8 \times SOXext_j \times RM_PROXY2_j + \alpha_9 \times BONUS_j + \alpha_{10} \times BONUS_j \times SCA_j + \alpha_{11} \times BONUS_j \times SOXext_j + \alpha_{12} \times UN_OPTION_j + \alpha_{13} \times UN_OPTION_j \times SCA_j + \alpha_{14} \times UN_OPTION_j \times SOXext_j + \alpha_{15} \times GRANT_OPTION_j + \alpha_{16} \times GRANT_OPTION_j \times SCA_j + \alpha_{17} \times GRANT_OPTION_j \times SOXext_j + \alpha_{18} \times EX_OPTION_j + \alpha_{19} \times EX_OPTION_j \times SCA_j + \alpha_{20} \times EX_OPTION_j \times SOXext_j + \alpha_{21} \times OWNER_j + \alpha_{22} \times OWNER_j \times SCA_j + \alpha_{23} \times OWNER_j \times SOXext_j + \alpha_{24} \times MWEAK_j + \varepsilon_j.$$

Model (2b) is the same as model (2a) except we replace SOXext with the four sub-SOX period indicators (SOX1-SOX4) and change the interaction terms with the ownership and each compensation variable by interacting each earnings management determinant with each of the four SOX sub-period indicator variables (SOX1-SOX4).

^c Model (3a) is the same as model (2a) except Dep is RM_PROXY2 and the earnings management control variable is ABS_DA in place of RM_PROXY2.

^d Model (3b) is the same as model (3a) except we replace SOXext with the four sub-SOX period indicators (SOX1-SOX4) and change the interaction terms with the ownership and each compensation variable by interacting each earnings management determinant with each of the four SOX sub-period indicator variables (SOX1-SOX4).

Note 1: Replacing ABS_DA_PM with ABS_DA or replacing RM_PROXY2 with RM_PROXY2_PM yields qualitatively identical results.

Note 2: Details of regression results are in Online Appendix OL Table 10 for Panel A (Panels A2a, A2b) and for Panel B (Panels B3a, B3b).

***, **, * indicate, respectively, a two-tail significance level of 1%, 5%, or 10%. We cluster standard errors by firm and year.

Table 9

Replication of CDL's Earnings Management Metrics: Time Trends and Correlation Matrix

Time Trend Analysis of Earnings Management Metrics over Time, 1987-2005
(n=89,806)

Dependent Variables	$Dep_{jt} = a + b * Time + c * SCA + d * SOX$				Adjusted R ²
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{d}$	
ABS_DA	0.072***	0.003***	0.027***	0.001	0.014
Positive_DA	0.071***	0.002***	0.016***	-0.002	0.010
Negative_DA	-0.074***	-0.004***	-0.050***	-0.015***	0.025
R_CFO	-0.052***	-0.001***	0.009**	0.008*	0.001
R_PROD	0.010***	-0.002***	-0.015***	0.008**	0.003
R_DISX	0.023***	0.006***	0.015***	-0.020***	0.006
RM_PROXY	-0.019***	0.003***	0.009***	-0.005	0.003
RM_PROXY2	-0.076***	-0.008***	-0.008	0.028***	0.003