UNSW Business School

Accounting Research Seminar Series
Term 1, 2019

The Impact of Year-to-Year EAR Modifications on Audit Effort

Greg Shailer

The Australian National University

Date: Friday March 22, 2019
Time: 3.00pm – 4.00pm
Venue: BUS 115

The Impact of Year-to-Year EAR Modifications on Audit Effort

Penny Zhang & Greg Shailer

Research School of Accounting

Australian National Centre for Audit and Assurance Research

The Australian National University

Canberra

ABSTRACT: This paper investigates whether auditors' year-to-year modifications in EARs signal differences in the underlying audit effort, proxied by changes in audit fees. To improve audit report transparency, regulators worldwide have adopted enhanced reporting standards which require auditors to report key (or critical) audit matters for an engagement in a particular year (EC 2014; FRC 2013; IAASB 2013; PCAOB 2017). We examine the changes in EARs over time from both language usage and KAM identification perspectives, and test whether the differences in year-to-year modifications are a consequence of differences in audit effort. We find that audit fee changes are negatively associated with KAMs that are "repeated" in consecutive years, and positively associated with "removed" and "added" KAMs, where the latter is be driven by "new" KAMs that are not borrowed from prior year's audit committee report, as distinct from :added" KAMS that previously appeared in the audit committee's report. Overall, the results support the argument that differences in year-over-year modifications in EARs are associated with differences in underlying audit effort.

Keywords: extended audit report; audit transparency; audit fee; textual analysis

INTRODUCTION

This paper investigates the extent to which extended audit report (EAR) disclosures are modified from year to year, and the association between the year-to-year EAR changes and audit effort. The most significant audit matters in a financial statement audit are now required as key narrative disclosures in the EARs (EU 2014; FRC 2013; IAASB 2015; PCAOB 2017).[1] According to the relevant auditing standards (e.g., ISA 701; ISA (UK and Ireland) 700), auditors are required to describe each reported significant audit matters and explain the way the particular matter was addressed in the audit. To enhance audit report transparency and provide financial statement users with useful and valuable information, EAR disclosures are expected to be entity-specific and fiscal-period-specific (FRC 2013; IAASB 2015; PCAOB 2017). However, it is up to the auditors to determine the breadth and depth of what is discussed in the reports. The objectives for EARs are unlikely to be achieved if disclosures included in the EARs are largely unchanged across years.

Although substantial year-to-year EAR modifications are not required under the auditing standards, regulators and investors have expressed concerns about the lack of changes in EARs over time, especially regarding the changes in KAM identification between years (FRC 2015, 2016b; PCAOB 2017). KAMs do change from year to year, but limited explanation

[1] The IAASB changes to the auditor's report are effective for periods ending on or after December 15, 2016, the EU regulations have to be adopted from June 17, 2016, and the revised auditing standards required by the PCAOB will become effective for periods ending on or after June 30, 2019, and December 15, 2020, for large accelerated filers and all other companies respectively. (EU 2014; IAASB 2015; PCAOB 2017).

is given by auditors, and both regulators and financial statement users have stated that they expect more to be discovered in this regard (e.g., FRC 2015; PCAOB 2017). In addition, accounting practitioners have expressed concerns about repetitive disclosures over time. Auditors have claimed that, while EARs for the same entity over years inevitably have some similarities, the overuse of standardized wording to describe similar risks and responses makes EARs become longer with more boilerplate language (e.g., PwC 2015).[2] In light of these concerns, it could be useful for policy makers, regulators, and practitioners to understand the implications of the year-to-year EAR modifications, from both the wording usage and the KAM identification perspectives.

In one of the early studies of EARs, Smith (2017) concludes that, compared with traditional binary audit reports, EARs are easier to read and can better capture client-specific audit risks; this conclusion is based on increases in readability scores and negative and uncertain tone. Moreover, the reporting change appears to influences financial statement users, as evidenced by negative associations between readability and analyst forecast dispersion (Smith 2017), and positive associations between the number of KAMs and loan contracting terms (Porumb, Karaibrahimoglu, Lobo, Hooghiemstra, and de Waard 2018) but, thus far, there

[2] The concept of material risks during our sample period (30 Sep 2013-30 Sep 2016) is "risks of material misstatement that had the greatest effect on the overall audit strategy; the allocation of resources in the audit; and directing the efforts of the engagement team" (FRC 2013, 6). For EARs issued for accounting periods ending on or after 18 June 2017, the FRC has extended the concept of material risks to incorporate the EU and the IAASB requirements, and risks are discussed under the heading "KAMs" (FRC 2016a). We use "KAMs" in this study, because the FRC states that the incorporation of international standards does not expect to result in a significant change on risk identifications, and "KAMs" and material risks are broadly equivalent (FRC 2016b).

is no supporting evidence based on equity market responses (Lennox, Schmidt, and Thompson 2018).

Gutierrez, Minutti-Meza, Tatum, and Vulcheva (2018) and Reid, Carcello, Li, and Neal (2018) find that the implementation of EARs had no effect on audit fees. However, the positive associations between audit fees and the EAR length and the total number of KAMs reported in Gutierrez et al. (2018) suggest that, after the implementation of EARs, more transparent disclosures may signal greater audit effort. We extend this literature by examining the extent to which auditors modify their EAR disclosures from year to year, and investigate whether any changes are a consequence of greater audit effort, as reflected in audit fees.

The year-to-year EAR modifications are measured from both wording usage and KAM identification perspectives. Wording usage is based on textual analysis, using the vector space model (VSM), the VSM with TF-IDF weighting function, and Trigram approaches to compute year-to-year document similarity scores for an audited entity's EARs. These measures capture the extent to which individual words or three-word phrases in EARs are repeated from one year to the next. KAM changes are grouped into three categories: removed, repeated, and added. Although KAM identification has received prior research attention, there is little established understanding of their use over time. It could be that repeating KAMs from year to year means that these risks are particularly important and that these issues attract the greatest audit effort. Conversely, they could represent a standardized approach in auditing. However, under both of the circumstances, more "repeated" KAMs imply relatively stable economic condition, and we

therefore predict no association between repeated KAMs and audit fee changes. In addition, both "removed" and "added" KAMs indicate greater larger degree of EAR modifications.

Substantial changes in year-to-year EAR disclosures may arise because of changes in the economic condition an entity faced or because of auditors seeking higher reporting quality. We argue that if auditors provide reliable disclosures regarding the client's risks of material misstatements, the degree of EAR modifications is expected to be greater for firms facing substantial economic changes. If that is the case, greater audit effort is expected as auditors cannot rely much on the prior audit approach and need to substantially revise prior year's audit plan, which thus leading to increases in audit fees. Therefore, a positive association between year-to-year EAR modifications and audit fee changes is expected.

Overall, the empirical results provide evidence that audit fees increase more when auditors made greater modifications of EAR disclosures in terms of both wording usage and KAM identifications. Audit fee changes are found to be negatively associated with EAR year-to-year similarity scores and "repeated" KAMs, and positively associated with "removed" and "added" KAMs. Further, we find that the "added" KAMs effect is driven by those "new" KAMs that are not borrowed from the prior year's EACR.

This study contributes to policy development and the audit research literature. First, it provides timely evidence with regard to the potential effect of the new extended audit reporting regime on auditors' reporting behavior. We find that disclosing less repetitive and more entity-specific information is associated with more audit effort. The results have policy implications

and may help regulators in other jurisdictions better monitor the implementation of EARs, owing to the narrower approach that the FRC has adopted. Second, by examining the relations between year-to-year EAR modifications and changes in audit fees, this study contributes to the audit reporting literature and enriches the ongoing discussion regarding the consequences of implementing the EAR regime. We add to the relevant literature by examining auditors' narrative disclosures from both KAM identification and language usage perspectives, and testing whether providing less repetitive information in the EAR is a consequence of greater audit effort. Third, the study contributes to the literature on textual analysis. Although research concerned with narrative disclosures has been growing rapidly in recent years (Loughran and McDonald 2016), very few studies apply textual analysis to audit reports. By analyzing auditors' wording usage and KAM identification in EARs, this study complements the emerging research pertaining to auditors' contextual disclosures.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Since Simunic (1980), the audit fee studies have consistently found that audit fees are positively associated with firm size, client complexity and engagement risk. Factors such as the number of subsidiaries, litigation risk, and regulatory requirements, are found to determine the amount of effort that auditors spend in an audit and thus lead to higher audit fees (e.g., Simunic 1980; Simunic and Stein 1996). According to Hay, Knechel, and Wong (2006), those factors can be grouped into client attributes, auditor attributes, and engagement attributes.

Regarding client attributes, client size, entity complexity, inherent risk, profitability, and leverage are found to affect audit fees significantly (e.g., Francis 2011; Simunic 1980). Client size is the dominant determinant of audit fees, but fees are also increased by entity complexity and various risk factors (as detailed in the meta-analysis in Hay et al. 2006 and subsequent studies). Areas that are difficult to audit and are likely to contain errors, such as inventory and accounts receivable, require specialized audit procedures and are positively related to audit fees. Profitability and leverage, which may signal inherent risks of misstatements or the extent to which auditors are consequentially exposed to risk, also increase audit fees. Overall, audit fees increase when auditors are required to make estimates that are more complex and to use greater professional judgment and decrease when auditors can plan and process the audit more efficiently (e.g., Kim, Liu, and Zheng 2012; Zhang 2018).

Prior studies of EARs find that riskier, more complex and larger entities have comparatively more KAMs discussed in their EARs (Lennox et al. 2018), and that banks incorporate this into their risk assessments (Porumb et al. 2018). These findings imply that auditors' disclosures in EARs reliably capture client firms' financial reporting risks. This argument is also partially supported by prior research of corporate disclosures. The literature concerned with managers' disclosures suggests a positive association between risk factor disclosures and firm risks (e.g., Brown and Tucker 2011; Campbell, Chen, Dhaliwal, Lu, and Steele 2014). For example, Campbell et al. (2014), using thirteen proxies to capture risks an entity faced, report that managers increase their disclosures in the "risk factor" Section of the 10-K file when firms experience greater expected returns, higher leverage, larger stock return

volatility and turnover, have a Big N auditor, have greater analyst following, experience lower profitability, lower effective tax rates, and lower stock return skewness. Moreover, Brown and Tucker (2011) find that managers modify entities' Management Discussion and Analysis (MD&A) disclosures more from the prior years' when entities are facing substantial changes in operations, liquidity, capital resources, risk exposure, and business components.

Reid et al. (2018) and Gutierrez et al. (2018) examine the introduction of EARs. Their results based on pre-post tests do not suggest that the introduction of EARs affected audit fees; this remained the case even for reports with more total KAMs, more "unmatched" KAMs (those that were not mentioned in that entity's EACR), or greater materiality thresholds. However, within the sample of EARs, Gutierrez et al. (2018) found positive associations between audit fees and both the length of EARs and the number of reported KAMs, which suggests that more disclosures may entail more audit effort.

Yang, Yu, Liu, and Wu (2018), a relevant non-EAR study, find that audit fees are positively associated with the clients' risk disclosures. Using textual analysis based on Natural Language Processing to capture entity-specific risk disclosures and the risk management standards of Institute of Risk Management (2002), Yang et al. (2018) examine client self-identified risk disclosures from financial, strategic, operational, and hazard perspectives, and find that audit fees are positively associated with each of the four types of risk disclosures and an overall risk score that is generated using the four risk types. The results of Yang et al. (2018) can be related to the results in Lennox et al. (2018), who report that KAM disclosures in the first year of EARs in the UK largely pertained to risks that were previously disclosed by the

clients and that the audit fees already reflected these factors. In further analysis for the second year of EARS, Lennox et al. found that 72% of KAMs were repeated from the previous year, but they did not evaluate how this impacted audit fees. Consequently, the prior studies do not provide any evidence regarding how *changes* in auditors' disclosures in EARs might pertain to audit effort or fees.

If a company experiences relatively stable underlying economic fundamentals and operating conditions over time, then the auditor might not need to make substantial revisions to the existing audit plan and might rely more on its prior approach to conduct the audit. This may be reflected in more consistent risk disclosures, including fewer wording changes and more "repeated" KAMs reported in the next year's EARs. If EAR disclosures reliably reflect the audit concerns, then greater changes to EARs disclosures from year to year should be related to changes in auditors' concerns. Identifying, evaluating, and addressing the risks underlying the changes in auditors' disclosures should require audit effort. The more the auditor needs to revise the previous audit plan and place less reliance on the prior audit approach, the greater the expected impact on audit effort. We expect these engagements to have greater EAR modifications, reflected in more wording changes and more "removed" and "added" KAMs in the following year's EARs. These arguments are reflected in the following hypothesis:

HYPOTHESIS. Year-to-year EAR modifications are positively associated with audit fee changes.

Testing the predicted positive association between year-to-year EAR modifications and audit fee changes might be confounded if substantial changes in EAR disclosures are used strategically by auditors. If auditors modify their disclosures in EARs as a deliberate differentiation or obfuscation strategy without expending more audit effort, we are less likely to observe significant associations between changes in audit fees and changes in disclosures. Furthermore, fewer year-to-year EARs modifications might occur when auditors invest less effort in their reporting behavior, irrespective of their underlying audit effort. If the extra work auditors conducted to address their concerns in relation to companies' substantial economic changes are not reflected in their EAR disclosures, the expected association between year-to-year EAR modifications and changes in audit fees might not be evident.

MEASUREMENT AND SAMPLE

Measurement of Year-Over-Year EAR Modifications

The changes in EARs from year-to-year are examined using textual analysis methods (applied to both the full EARs and the KAMs sections of EARs) and by examining changes in specific KAMs

Textual Analysis Methods

To examine year-to-year changes in text, we use the vector space model (VSM) method, with and without the TF-IDF weighting function, and Trigram methods, to compute the

similarity score of a particular company's EARs in two consecutive years for both KAM sections and the full report.

The VSM has been widely used in accounting research. For example, Brown and Tucker (2011) adopt the method to capture the extent to which firms modify the management discussion and analysis (MD&A) section of the 10-K filing from year to year. Peterson, Schmardebeck, and Wilks (2015) apply the VSM to evaluate the extent to which an entity's accounting policy disclosures are consistent over years.

In the VSM, the m-dimensional vector for an EAR (or the KAM section) of company *i* is represented as:

$$v_i = (w_1, w_2, \ldots, w_{m-1}, w_m) \quad (1)$$

where *w* is the frequency of (each) word in the full EAR (or KAM section) for company *i*.[3]

To address the concern that all words should be treated equally under the VSM, we follow prior studies, such as Brown and Tucker (2011) and Hoberg and Phillips (2016), and modify the calculation of similarity scores in the traditional VSM approach by weighting words

[3] m is the number of unique words after removing "stop words". Consistent with prior literature, "stop words" are common words, such as "a" "is" "the" "will" (e.g., Li 2010; Perterson et al. 2015), and words with a common stem are treated as the same word; e.g., "auditing" and "audited" are both stemmed to "audit, and "audit" "auditing" and "audited" are treated as the same word (e.g., Merkley 2014; Lang and Stice-Lawrence 2015). Some studies use complex text terms to generate similarity measures, rather than single-word identifiers, but it has been shown that the single-word basis is preferable (e.g., Salton and Buckley 1988).

using the TF-IDF approach. The TF-IDF weighting, expressed as [$w_{mi} \times \log(N/nm)$], assigns greater weight to words that are used less frequently and lesser weights to common words, with 0 weight for a word that appears in every document.

Both the VSM, with and without TF-IDF weighting, use the same model to compute the similarity score. The similarity degree between EARs, v_i and v_j, is stated as:

$$Similarity_{ij} = (v_i \times v_j) / (\|v_i\| \times \|v_j\|) \quad (2)$$

where $v_i \times v_j$ yields the scalar product of v_i and v_j, and $\|v_i\|$ and $\|v_j\|$ represent the vector lengths.[4]

The resulting variables, *SIMKAM* (*SIMEAR*) and *SIMKAM_IDF* (*SIMEAR_IDF*) are the test variables. They are the similarity scores between a company's current year KAM section (full EAR) and that for the previous year, estimated using the traditional VSM and VSM with TF-IDF weighting function, respectively.

In addition, following prior accounting research (e.g., Lang and Stice-Lawrence 2015; Nelson and Pritchard 2007), the Trigram approach is adopted to compute the year-to-year KAM and EAR modifications, based on the similarity of three-word phrases across the two reports. Each KAM section and full EAR are converted into sets of overlapping trigrams. Then

[4] The inherent normalization makes it unnecessary to further control for EAR length (see Hoberg and Phillips 2010, 2016).

the similarity score between the EARs of companies i and j is computed by dividing the intersection of the two sets (one for each company) by the union of the two sets of trigrams, as follows:

$$Similarity_{ij} = |S(i) \cap S(j)| \,/\, |S(i) \cup S(j)| \tag{3}$$

where $S(i)$ and $S(j)$ represent the sets of trigrams for KAM(i) and KAM(j) or EAR(i) and EAR(j), respectively. $Similarity_{ij}$ is bounded within (0, 1), and the score approaches 1 as the similarity of two documents for companies i and j increases.

The resulting variables are therefore denoted as *SIMKAM_Tri* and *SIMEAR_Tri*.

KAM Categories

We examine reported KAMs according to whether specific KAMs are repeated, removed, or added in one year compared to the preceding year. "Removed" KAMs are those that were included in the prior year's EAR but are not included in the current year's EAR. "Repeated" KAMs are those that are included in both the current and prior years' EARs. "Added" KAMs are those that are included in the current year's EAR were not included in the prior year's EAR. The number of "removed," "repeated," and "added" KAMs are scaled by the total KAMs in current year's EAR to yield the test variables *%REMOVE*, *%REPEAT*, and *%ADDED*.

Model for Testing the Hypothesis

To tests our hypothesis, we variously estimate the following model for our different measures of changes in EARs:

$$\Delta AFEE = \beta_0 + \beta_1 \Delta EAR\_CHG + \beta_2 \Delta SIZE + \beta_3 \Delta MTB + \beta_4 \Delta CFO + \beta_5 \Delta INV + \beta_6 \Delta REC + \beta_7 \Delta SALESVOL + \beta_8 LOSS + \beta_9 \Delta LEV + \beta_{10} BUSY + \beta_{11} \Delta ROA + \beta_{12} AFEE_{t-1} + \Sigma\beta_j industry + \Sigma\beta_k year \quad (4)$$

where *ΔAFEE* is the change in log of audit fees from *t*-1 to *t*; and *ΔEAR_CHG* is the variable of interest, and represents each of *SIMKAM*, *SIMEAR, SIMKAM_IDF*, *SIMEAR_IDF*, *SIMKAM_Tri*, *SIMEAR_Tri*, *%REMOVE, %REPEAT,* and *%ADDED* in different of the model.

We expect the coefficient of Δ*EAR_CHG* to be negative when it represents *SIMKAM*, *SIMEAR, SIMKAM_IDF*, *SIMEAR_IDF*, *SIMKAM_Tri*, *SIMEAR_Tri*, or *%REPEAT*, and positive when it represents %*REMOVE* or *%ADDED*.

Consistent with prior studies, we control for the effects on audit fees of firm size (*SIZE*), market-to-book ratio (*MTB*), cash flow from operations (*CFO*), inventory (*INV*), receivables (*REC*), sales volatility (*SALESVOL*), loss (*LOSS*), leverage (*LEV*), busy season (*BUSY*), and

return on assets (*ROA*).[5] Except for *LOSS* and *BUSY*, the controls are calculated as changes from year *t*-1 to *t*. We also control for the previous year's level of audit fees, $AFEE_{t-1}$, and industry and year fixed effects. For convenience, all variable definitions are tabulated in Appendix 1.

Sample Selection

The sample is selected from all UK non-financial entities that reported on the application of the UK Corporate Governance Code for the fiscal years between 30 September 2013 and 30 September 2016.[6] We start with all available annual reports of relevant companies, yielding 1,244 firm-year observations of changes in EARs. We then manually collect the observations' audit fee data and lose 14 observations during this process. The data are then merged with DataStream to obtain the control variables and we lose another 391 observations because of the missing data. The final sample is comprised of 839 firm-year observations, as shown in Table 1.[7]

[INSERT TABLE 1 HERE]

[5] We do not control for changes in Big4 auditors, because no company changed between a Big4 auditor and non-Big4 auditor during the study period.
[6] Two early adopters of EARs are included in the sample and the results remain unchanged after removing them from the analyses.
[7] The final sample size is reasonable, compared to other current EAR studies in the UK. For example, Gutierrez et al. (2018) and Reid et al. (2018) report 872 and 1,088 firm-year observations, respectively, for four-year period pre-post analyses.

EMPIRICAL RESULTS

Descriptive Statistics

Descriptive statistics for the main variables are presented in Table 2.[8] The average change of audit fees in the natural log format is −0.043 and it is expected that the mean and median of *ΔAFEE* would be centered around zero, as audit fees are relatively stable over time. The mean of lagged $AFEE_{t-1}$ is 12.707, consistent with 13.084 in Gutierrez et al. (2018) and 13.485 Reid et al. (2018). The mean values of KAM-related scores (*SIMKAM, SIMKAM_IDF,* and *SIMKAM_Tri*) are consistently lower than EAR-related scores (*SIMEAR, SIMEAR_IDF,* and *SIMEAR_Tri*). This is expected because disclosures in the KAM section are expected to be the section of greatest modifications. Similarity scores computed using the VSM (without TF-IDF) measure (*SIMKAM* and *SIMEAR*) are similar to those using the VSM with TF-IDF weighted measure (*SIMKAM_IDF* and *SIMEAR_IDF*) in terms of scale and variability; while the scores using the Trigram measure (*SIMKAM_Tri* and *SIMEAR_Tri*) exhibit much more variability. In the sample, around 18.2 per cent of KAMs reported in year *t*-1 are removed in year *t* (*%REMOVE*). About 84.2 per cent KAMs appear in EARs for two consecutive years (*%REPEAT*), while only 7.0 per cent of year *t* KAMs are newly added (*%ADDED*).

[8] To mitigate the effect of outliers, all continuous variables are winsorized at 0.01 (Francis and Yu 2009). As a robustness check, we estimate the models without winsorizing and do not observe any significant differences in the regression results.

[INSERT TABLE 2 HERE]

The control variables do not exhibit any unusual characteristics. The mean change in *SIZE* is positive and the mean change in *MTB* is negative. Changes in *CFO*, *INV*, *REC*, *SALESVOL*, *LEV*, and *ROA* have means close to zero. Approximately 18 per cent of cases report a loss and 43 per cent of reports are issued in the busy period.

Pearson Correlations

Pearson correlation coefficients between all regression variables are reported in Table 3. There are high correlations between each VSM measure (either with or without the TF-IDF weighting function) and the corresponding Trigram measure. For example, the correlation coefficient of *SIM_KAM* and *SIM_KAM_IDF* is 0.97, suggesting we are likely to observe similar results for the different measures. High correlations among *%REMOVE*, *%REPEAT*, and *%ADDED* are expected because, all else being equal, more KAMs should be repeated when fewer prior year's KAMs are dropped (-0.91) and fewer new KAMs are added (−0.67) to the current year's EAR. Similarly, when more prior year's KAMs are dropped, more new KAMs are likely to be added to this year's EARs (0.76). Regarding the correlations among the control variables, no particular concerns are raised by the pairwise correlation. To assess multicollinearity concerns for each regression, VIFs are computed. The largest VIF is 2.03, far below the orthodox benchmark level of 10, which do not suggest any multicollinearity problems (Neter, Kutner, Nachtsheim, and Wasserman 1996).

[INSERT TABLE 3 HERE]

Regression Results

The regression results examining annual changes in auditors' wording usage and KAM identifications are reported in Panels A and B of Table 4, respectively. We estimate Model 4 with each of *SIMKAM, SIMEAR, SIMKAM_IDF, SIMEAR_IDF, SIMKAM_Tri,* and *SIMEAR_Tri* as the test variables, as reported in Panel A of Table 4.[9] Each measure of textual similarity is significantly negatively related to audit fee changes, supporting the hypothesis. The results suggest that greater changes in auditors' wording usage from the prior year, implying more entity-specific and fiscal-period-specific information, signal higher audit effort. It is worth noting that similarity scores captured over full EARs have stronger effects on audit fees, compared with those captured in KAM sections only. This could be because the UK auditing standards require auditors to disclose additional information on not only KAMs, but also on materiality and audit scope (FRC 2013).[10] Sections other than KAMs have significant variations over time as well and receive additional audit effort.

[INSERT TABLE 4 HERE]

[9] The results indicate that most control variables are non-significant in relation to the changes in audit fees. To assess the relevance of these variables to audit fees for this sample, we estimate the basic audit fee model ($AFEE = \beta_0 + \beta_1 SIZE + \beta_2 MTB + \beta_3 CFO + \beta_4 INV + \beta_5 REC + \beta_6 SALESVOL + \beta_7 LOSS + \beta_8 LEV + \beta_9 BUSY + \beta_{10} ROA + \Sigma\beta_j$ industry+ $\Sigma\beta_k$ year), and obtain significant coefficients on *SIZE, MTB, INV, REC, BUSY*, and *ROA*, consistent with prior studies, and adjusted R-squared of 0.83.

[10] Further investigation of materiality and audit scope effects is not practical at this stage and is left to further research.

The regression results investigating KAM changes over time are presented in Panel B of Table 4. Audit fee changes are significantly positively associated with *%REMOVE* and *%ADDED* KAMs (at p-values < 0.01), and negatively related to *%REPEAT* KAMs (at p-values < 0.05). The magnitude of increases in audit transparency, reflected as greater changes in risk assessments and accounting judgments, is associated with additional audit effort, supporting the hypothesis. It appears that, on average, additional audit effort is incurred when either adding or dropping KAMs from one year to the next.

The evidence supports the hypothesis that audit fees increase more when auditors make greater modifications to EAR disclosures (in terms of both wording usage and reported KAMs). The results suggest that the inclusion of entity-specific and fiscal-period-specific information could be a consequence of extra audit effort, or that the issues to which the changes pertain require more audit effort.

ROBUSTNESS TESTS

Analyses Using Cross-section Similarity Measures

In the main analysis, we capture the year-to-year EAR modifications by comparing auditors' wording between years. We repeat the analysis replaing the between-years similarity measures with mean cross-sectional measures of similarity, estimated between clients' EARs within an auditor's portfolio. Because disclosures that are provided in most EARs are likely to be generic and so convey less firm-specific or fiscal-period-specific information, we investigate the relationship between the changes in the EAR cross-section similarity scores and

audit fees to test the extent to which standardization across the client's EARs is a consequence of deficiencies in the audit effort. We calculate $Similarity_{ij}$ for each pair of companies (*i* and *j)* in an audit firm's portfolio in a given year, for both the KAM section and the full EAR, using equation 2. The cross-sectional *BaseSIMKAM* (or *BaseSIMEAR*) score for company *i* is then computed as the average of its relevant *Similarity* scores in relation to all other (*n*-1) companies in the same audit firm portfolio (of *n* companies) in the relevant year. For example, the KAM similarity score for company *i* is expressed as:

$$BaseSIMKAM_{it} = \Sigma_{j} KAM\ Similarity_{ij} / (n\text{-}1) \quad (5)$$

where *n* is the number of companies in the relevant audit firm's portfolio in year *t.* The same method is used to obtain $BaseSIMEAR_{it}$. Same as the main analyses, we also adopt both the VSM with TF-IDF weighting function and Trigram methods to compute the document similarity, yielding *BaseSIMKAM_IDF, BaseSIMEAR_IDF*, *BaseSIMKAM_Tri* and *BaseSIMEAR_Tri.* The annual change of *BaseSIMKAM*, *BaseSIMEAR, BaseSIMKAM_IDF, BaseSIMEAR_IDF*, *BaseSIMKAM_Tri* and *BaseSIMEAR_Tri* are the test variables in the test, and positive coefficients are expected on these variables.

The underlying assumption of this measure is that standardized disclosures, which are used by auditors across clients, fail to convey entity-specific information because they make "unlike things look alike". However, if the standardization across reports makes "like things look alike", similarities are expected to enhance comparability (Financial Accounting

Standards Board 2010; Lang and Stice-Lawrence 2015).[11] To address this concern, we then modify *BaseSIMKAM* and *BaseSIMEAR* to take account of expected standardized wording (where "like things *should* look alike") based on industry comparability effects. The adjustment applied to each company's *BaseSIMKAM* or *BaseSIMEAR* is calculated using the same procedure as described above, except that the defined set of documents is the disclosures for all companies in the same industry that are audited by the same audit firm in the relevant year. Thus, company *j* must be in the same industry as company *i*.[12] Based on equation 5, the adjustment for industry-based comparability effects for the KAMs section of company *i* is:

$$IndSIMKAM_{it} = \sum_{j} KAM\ Similarity_{ij} / (m\text{-}1) \qquad (6)$$

where *m* is total number of companies in the industry of company *i* in the relevant audit firm's portfolio in year *t*. Same method is used to obtain *IndSIMEAR*.

We then use the following two methods to adjust the similarity scores for the expected industry effects from equation 6. First, to obtain the unexpected similarity score, we regress the base similarity scores (from equation 5) against the industry-based scores for each year (from equation 6). For the KAM scores, this can be expressed as:

---

[11] Comparability is defined by Financial Accounting Standards Board (FASB) as a qualitative characteristic of disclosure to enable users to identify and understand similarities in and differences among items. FASB (2010) states that "For information to be comparable, like things must look alike and different things must look different. Comparability of financial information is not enhanced by making unlike things look alike any more than it is enhanced by making like things look different" (FASB 2010, 19-20).

[12] This is similar to the grouping method used in Brown and Knechel (2016) to compare the similarity of the narrative content in 10-K filings.

$$BaseSIMKAM_{it} = b_1 IndSIMKAM_{it} + \varepsilon_{it} \quad (7)$$

and save the residual ε_{it} as the unexpected similarity score for the KAMs section for company i in year t, SIM_KAMRES_{it}. The same method is used to obtain unexpected similarity scores for the full EARs, *SIM_EARRES*. In addition, we also obtain the adjusted similarity score deducting the industry-based score *IndSIMKAM* from the base similarity scores *BaseSIMKAM* to obtain *SIM_KAMDIF*, which thus can be expressed as:

$$SIM\_KAMDIF_{it} = BaseSIMKAM_{it} - IndSIMKAM_{it} \quad (8)$$

The same method is used to obtain adjusted similarity scores for the full EARs, *SIM_EARDIF*.

Consistently, both the VSM with TF-IDF weighting function and Trigram methods are adopted, and following the same procedures as discussed above, we yield *SIM_KAMRES_IDF, SIM_KAMRES_Tri, SIM_EARRES_IDF* and *SIM_EARRES_Tri* when using the residual method and the *SIM_KAMDIF_IDF, SIM_KAMDIF_Tri, SIM_EARDIF_IDF* and *SIM_EARDIF_Tri* when using the differencing method. The annual change of *SIM_KAMRES_IDF, SIM_KAMRES_Tri, SIM_EARRES_IDF*, *SIM_EARRES_Tri, SIM_KAMDIF_IDF, SIM_KAMDIF_Tri, SIM_EARDIF_IDF* and *SIM_EARDIF_Tri* are therefore the test variables in the test, with positive coefficients expected.

The regression results using base similarity scores, residual adjusted scores and differencing adjusted similarity scores are reported in Panels A, B, and C of Table 5,

respectively. None of the coefficients for these variables are significant. The non-significant coefficients may be resulted from the lack of variation among cross-section measures from year to year. For example, the mean value and the standard deviation of *ΔSIM_KAMRES* are 0.000 and 0.035, respectively. Nonetheless, the results do not suggest that annual changes in standardization across EARs are related to changes in audit effort.

[INSERT TABLE 5 HERE]

FURTHER ANALYSES

Analysis on Added KAMs

The main analysis reveals that audit fees increased more from the previous year's fees when more KAMs are added in the current year's EAR. Among the KAMs that were added in current year's EARs, some of them are prior year's "unmatched" KAMs (that is, they were reported in that entity's EACR, but not in the EAR). If newly disclosed KAMs are "borrowed" from the AC's prior disclosures, then the audit production cost and the amount of extra effort the auditors needed to spend is expected to be lower. Therefore, we investigate whether the effect of "added" KAMs is mainly driven by the "newly added" KAMs.

Particular attention is paid to the communications between the ACs and the auditors. This is because, although the approaches to determine what KAMs auditors should disclose differ in detail across jurisdictions, expecting disclosed KAMs to be selected from matters that were communicated with ACs is consistent, and early and effective engagements between

external auditors and ACs are critical in this new reporting regime (Deloitte 2016; IAASB 2015; PCAOB 2017). Moreover, because the UK requires both auditors and ACs to report material matters in their respective extended reports, we are able to investigate the "borrowed" and "newly added" KAMs separately.

We re-estimate Model 4 using "*%BORROW*" and "*%ADD_NEW*" as the test variables. The results are reported in Table 6. Significant positive coefficients are obtained on "*%ADD_NEW*" (at p-values < 0.01), but not "*%BORROW*." The results imply that the "new" KAMs added in the current period's EARs receive more audit effort than those prior "ignored" matters that were disclosed in AC reports (as reflected in greater audit fee increases).

[INSERT TABLE 6 HERE]

CONCLUSION

Regulators have responded to concerns over the limited information content in the traditional auditor's report by developing new audit reporting standards that require disclosures of the most significant audit matters in that year's financial statement audit (EU 2014; FRC 2013; IAASB 2015; PCAOB 2017). EARs are intended to increase the transparency of the underlying audit work (FRC 2016a). However, we observe a seemingly low level of changes in EARs for our sample period. The generalization is that EARs may be less informative if it looks very similar to that of the previous year.

This study focuses on auditors' year-to-year modification in EARs during the UK's first three years of EARs implementation and examines whether the year-to-year variations in auditors' disclosures are associated with changes in the underlying audit effort. Changes in EAR disclosures are captured from both wording usage and KAM identification perspectives.

The results support the proposition that audit effort increases more when greater modifications to EAR disclosures are made, both in terms of wording usage and reported KAM, and for both adding new KAMs or dropping previous KAMs. Year-to-year similarity scores are negatively related to audit fee changes and this result is robust to measures using the VSM (with and without the TF-IDF weighting function) and Trigram methods. Audit fee changes are negatively associated with "repeated" KAMs, and positively associated with "removed" and "added" KAMs. Further, we find that "added" KAMs effect is driven by those "new" KAMs identified by the auditor that are not "borrowed" from the prior year's AC report.

REFERENCES

Brown, S. V., and W. R. Knechel. 2016. Auditor–client compatibility and audit firm selection. *Journal of Accounting Research* 54(3): 725–775.

Brown, S. V., and J. W. Tucker. 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research* 49: 309–346.

Campbell, J. L., H. Chen, D. S. Dhaliwal, H. M. Lu, and L. B. Steele. 2014. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19(1), 396–455.

Choi, J. H., J. B. Kim, X. Liu, and D. A. Simunic. 2008. Audit pricing, legal liability regimes, and Big 4 premiums: Theory and cross-country evidence. *Contemporary Accounting Research,* 25(1), 55–99.

Deloitte. 2016. *Clear, Transparent Reporting. The New Auditor's Report.* Available at: https://www2.deloitte.com/content/dam/Deloitte/ch/Documents/audit/ch-en-audit-new-auditors-report.pdf

European Union (EU). 2014. *Regulation (EU) No. 537/2014 of the European Parliament and of the Council of 16 April 2014 on Specific Requirements Regarding Statutory Audit of Public-interest Entities and Repealing Commission Decision 2005/909/EC*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0537&from=en

Financial Accounting Standards Board (FASB). 2010. *Statement of Financial Accounting Concepts No. 8 September 2010.* Available at: http://www.fasb.org/resources/ccurl/515/412/Concepts%20 Statement%20No%208.pdf

Financial Reporting Council (FRC). 2013. *International Standard on Auditing (UK and Ireland) 700: The Independent Auditor's Report on Financial Statements.* Available at: https://www.frc.org.uk/Our-Work/Publications/Audit-and-Assurance-Team/ISA-700-(UK-and-Ireland)-700-(Revised)-File.pdf

Financial Reporting Council (FRC). 2015. *Extended Auditors' Reports: A Review of Experience in the First Year*. Available at: https://frc.org.uk/Extended-auditors-reports.pdf

Financial Reporting Council (FRC). 2016a. *Final Draft, International Standards on Auditing (UK and Ireland) 701, Communicating Key Audit Matters in the Independent Auditor's Report.* Available at: https://www.frc.org.uk/Our-Work/Publications/Audit-and-Assurance-Team/Final-Draft-ISA-(UK-and-Ireland)-701.pdf

Financial Reporting Council (FRC). 2016b. *Extended Auditor's Reports: A Further Review of Experience.* Available at: https://www.frc.org.uk/getattachment/ 76641d68-c739-45ac-a251-cabbfd2397e0/Report-on-the-Second-Year-Experience-of-Extended-Auditors-Reports-Jan-2016.pdf

Francis, J. R. 2011. A framework for understanding and researching audit quality. *Auditing: A Journal of Practice & Theory* 30: 125–152.

Francis, J. R., and M. D. Yu. 2009. Big 4 office size and audit quality. *The Accounting Review* 84(5): 1521–1552.

Gutierrez, E., M. Minutti-Meza, K. W. Tatum, and M. Vulcheva. 2018. Consequences of adopting an expanded auditor's report in the United Kingdom. *Review of Accounting Studies* 23: 1543–1587.

Hay, D. C., W. R. Knechel, and N. Wong. 2006. Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research* 23(1): 141-191.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23: 3773–3811.

Hoberg, G., and G. Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124(5): 1423–1465.

International Auditing and Assurance Standards Board (IAASB). 2015. *International Standard on Auditing 701. Communicating Key Audit Matters in the Independent Auditor's Report.* Available at: https://www.iaasb.org/ system/files/publications/files/ISA-701_2.pdf

Ireland, J. C., and Lennox, C. S. 2002. The large audit firm fee premium: A case of selectivity bias? J*ournal of Accounting, Auditing & Finance*, 17(1), 73–91.

Kim, J. B., X. Liu, and L. Zheng. 2012. The impact of mandatory IFRS adoption on audit fees: Theory and evidence. *The Accounting Review* 87(6): 2061–2094.

Lang, M., and L. Stice-Lawrence. 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60: 110–135.

Lennox, C. S., J. J. Schmidt, and A. M. Thompson. 2018. *Is the Expanded Model of Audit Reporting Informative to Investors? Evidence from the UK.* Working paper, University of Southern California, University of Texas at Austin, and University of Illinois at Urbana-Champaign.

Li, F. 2010. The information content of forward-looking statements in corporate filings: A naïve Bayesian machine learning approach. *Journal of Accounting* Research 48(5): 1049–1102.

Merkley, K. J. 2014. Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review* 89(2): 725-–757.

Nelson, K. K., and A. C. Pritchard. 2007. *Litigation Risk and Voluntary Disclosure: The Use of Meaningful Cautionary Language.* Working paper, Texas Christian University and University of Michigan.

Neter, J., M. Kutner, C. Nachtsheim, and W. Wasserman. 1996. *Applied Linear Statistical Models*. 4th ed. New York, NY: McGraw-Hill.

Peterson, K., R. Schmardebeck, and T. J. Wilks. 2015. The earnings quality and information processing effects of accounting consistency. *The Accounting Review* 90(6): 2483–2514.

Porumb, V. A., Y. Z. Karaibrahimoglu, G. J. Lobo, R. Hooghiemstra, and D. de Waard. 2018. *Is More Always Better? Disclosures in the Expanded Audit Report and their Impact on Loan Contracting.* Working paper, University of Groningen, and University of Houston.

PricewaterhouseCoopers (PwC). 2015. *Delivering the Value of the Audit: New Insightful Audit Reports.* Available at: https://www.pwc.com/gx/en/audit-services/publications/assets/pwc-auditing-report-new-insightful.pdf

Public Company Accounting Oversight Board (PCAOB). 2017. *The Auditor's Report on an Audit of Financial Statements when the Auditor Expresses an Unqualified Opinion and Related Amendments to PCAOB Standards.* Available at: https://pcaobus.org/Rulemaking/Docket034/2017-001-auditors-report-final-rule.pdf

Reid, L. C., J. V. Carcello, C. Li, and T. L. Neal. 2018. Impact of auditor and audit committee report changes on audit quality and costs: Evidence from the United Kingdom. *Contemporary Accounting Research* (forthcoming).

Salton, G., and C. Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing and Management* 24: 513–523.

Simunic, D. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research* 18 (1): 161–90.

Simunic, D., and M. T. Stein. 1996. Impact of litigation risk on audit pricing: A review of the economics and the evidence. *Auditing: A Journal of Practice & Theory* 15 (Supplement): 119–134.

Smith, K. W. 2017. *Tell me More: A Content Analysis of Expanded Auditor Reporting in the United Kingdom.* Working paper, Virginia Polytechnic Institute and State University.

Yang, R., Y. Yu, M. Liu, and K. Wu. 2018. Corporate risk disclosure and audit fee: A text mining approach. *European Accounting Review* 27(3): 583–594.

Zhang, J. H. 2018. Accounting comparability, audit effort, and audit outcomes. *Contemporary Accounting Research* 35(1): 245–276.

APPENDIX A – Variable Definitions

%ADD_NEW	=	The number of new KAMs that auditors disclose in current year's EARs, after removing those included in prior year's AC report, scaled by the current year's total KAMs.
%ADDED	=	The number of new KAMs that auditors disclose in the current year's EARs, scaled by the current year's total KAMs.
%BORROW	=	The number of new KAMs in the current year's EAR that are the same as those disclosed in that entity's prior year's AC report, scaled by the current year's total KAMs.
%REMOVE	=	The number of removed KAMs from the prior year's EARs, scaled by the prior year's total KAMs.
%REPEAT	=	The number of KAMs included in both the prior year's and current year's EARs, scaled by the current year's total KAMs.
△REC	=	The change in accounts receivables from year t-1 to year t, scaled by total assets at year t.
AFEE	=	The natural log of total audit fees.
BaseSIMEAR	=	The mean similarity score of a firm's EAR disclosure to all of the other firms' disclosures, audited by the same audit firm in year t, using the VSM.
BaseSIMEAR_IDF	=	The mean similarity score of a firm's EAR disclosure to all of the other firms' disclosures, audited by the same audit firm in year t, using the VSM with TF-IDF weighting function.
BaseSIMEAR_Tri	=	The mean similarity score of a firm's EAR disclosure to all of the other firms' disclosures, audited by the same audit firm in year t, using the Trigram method.
BaseSIMKAM	=	The mean similarity score of a firm's KAM disclosure to all of the other firms' disclosures, audited by the same audit firm in year t, using the VSM.
BaseSIMKAM_IDF	=	The mean similarity score of a firm's KAM disclosure to all of the other firms' disclosures, audited by the same audit firm in year t, using the VSM with TF-IDF weighting function.
BaseSIMKAM_Tri	=	The mean similarity score of a firm's KAM disclosure to all of the other firms' disclosures, audited by the same audit firm in year t, using the Trigram method.
BUSY	=	1 if the company's fiscal year end is during the month of December, otherwise 0.
CFO	=	The cash flow from operations divided by total assets at the end of year t.
INV	=	Inventory scaled by total assets in year t.
LEVERAGE	=	The ratio of debt to total assets at the end of year t.
LOSS	=	1 if the company reports a negative net income, otherwise 0.
MTB	=	The market value of equity divided by the book value of equity at the end of year t.
REC	=	Accounts receivables scaled by total assets in year t.
ROA	=	The net income before extraordinary items divided by total assets at the end of year t.

SALESVOL	=	The standard deviation of annual sales over the prior three years.
SIM_EARDIF	=	The difference between industry-based similarity scores and *BaseSIMEAR*, with the similarity scores computed using the VSM.
SIM_EARDIF_IDF	=	The difference between industry-based similarity scores and *BaseSIMEAR*, with the similarity scores computed using the VSM with TF-IDF weighting function.
SIM_EARDIF_Tri	=	The difference between industry-based similarity scores and *BaseSIMEAR*, with the similarity scores computed using the Trigram method.
SIM_EARRES	=	The regression residual of industry-based similarity scores and *BaseSIMEAR*, with the similarity scores computed using the VSM.
SIM_EARRES_IDF	=	The regression residual of industry-based similarity scores and *BaseSIMEAR*, with the similarity scores computed using the VSM with TF-IDF weighting function.
SIM_EARRES_Tri	=	The regression residual of industry-based similarity scores and *BaseSIMEAR*, with the similarity scores computed using the Trigram method.
SIM_KAMDIF	=	The difference between industry-based scores and *BaseSIMKAM*, with the similarity scores computed using the VSM.
SIM_KAMDIF_IDF	=	The difference between industry-based scores and *BaseSIMKAM*, with the similarity scores computed using the VSM with TF-IDF weighting function.
SIM_KAMDIF_Tri	=	The difference between industry-based scores and *BaseSIMKAM*, with the similarity scores computed using the Trigram method.
SIM_KAMRES	=	The regression residual of industry-based scores and *BaseSIMKAM*, with the similarity scores computed using the VSM.
SIM_KAMRES_IDF	=	The regression residual of industry-based scores and *BaseSIMKAM*, with the similarity scores computed using the VSM with TF-IDF weighting function.
SIM_KAMRES_Tri	=	The regression residual of industry-based scores and *BaseSIMKAM*, with the similarity scores computed using the Trigram method.
SIMEAR	=	The similarity score of a firm's current year's EAR disclosure to that of its prior year's disclosures, using the VSM.
SIMEAR_IDF	=	The similarity score of a firm's current year's EAR disclosure to that of its prior year's disclosures, using the VSM with TF-IDF weighting function.
SIMEAR_IDF	=	The similarity score of a firm's current year's EAR disclosure to that of its prior year's disclosures, using the Trigram method.
SIMKAM	=	The similarity score of a firm's current year's KAM disclosure to that of its prior year's disclosures, using the VSM;
SIMKAM_IDF	=	The similarity score of a firm's current year's KAM disclosure to that of its prior year's disclosures, using the VSM with TF-IDF weighting function.
SIMKAM_Tri	=	The similarity score of a firm's current year's KAM disclosure to that of its prior year's disclosures, using the Trigram method.
SIZE	=	The natural log of total assets (in £,000) at the end of year t.

TABLE 1
Sample Selection

	Firm–Year Observations
Total available EARs to calculate changes for fiscal years ending between 30 September 2014 and 30 September 2016	1,244
Less: Observations missing *AFEE*	(14)
Less: Observations missing control variables in DataStream	(391)
Final Sample	839

TABLE 2
Descriptive Statistics

Variable	N	Mean	SD	p25	p50	p75
ΔAFEE	839	-0.043	0.650	-0.007	0.000	0.083
SIMKAM	839	0.838	0.116	0.761	0.870	0.930
SIMEAR	839	0.955	0.029	0.941	0.962	0.976
SIMKAM_IDF	839	0.822	0.124	0.746	0.852	0.921
SIMEAR_IDF	839	0.950	0.033	0.937	0.957	0.972
SIMKAM_Tri	839	0.391	0.239	0.176	0.369	0.570
SIMEAR_Tri	839	0.546	0.160	0.428	0.536	0.661
%REMOVE	839	0.182	0.280	0.000	0.000	0.333
%REPEAT	839	0.842	0.223	0.667	1.000	1.000
%ADDED	839	0.070	0.451	0.000	0.000	0.333
ΔSIZE	839	0.060	0.182	-0.017	0.048	0.122
ΔMTB	839	-0.140	1.980	-0.236	-0.009	0.128
ΔCFO	839	0.000	0.049	-0.016	0.000	0.014
ΔINV	839	-0.001	0.015	-0.001	0.000	0.000
ΔREC	839	-0.002	0.033	-0.008	0.000	0.006
ΔSALESVOL	839	-0.004	0.043	-0.010	0.000	0.006
LOSS	839	0.184	0.387	0.000	0.000	0.000
ΔLEV	839	0.007	0.060	-0.009	0.000	0.019
BUSY	839	0.430	0.495	0.000	0.000	1.000
ΔROA	839	-0.015	0.103	-0.047	-0.002	0.024
AFEE(t-1)	839	12.707	1.899	10.758	12.899	13.998

Variable definitions are provided in Appendix A.

TABLE 3

Pearson Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		
1 *ΔAFEE*	1.00																				
2 *SIMKAM*	-0.01	1.00																			
3 *SIMEAR*	-0.05	0.65*	1.00																		
4 *SIMKAM_IDF*	-0.01	0.97*	0.64*	1.00																	
5 *SIMEAR_IDF*	-0.05	0.54*	0.93*	0.55*	1.00																
6 *SIMKAM_Tri*	-0.01	0.87*	0.69*	0.86*	0.60*	1.00															
7 *SIMEAR_Tri*	-0.05	0.70*	0.81*	0.70*	0.74*	0.87*	1.00														
8 *%REMOVE*	0.03	-0.52*	-0.45*	-0.54*	-0.38*	-0.43*	-0.41*	1.00													
9 *%REPEAT*	-0.03	0.56*	0.47*	0.58*	0.39*	0.48*	0.46*	-0.91*	1.00												
10 *%ADDED*	0.04	-0.22*	-0.25*	-0.25*	-0.22*	-0.17*	-0.19*	0.76*	-0.67*	1.00											
11 *ΔSIZE*	0.06	-0.04	-0.06	-0.04	-0.07*	-0.06	-0.05	0.04	-0.06	0.03	1.00										
12 *ΔMTB*	-0.03	-0.00	0.04	-0.00	0.04	0.00	0.02	-0.05	0.04	-0.04	-0.05	1.00									
13 *ΔCFO*	0.01	0.07	0.06	0.06	0.06	0.09*	0.08*	-0.09*	0.10*	-0.10*	-0.13*	-0.07*	1.00								
14 *ΔINV*	0.02	0.02	0.01	0.02	0.01	0.03	0.03	-0.01	0.00	-0.04	-0.13*	0.03	-0.22*	1.00							
15 *ΔREC*	-0.03	0.01	0.05	0.02	0.06	0.01	0.01	-0.04	0.02	-0.04	-0.18*	0.12*	-0.07*	0.12*	1.00						
16 *ΔSALESVOL*	0.01	-0.05	0.01	-0.03	0.03	-0.01	0.02	0.02	-0.01	0.03	0.06	0.02	-0.05	0.04	0.04	1.00					
17 *LOSS*	0.01	-0.04	-0.11*	-0.03	-0.10*	-0.04	-0.07	0.07	-0.04	0.07	-0.37*	0.05	-0.03	0.06	0.02	0.05	1.00				
18 *ΔLEV*	0.05	-0.02	-0.07*	-0.02	-0.10*	-0.04	-0.05	0.05	-0.04	0.06	0.23*	0.12*	-0.19*	-0.05	-0.02*	-0.05	0.11*	1.00			
19 *BUSY*	0.04	0.01	-0.05	-0.01	-0.08*	-0.05	-0.08*	0.04	-0.01	0.04	-0.03	0.09*	-0.06	0.03	-0.03	-0.00	0.04	0.12*	1.00		
20 *ΔROA*	-0.01	0.01	-0.01	-0.01	-0.01	0.01	-0.01	-0.00	-0.00	-0.01	0.25*	-0.02	0.10*	-0.03	-0.02	-0.05	-0.37*	-0.14*	-0.00	1.00	
21 $AFEE_{t-1}$	-0.20*	0.01	-0.22*	-0.04	-0.29*	-0.13*	-0.26*	0.10*	-0.13*	0.08*	0.04	-0.05	-0.08*	-0.06	-0.03	-0.11*	-0.08*	0.09*	0.32*	0.09*	1.00

Variable definitions are provided in Appendix A.

TABLE 4

Panel A: Regression Results (SIMKAM, SIMEAR, SIMKAM_IDF, SIMEAR_IDF, SIMKAM_Tri, and SIMEAR_Tri)

	ΔAFEE					
	(1)	(2)	(3)	(4)	(5)	(6)
SIMKAM	-0.348*	-	-	-	-	-
	(-1.92)					
SIMEAR	-	-2.344***	-	-	-	-
		(-3.25)				
SIMKAM_IDF	-	-	-0.354**	-	-	-
			(-2.10)			
SIMEAR_IDF	-	-	-	-2.061***	-	-
				(-3.19)		
SIMKAM_Tri	-	-	-	-	-0.209**	-
					(-2.38)	
SIMEAR_Tri	-	-	-	-	-	-0.450***
						(-3.47)
ΔSIZE	0.237*	0.210	0.238*	0.205	0.225*	0.213
	(1.78)	(1.58)	(1.78)	(1.54)	(1.69)	(1.61)
ΔMTB	-0.009	-0.009	-0.010	-0.009	-0.010	-0.010
	(-0.91)	(-0.86)	(-0.92)	(-0.89)	(-0.92)	(-0.93)
ΔCFO	-0.005	-0.019	-0.009	-0.036	0.011	0.014
	(-0.01)	(-0.04)	(-0.02)	(-0.08)	(0.03)	(0.03)
ΔINV	-0.033	-0.118	-0.046	-0.164	-0.017	-0.038
	(-0.02)	(-0.08)	(-0.03)	(-0.12)	(-0.01)	(-0.03)
ΔREC	-0.192	-0.150	-0.177	-0.138	-0.210	-0.220
	(-0.31)	(-0.24)	(-0.28)	(-0.22)	(-0.34)	(-0.35)
ΔSALESVOL	0.437	0.404	0.433	0.417	0.460	0.432
	(1.35)	(1.25)	(1.33)	(1.29)	(1.42)	(1.34)
LOSS	-0.021	-0.043	-0.021	-0.042	-0.024	-0.034
	(-0.35)	(-0.71)	(-0.35)	(-0.69)	(-0.40)	(-0.57)
ΔLEV	0.186	0.166	0.186	0.152	0.193	0.199
	(0.50)	(0.45)	(0.50)	(0.41)	(0.52)	(0.54)
BUSY	0.127***	0.131***	0.127***	0.131***	0.127***	0.129***
	(2.85)	(2.94)	(2.85)	(2.93)	(2.85)	(2.91)
ΔROA	-0.343	-0.367*	-0.351	-0.353	-0.335	-0.348
	(-1.60)	(-1.71)	(-1.63)	(-1.65)	(-1.56)	(-1.63)
$AFEE_{t-1}$	-0.251***	-0.259***	-0.252***	-0.261***	-0.255***	-0.260***
	(-14.78)	(-15.14)	(-14.83)	(-15.18)	(-14.96)	(-15.21)
Constant	3.542***	5.569***	3.547***	5.328***	3.370***	3.590***
	(11.99)	(7.39)	(12.19)	(7.67)	(12.84)	(13.10)
Industry, Year FE	Included	Included	Included	Included	Included	Included
N	839	839	839	839	839	839
Adjusted R^2	0.234	0.241	0.235	0.240	0.236	0.242

TABLE 4

Panel B: Regression Results (*%REMOVE*, *%REPEAT*, and *%ADDED*)

	ΔAFEE		
	(1)	(2)	(3)
%REMOVE	0.195***	-	-
	(3.05)		
%REPEAT	-	-0.202**	-
		(-2.33)	
%ADDED	-	-	0.122***
			(2.96)
ΔSIZE	0.235	0.231	0.238
	(1.60)	(1.56)	(1.62)
ΔMTB	-0.008	-0.009	-0.008
	(-1.17)	(-1.25)	(-1.22)
ΔCFO	0.023	0.015	0.056
	(0.05)	(0.03)	(0.11)
ΔINV	-0.051	-0.101	0.079
	(-0.05)	(-0.10)	(0.07)
ΔREC	-0.159	-0.198	-0.153
	(-0.28)	(-0.35)	(-0.27)
ΔSALESVOL	0.423	0.437	0.407
	(1.46)	(1.51)	(1.40)
LOSS	-0.029	-0.023	-0.026
	(-0.47)	(-0.38)	(-0.43)
ΔLEV	0.165	0.182	0.171
	(0.48)	(0.53)	(0.49)
BUSY	0.125**	0.130**	0.125**
	(2.29)	(2.39)	(2.29)
ΔROA	-0.357	-0.345	-0.355
	(-1.32)	(-1.27)	(-1.31)
$AFEE_{t-1}$	-0.255***	-0.254***	-0.255***
	(-5.76)	(-5.73)	(-5.78)
Constant	2.978***	3.171***	3.013***
	(5.68)	(5.76)	(5.74)
Industry, Year FE	Included	Included	Included
N	839	839	839
Adjusted R^2	0.238	0.235	0.238

***, **, * indicate significance at 1 per cent, 5 per cent, and 10 per cent levels based on a two-tailed test, respectively. t-statistics are reported below the coefficients. Variable definitions are provided in Appendix A.

TABLE 5

Panel A: Regression Results Using Base Similarity Scores

	ΔAFEE					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔBaseSIMKAM	0.193	-	-	-	-	-
	(0.66)					
ΔBaseSIMEAR	-	-1.117	-	-	-	-
		(-1.77)				
ΔBaseSIMKAM_IDF	-	-	0.232	-	-	-
			(0.84)			
ΔBaseSIMEAR_IDF	-	-	-	-1.352	-	-
				(-2.38)		
ΔBaseSIMKAM_Tri	-	-	-	-	0.081	-
					(0.23)	
ΔBaseSIMEAR_Tri	-	-	-	-	-	-0.426
						(-1.31)
ΔSIZE	0.226*	0.242*	0.224*	0.241*	0.229*	0.234*
	(1.82)	(1.95)	(1.80)	(1.95)	(1.84)	(1.89)
ΔMTB	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
	(-0.72)	(-0.73)	(-0.72)	(-0.72)	(-0.75)	(-0.70)
ΔCFO	-0.177	-0.127	-0.181	-0.122	-0.180	-0.149
	(-0.42)	(-0.31)	(-0.43)	(-0.29)	(-0.43)	(-0.36)
ΔINV	-0.584	-0.526	-0.592	-0.574	-0.597	-0.536
	(-0.43)	(-0.39)	(-0.44)	(-0.43)	(-0.44)	(-0.40)
ΔREC	-0.187	-0.136	-0.179	-0.130	-0.190	-0.185
	(-0.31)	(-0.23)	(-0.30)	(-0.22)	(-0.32)	(-0.31)
ΔSALESVOL	0.356	0.403	0.357	0.414	0.356	0.378
	(1.20)	(1.35)	(1.20)	(1.39)	(1.20)	(1.27)
LOSS	-0.023	-0.025	-0.023	-0.029	-0.021	-0.023
	(-0.40)	(-0.43)	(-0.40)	(-0.51)	(-0.37)	(-0.40)
ΔLEV	0.296	0.269	0.297	0.266	0.295	0.272
	(0.83)	(0.75)	(0.83)	(0.75)	(0.83)	(0.76)
BUSY	0.121***	0.124***	0.120***	0.124***	0.123***	0.123***
	(2.85)	(2.94)	(2.84)	(2.94)	(2.90)	(2.91)
ΔROA	-0.160	-0.162	-0.157	-0.163	-0.156	-0.166
	(-0.79)	(-0.80)	(-0.78)	(-0.81)	(-0.77)	(-0.82)
$AFEE_{t-1}$	-0.255***	-0.256***	-0.256***	-0.256***	-0.255***	-0.257***
	(-16.11)	(-16.14)	(-16.12)	(-16.18)	(-16.09)	(-16.17)
Constant	3.534***	3.328***	3.532***	3.333***	3.508***	3.519***
Industry, Year FE	Included	Included	Included	Included	Included	Included
N	975	975	975	975	975	975
Adjusted R^2	0.241	0.243	0.241	0.245	0.241	0.242

***, **, * indicate significance at 1 per cent, 5 per cent, and 10 per cent levels based on a two-tailed test, respectively. t-statistics are reported below the coefficients. Variable definitions are provided in Appendix A..

Panel B: Regression Results Using Adjusted Similarity Scores (Residual Method)

	ΔAFEE					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔSIM_KAMRES	0.500	-	-	-	-	-
	(1.14)					
ΔSIM_EARRES	-	0.215	-	-	-	-
		(0.22)				
ΔSIM_KAMRES_IDF	-	-	0.362	-	-	-
			(0.90)			
ΔSIM_EARRES_IDF	-	-	-	-0.376	-	-
				(-0.45)		
ΔSIM_KAMRES_Tri	-	-	-	-	0.478	-
					(0.89)	
ΔSIM_EARRES_Tri	-	-	-	-	-	0.170
						(0.31)
ΔSIZE	0.208	0.208	0.207	0.207	0.208	0.208
	(1.63)	(1.63)	(1.62)	(1.63)	(1.64)	(1.63)
ΔMTB	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
	(-0.74)	(-0.73)	(-0.73)	(-0.73)	(-0.74)	(-0.73)
ΔCFO	-0.095	-0.115	-0.093	-0.110	-0.106	-0.111
	(-0.22)	(-0.27)	(-0.22)	(-0.26)	(-0.25)	(-0.26)
ΔINV	-1.564	-1.620	-1.568	-1.606	-1.605	-1.612
	(-0.99)	(-1.02)	(-0.99)	(-1.01)	(-1.01)	(-1.02)
ΔREC	-0.282	-0.295	-0.301	-0.280	-0.284	-0.293
	(-0.44)	(-0.46)	(-0.47)	(-0.44)	(-0.44)	(-0.46)
ΔSALESVOL	0.453	0.438	0.445	0.447	0.444	0.439
	(1.46)	(1.41)	(1.43)	(1.44)	(1.43)	(1.41)
LOSS	-0.010	-0.011	-0.010	-0.011	-0.011	-0.011
	(-0.17)	(-0.18)	(-0.18)	(-0.19)	(-0.19)	(-0.18)
ΔLEV	0.267	0.263	0.274	0.261	0.266	0.263
	(0.74)	(0.73)	(0.76)	(0.73)	(0.74)	(0.73)
BUSY	0.116***	0.119***	0.117***	0.121***	0.119***	0.119***
	(2.68)	(2.76)	(2.72)	(2.81)	(2.77)	(2.77)
ΔROA	-0.155	-0.157	-0.156	-0.156	-0.159	-0.158
	(-0.76)	(-0.77)	(-0.76)	(-0.76)	(-0.78)	(-0.77)
$AFEE_{t-1}$	-0.267***	-0.268***	-0.267***	-0.268***	-0.267***	-0.267***
	(-16.31)	(-16.35)	(-16.26)	(-16.35)	(-16.29)	(-16.32)
Constant	3.540***	3.547***	3.535***	3.554***	3.539***	3.543***
	(14.23)	(14.26)	(14.19)	(14.26)	(14.22)	(14.22)
Industry, Year FE	Included	Included	Included	Included	Included	Included
N	950	950	950	950	950	950
Adjusted R^2	0.251	0.250	0.250	0.250	0.250	0.250

***, **, * indicate significance at 1 per cent, 5 per cent, and 10 per cent levels based on a two-tailed test, respectively. t-statistics are reported below the coefficients. Variable definitions are provided in Appendix A..

Panel C: Regression Results Using Adjusted Similarity Scores (Differencing Method)

	ΔAFEE					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔSIM_KAMDIF	0.400	-	-	-	-	-
	(1.03)					
ΔSIM_EARDIF	-	0.371	-	-	-	-
		(0.39)				
ΔSIM_KAMDIF_IDF	-	-	0.269	-	-	-
			(0.77)			
ΔSIM_EARDIF_IDF	-	-	-	-0.264	-	-
				(-0.32)		
ΔSIM_KAMDIF_Tri	-	-	-	-	0.362	-
					(0.86)	
ΔSIM_EARDIF_Tri	-	-	-	-	-	0.435
						(0.87)
ΔSIZE	0.210	0.208	0.209	0.207	0.210*	0.209
	(1.65)	(1.64)	(1.64)	(1.63)	(1.65)	(1.64)
ΔMTB	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
	(-0.75)	(-0.73)	(-0.74)	(-0.73)	(-0.73)	(-0.72)
ΔCFO	-0.097	-0.118	-0.094	-0.110	-0.103	-0.107
	(-0.23)	(-0.28)	(-0.22)	(-0.26)	(-0.24)	(-0.25)
ΔINV	-1.567	-1.629	-1.569	-1.607	-1.592	-1.618
	(-0.99)	(-1.03)	(-0.99)	(-1.01)	(-1.01)	(-1.02)
ΔREC	-0.287	-0.296	-0.305	-0.284	-0.288	-0.291
	(-0.45)	(-0.46)	(-0.48)	(-0.44)	(-0.45)	(-0.46)
ΔSALESVOL	0.452	0.437	0.445	0.445	0.446	0.439
	(1.46)	(1.41)	(1.43)	(1.43)	(1.44)	(1.41)
LOSS	-0.009	-0.011	-0.010	-0.011	-0.010	-0.011
	(-0.15)	(-0.18)	(-0.17)	(-0.19)	(-0.17)	(-0.19)
ΔLEV	0.265	0.263	0.272	0.262	0.263	0.262
	(0.74)	(0.73)	(0.75)	(0.73)	(0.73)	(0.73)
BUSY	0.117***	0.118***	0.118***	0.121***	0.119***	0.118***
	(2.70)	(2.74)	(2.74)	(2.80)	(2.76)	(2.75)
ΔROA	-0.154	-0.158	-0.156	-0.156	-0.161	-0.162
	(-0.75)	(-0.77)	(-0.76)	(-0.76)	(-0.78)	(-0.79)
$AFEE_{t-1}$	-0.267***	-0.267***	-0.267***	-0.268***	-0.267***	-0.267***
	(-16.29)	(-16.34)	(-16.24)	(-16.34)	(-16.32)	(-16.31)
Constant	3.541***	3.551***	3.537***	3.550***	3.535***	3.537***
	(14.23)	(14.26)	(14.20)	(14.26)	(14.19)	(14.21)
Industry, Year FE	Included	Included	Included	Included	Included	Included
N	950	950	950	950	950	950
Adjusted R2	0.251	0.250	0.250	0.250	0.250	0.250

***, **, * indicate significance at 1 per cent, 5 per cent, and 10 per cent levels based on a two-tailed test, respectively. t-statistics are reported below the coefficients. Variable definitions are provided in Appendix A..

TABLE 6
Regression Results for Analysis on Added KAMs

	ΔAFEE	
	(1)	(2)
%BORROW	0.122	-
	(0.72)	
%ADD_NEW	-	0.124***
		(2.64)
ΔSIZE	0.240*	0.240*
	(1.80)	(1.81)
ΔMTB	-0.010	-0.009
	(-0.93)	(-0.85)
ΔCFO	-0.049	0.058
	(-0.11)	(0.13)
ΔINV	-0.141	0.086
	(-0.10)	(0.06)
ΔREC	-0.247	-0.103
	(-0.39)	(-0.16)
ΔSALESVOL	0.454	0.402
	(1.40)	(1.24)
LOSS	-0.016	-0.024
	(-0.26)	(-0.40)
ΔLEVERAGE	0.182	0.180
	(0.49)	(0.48)
BUSY	0.130***	0.124***
	(2.90)	(2.77)
ΔROA	-0.340	-0.353
	(-1.58)	(-1.64)
$AFEE_{t-1}$	-0.251***	-0.256***
	(-14.74)	(-15.00)
Constant	3.257***	3.325***
	(12.49)	(12.80)
Industry, Year FE	Included	Included
N	839	839
Adjusted R^2	0.231	0.237

***, **, * indicate significance at 1 per cent, 5 per cent, and 10 per cent levels based on a two-tailed test, respectively. t-statistics are reported below the coefficients. Variable definitions are provided in Appendix A.