ORIGINAL RESEARCH

The context of earnings management and its ability to predict future stock returns

Nguyet T. M. Nguyen[1] · Abdullah Iqbal[2] · Radha K. Shiwakoti[3]

Abstract

This paper constructs a signal-based composite index, namely ESCORE, which captures the context of earnings management. Specifically, ESCORE aggregates 15 individual signals related to both accrual and real earnings management based on prior relevant literature. After establishing that ESCORE is capable of capturing the context in which earnings management is more likely to occur, the study finds that low ESCORE firms outperform those with high ESCORE by an average of 1.37% per month after controlling for risk loadings on the market, size, book-to-market and momentum factors up to one year after portfolio formation in the UK. This finding implies that investors tend to ignore the observable context of earnings management. In addition, with ESCORE model, investors do not need to estimate the magnitude of earnings management, rather it is sufficient to look at the surrounding context to differentiate between low and high earnings management firms. Finally, when tested using the US data, most of the main results of the study appear to hold.

✉ Abdullah Iqbal
  A.Iqbal@kent.ac.uk

  Nguyet T. M. Nguyen
  nguyet.nguyen@roehampton.ac.uk

  Radha K. Shiwakoti
  Radha.Shiwakoti@brunel.ac.uk

[1] Faculty of Business and Law, University of Roehampton, Roehampton Lane, London SW15 5SZ, UK

[2] Kent Business School, University of Kent, Chatham ME4 4TE, Kent, UK

[3] Brunel Business School, Brunel University London, Uxbridge UB8 3PH, Middlesex, UK

1 Introduction

Prior literature shows that stock markets overprice total and/or discretionary accrual component of earnings (see, for example, Sloan 1996; Teoh et al. 1998a, b; Xie 2001; Desai et al. 2004; Iqbal et al. 2009; Iqbal and Strong 2010; Wu et al. 2010; Simlai 2021). Similarly, accruals, especially the discretionary component over which managers have the discretion to manipulate, are widely used as a proxy for earnings management. Hence, the mispricing of accruals could be attributable to investors failing to fully reflect on the 'true' earnings that a manager knows but does not truthfully reveal to the market.

One can arguably question the intuition of the above story. Earnings management, of course, does not happen for no reason. Assuming managers could exercise their discretion to successfully hide the 'true' earnings and subsequently 'fool' the market, the manipulation of accounts would have happened in a context which the managers could not conceal, be it a personal motivation, a benefit to the shareholders, a pressure, or suitable opportunity for managing earnings. The market might not observe the actual earnings manipulation, but they could sense the existence of such manipulation if the general context surrounding the firm is susceptible. Since it is hard for managers to conceal the general context surrounding their firms, if investors are still 'fooled' by earnings management, they must have underreacted to the information contained in the context. This paper develops an empirical proxy for the context of earnings management and empirically tests the hypothesis that markets underreact to that context.

The main objective of this paper is to test whether the context of earnings management is mispriced. To achieve this, we begin by developing a model that captures the context in which earnings management is likely to occur. The model accumulates various signals which are extracted mainly from annual financial statements. The model generates a composite score, namely ESCORE, which accumulates 15 individual binary scores based on the rich extant literature on likely signals of earnings management. We group these signals into four broad categories: (i) the incentives for earnings management, (ii) the pressures on managers, (iii) the constraints to manage earnings, and (iv) the firm's innate characteristics. Next, using a sample of UK listed firms during the period 1995–2011, we test the effectiveness of ESCORE in capturing the context of earnings management by applying existing models of earnings management from the literature. We find that high-ESCORE firms engage in (both accruals and real) earnings management in larger magnitude and are more likely to engage in aggressive (i.e. income increasing) earnings management practices. We also find that firms which are required to restate their annual accounts by the Financial Reporting Review Panel (FRRP henceforth) have higher ESCORE in the year to which the restatement is related. Furthermore, we present evidence showing that the distribution of the first digits of figures on financial statements released by firms with higher ESCORE deviate significantly from the distribution expected by the Benford's law. According to this law, the first digits of accounting figures, like many other natural datasets, would be distributed in a way that makes smaller values have higher probability to occur (Amiram et al. 2015). Our evidence tends to suggest that higher ESCORE is associated with more anomalous distribution of the first digits of figures on financial statements, a sign suggesting the presence of earnings management. Having established that ESCORE is capable of capturing the context in which earnings management is more likely to occur, we further investigate whether ESCORE could predict one-year-ahead stock returns. The results show that a zero-investment hedge portfolio that takes long position in low ESCORE stocks and short position in high ESCORE stocks would earn an average abnormal return of 1.37%

per month after adjusting for market, size, book-to-market and momentum factors in up to one year after portfolio formation. In multivariate regressions, ESCORE is negatively and significantly related to one-year-ahead buy-and-hold returns after controlling for other existing market 'anomalies', including the mispricing of discretionary accruals. The results are robust across different portfolio weighting schemes and models to estimate abnormal returns. We also show, using US data, that the main conclusions of the paper are generalizable to the US market. Overall, the evidence strongly suggests that the context of earnings management can be used to predict future stock returns.

This paper contributes to the literature in several ways. First, the ESCORE is an alternative model to detect earnings management which accumulates the signals of earnings management. Although most existing studies in the earning management detection area are 'contextual' in nature with attempts to link one or some characteristics of the context to (often a particular type of) earnings management, none has accumulated those signals to construct an index which could capture the 'general context' which is associated with various earnings management strategies as the ESCORE does. For example, the MSCORE (Beneish 1999) and FSCORE (Dechow et al. 2011) are both constructed based on various financial statement information, but they are developed to specifically detect violations of generally accepted accounting principles (GAAP hereafter) and earnings restatements, respectively. By design, therefore, those indices are silent about earnings management strategies which are within GAAP, such as accruals and real earnings management. The ESCORE, on the other hand, could infer the probability of earnings management, both accruals and real earnings management, and even violations of financial reporting rules which result in restatements or errors. The model is, hence, particularly useful for subsequent studies which aim to detect earnings management but do not make a prediction about which methods have been used to manipulate earnings or what motives are behind such manipulation. Second and most important, we contribute new evidence to the 'market anomalies' literature showing that not only the market misprices earnings management, it also does not fully appreciate the information contained in the context surrounding such manipulation. Although previous studies document that the market under-reacts to fundamental-based composite scores which accumulate individual signals (e.g. Piotroski 2000; Mohanram 2005), our evidence clearly makes a significant incremental contribution. For example, Piotroski's FSCORE and Mohanram's MSCORE are developed for the particular settings of high and low book-to-market firms, respectively, and the selection of signals to create the composite scores is deliberately to pick up signals related to financial strengths. Therefore, while the abnormal returns earned by Piotroski's and Mohanram's models might suggest that the market under-reacts to firm's financial strength, it does not tell us how the market reacts to earnings management. Beneish et al. (2013) show that the MSCORE model, designed to detect firms which have been charged with or publicly admit to earnings manipulations, could also predict future returns. However, we develop ESCORE by selecting signals which are related to high likelihood of earnings management, regardless of whether it has been detected or not, and could be applied to any firm. Additionally, our evidence of abnormal returns earned by an ESCORE-based trading strategy suggests that the market also misprices another dimension of fundamental information which is related to the context of earnings management.

The rest of the paper is organised as follows. Section 2 reviews prior literature and develops the testable hypothesis. Section 3 presents the UK institutional settings and explains sample selection procedures. Section 4 outlines main methodologies employed in the paper. Section 5 presents and discusses the results. Section 6 provides some concluding remarks.

2 Related literature and hypothesis

The existing literature offers a handful number of earnings management detection models. The most popular method measures discretionary accruals as the deviation of actual accruals from a 'normal' level of accruals estimated using some firm-specific characteristics (Jones 1991; Dechow et al. 1995; Peasnell et al. 2000). The discretionary accruals model helps to detect one type of earnings management, namely managers exercise their discretion over accounting methods to influence reported earnings. Other researchers (Roychowdhury 2006; Athanasakou et al. 2009, 2011; Gunny 2010; Zang 2012; Sakaki et al. 2017) argue that to change reported earnings, managers do not necessarily resort to playing around with accounting methods and estimations, rather they could change real operation decisions, such as sales policies, production level, discretionary expense spending (e.g. advertising, R&D), etc. Such real earnings management has become increasingly more popular given the stricter financial reporting regulations (Cohen et al. 2008). To detect real operation management, the existing literature normally measures the deviation of the actual level of real activities with the expected level derived using some firm-specific information (Roychowdhury 2006; Srivastava 2019). Besides the aforementioned two methods, there are other models that detect other types of earnings management, such as timing of asset sales, classificatory shift, earnings guidance, etc. (Athanasakou et al. 2009, 2011; Gunny 2010; Twedt 2016).

Another strand of the literature develop earnings management detection models which are based on a combination of individual signals or firm characteristics. Beneish (1997), for example, develops a model, based on twelve signals which may reveal managerial incentives, to identify GAAP violators from accruals aggressors. Beneish (1999) provides an accounting-based index which could help to assess the likelihood of earnings overstatement. Dechow et al. (2011) develop a model, namely FSCORE, which can help predict the likelihood of earnings restatement. They start with an analysis of the characteristics of restated firms and employ a logistic regression to estimate the relation between firm's characteristics and the likelihood of misstatement. FSCORE is used as a 'barometer' for financial statement users to quickly and timely assess the likelihood of earnings misstatements. The models developed by Beneish (1997, 1999) and Dechow et al. (2011) are more practical and offer advancements to the earlier models (such as Jones 1991; Dechow et al. 1995, etc.) because they are validated using *ex-post* indicators of earnings management which typically have very low Type I error rate (Dechow et al. 2010). However, these models are not entirely free from limitations. One issue is that these studies focus on firms that are subject to enforcements by an authority, such as the Securities and Exchange Commission (SEC hereafter), and are typically large since government agents like SEC would normally aim to maximise public benefits given its constrained budget. In addition, Dechow et al. (2010) highlight that SEC is more likely to target egregious misstatements and avoid ambiguous cases of aggressive but within-GAAP earnings management. Thus, the *ex-post* indicators could potentially suffer from high Type II error rate and the predictive power of the resulting models can neither be generalised to other markets nor to the earnings management firms which are not enforced by SEC.

One way to go around the problems of high Type II error rate of the ex-post indicators of earnings management is to evaluate the likelihood of earnings management based solely on the expected distribution of figures reported in financial statements. A recent and interesting strand of the literature argues that financial statement numbers, like many other natural sets of numbers, should theoretically be distributed according to Benford's law. This law posits that

the probability the first digit of a figure reported on financial statements is one is biggest, and such probability decreases as the first digit gets larger. Hence, if a set of financial statements report figures which are distributed too differently from what is expected under Benford's law, it is a sign that those figures have been manipulated. Amiram et al. (2015) develop a measure, called FSD_SCORE, which captures the deviation of the distribution of financial statement figures of a firm in a given year from the theoretical distribution posited by Benford's law and provide evidence that FSD_SCORE is effective in flagging up financial statements which contain errors. Nguyen et al. (2021) confirm the effectiveness of using Benford's law in predicting earnings management in the UK context. The advantage of this approach is that FSD_SCORE does not require time-series or cross-sectional data to be estimated, nor does it rely on an external indicator of earnings management. Amiram et al. (2015, p. 1558) suggest that FSD_SCORE could be a 'useful tool to augment existing techniques to access accounting data quality'.

Sloan (1996) initiates another strand of accounting literature by showing that accruals are negatively related to future returns. Xie (2001) goes even further showing that it is the discretionary accrual component which mainly drives Sloan's results. The evidence seems to suggest that the market misprices the information contained in accruals, and especially the component over which managers could exercise their discretion to manipulate. To date, the evidence of market mispricing total and discretionary accruals remains one of the most persistent 'market anomaly' with many studies confirming its existence (e.g. Desai et al. 2004; Cheng and Thomas 2006; Mashruwala et al. 2006; Soares and Stark 2009; Choy et al. 2021). However, despite the established evidence of the existence of the accruals anomaly, the reason behind it is still controversial. Sloan (1996) attributes the accruals anomaly to the market's systematic overestimation of the persistence of accruals and underestimation of the persistence of cash flows. That the market irrationally misprices accruals is supported by, for example, Hirshleifer et al. (2012), Allen et al. (2013) and Papanastasopoulos (2020). On the other hand, many other authors suggest the returns to trading strategies designed to exploit the market mispricing of accruals are a fair compensation for risk (Khan 2008) or rational under the Q-theory of investment (Wu et al. 2010).

As accruals, and especially its discretionary component, is widely used as a measure of earnings management, the literature tends to suggest the market does not see through, hence underreacts to, earnings management. Nevertheless, one issue remains unexplored. While a manager can assumingly hide the 'true' earnings through earnings management, he or she cannot hide the surrounding context (for example, manager's motivations, pressures or opportunities for managing earnings, etc.). Hence investors should be able to 'sense' the existence of earnings management by observing the surrounding context. If investors are still 'fooled' by earnings management, it implies that they must have mispriced *both* the magnitude *and* the context in which such manipulation occurs. In other words, if we could document that the market also misprices the context of earnings management incrementally beyond the mispricing of discretionary accruals, it would offer more convincing evidence that the market could not see through earnings management, which is crucial to understand how the market processes publicly available financial statement information. Following this intuition, we form the hypothesis as follows:

Ceteris paribus, firms with a more (less) susceptible context of earnings management yield lower (higher) abnormal returns.

To empirically test this hypothesis, this paper develops a model that captures the context of earnings management and investigates if such model can predict future stock returns.

3 Sample selection

The sample comprises all UK listed stocks during the period from 1995 to 2012.[1] The paper builds upon a rich literature on earnings management and the market mispricing of accruals, of which many existing studies focus on the US market. Nevertheless, the choice of the UK market as the setting for this study is justifiable for a variety of reasons which make the paper an interesting and important contribution to the advancement of our knowledge on these topics. First, the UK market offers a unique setting shaped by several characteristics of the environment in which listed companies operate, ranging from financial reporting and corporate governance regulations, cultural factors and the norms in business and reporting practice, to the popular bases of share ownership in listed companies. For example, the UK business norms and financial reporting practices, especially when it comes to selecting the mechanism for earnings management, are quite different from other developed markets such as the US (see, for example, Bond 2000; Athanasakou et al. 2009, 2011). In addition, institutional holdings tend to be more prevalent in the UK compared to other markets. Institutional stockholders, especially financial institutions and professionally-managed funds, typically play a more active monitoring role, which in turns constraints managers' discretion over financial reporting practices (Chung et al. 2002). These characteristics suggest that using the UK market as a setting for research on earnings management detection would yield interesting and unique insights.

Second, the UK is one of the world's major economies with one of the largest stock markets. Hence, knowledge on how the UK market operates cannot be understated. Compared to the US, the UK-based literature on earnings management is remarkably thinner and that creates an important gap for further studies to fill in. This paper does not simply replicate a US study in the UK, rather it provides evidence which is directly relevant in the UK context with a number of implications for other developed markets such as the US. Having said that, while we would leave a full replication of the study (to the US or any other markets) for future research, we offer a reassurance for subsequent studies to build upon what we find by showing that most of the main conclusions from this paper could be generalised to, at least, the US market.

Finally, using the UK market allows us to make use of a unique dataset which could significantly add strength to our analysis as well as reinforce previous findings using US

---

[1] We acknowledge that the sample period may appear a bit old by the time the paper is published. However, given the nature of our study, which is to develop and test a model, we feel that the choice of the sample period would have little bearing on the results of the study. We have also conducted an out-of-sample test which employs US data during the period from 1987 to 2013 (see Subsect. 5.5). On balance, our results are shown to be generalizable, and the choice of sample period is unlikely to have significant impact on our findings.

data. One of the recent strands in the earnings management literature is to employ an *ex-post* indicator of earnings management, most popular in the US thanks to rich and readily available data on earnings restatements, such as the SEC's accounting and auditing enforcement releases or US Government Accountability Office's (GAO) releases of restatements. These *ex-post* measures of earnings management have significant advantages as well as drawbacks (see, for example, Dechow et al. 2010). One most notable pitfall of the *ex-post* measures is the sample selection biases. In particular, the SEC or GAO does not randomly select firms to investigate. Due to constrained resources, they follow specific strategies to target firms for investigation, such as to prioritise large companies, unambiguous cases, or serious frauds. As far as the accounting profession is concerned, such pitfalls could not be completely corrected. However, it could be mitigated by having more datasets where the investigated firms are selected by other authorities applying different sampling strategies. One of the analyses in this paper employs the sample of firms subjected to investigation by the UK Financial Reporting Review Panel (FRRP), which has a rather different sampling strategy compared to the SEC and GAO in the US (although the objectives might be similar). Hence, our evidence linking the FRRP-investigated firms with earnings management complements and further mitigates the concerns of using *ex-post* measures of earnings management. Moreover, our use of the FRRP data also makes the paper especially distinctive compared to previous UK-based studies.

To avoid survivorship bias, we include both live and dead stocks. We exclude both financial and utilities firms due to their distinct financial reporting requirements. Datastream is the main source of financial data, except for external auditor and merger and acquisition deals for which we use Bloomberg. Data from Datastream and Bloomberg are combined using the International Securities Identification Number (ISIN). Therefore, we also exclude firms which do not have an ISIN. For firms which have more than one type of common stocks, only one is included in the sample. To ensure comparability, we restrict our sample to only those firms which report financial results in British Pound Sterling and whose financial years have between 350 and 380 days. Our sample also excludes firms with market value of less than £1 million to avoid very small firms which are typically thinly traded in practice but can influence the returns on the equally-weighted portfolios. We also exclude stocks with negative market-to-book ratios. Finally, we require data to be available to calculate the variables as described in the Appendix (except for corporate governance and compensation variables) to arrive at the final main sample. The sample[2] consists of 11,920 firm-year observations from 1866 unique firms across 43 Datastream level-six industries. We winsorise all continuous variables at the 1st and 99th percentiles to mitigate the influence of extreme values.

---

[2] This sample is employed for all of the main analyses, except for (a) the multivariate regressions which are used to validate the ESCORE in Subsect. 5.2.2 (where our sample depends on the availability of corporate governance data); (b) the test using an ex-post measure of earnings management in Subsect. 5.2.3 (where we restrict the analyses to the firms covered by the FRRP); and (c) the test using Benford's law in Subsect. 5.2.3 (where we further restrict the sample to include only those firms with enough data to calculate FSD_SCORE).

4 Research methodologies

4.1 The construction of ESCORE

One of the key contributions of our study is the empirical measure of the context of earnings management. Watts and Zimmerman (1986) provides the theoretical foundation for the establishment of a context in which firms might act in certain ways when it comes to exercise discretion over the choice of accounting practices. In particular, earnings management in this framework is likely to exist in the presence of one or a combination of the following triggers: a bonus plan (the bonus plan hypothesis), high cost of renegotiation of debt (the debt/equity hypothesis) and high political cost (the political cost hypothesis). Subsequent studies in this large, and still growing, literature add significantly to the list of earnings management signals, which have outgrown the three broad categories suggested by Watts and Zimmerman (1986). The subsequent literature generally categorizes earnings management signals into the managerial incentives and pressures, with the added dimensions which cover the scope for earnings management to happen, including the practical constraints and other firm's innate factors (see, for example, Healey and Wahlen 1999, Beneish 2001, Dechow et al. 2011).

Within the scope of this study, we define the 'context of earnings management'[3] as (a) the incentives to manage earnings, (b) the pressures under which managers are more likely to resort to earnings management, (c) the constraints on earnings management, and (d) the innate factors of the firm which could indicate the existence of earnings management. To capture the context of earnings management, we construct a composite index named ESCORE which is the sum of 15 individual binary variables, each taking a value of one if a firm has a suspicious signal and zero otherwise. We select these signals based on the rich extant earnings management literature and discuss their construction in the sub-sections below. In selecting the individual signals, we focus on those signals that suggest higher likelihood of earnings management without preference to either aggressive or conservative practices. As a result, ESCORE is not a *signed* measure of earnings management. We conjecture that the relation between ESCORE and future returns comes from the power of the ESCORE to reveal the context in which earnings management is likely, and not from its ability to reveal the sign and magnitude of such manipulation.

4.1.1 Benchmark construction procedure

To construct ESCORE, we first need a 'benchmark' for each individual signal. For example, we know from prior literature that small firms are more susceptible to earnings management (Lang and Lundholm 1993; Dechow and Dichev 2002). However, we need a 'benchmark' to determine which firms should be classified as 'small'. Such 'benchmarks' should reflect the characteristics of the corresponding industry. We explain the procedures to construct these benchmarks below.

---

[3] Within this paper, 'context of earnings management' refers to the aggregation of individual signals which suggest the existence of earnings management rather than referring to the use of textual analysis that reads and summarizes the language, symbols, terminology, etc., present in the text of the financial statements of a firm.

First, for each industry-year,[4] we rank firms based on γ (where γ is substituted by the relevant financial signals used in this study). Next, we use the 20th and 80th percentiles in each industry-year as the lower and upper benchmarks, denoted as $\gamma_{k,t}^{20}$ and $\gamma_{k,t}^{80}$ respectively, where $k = 1 \ldots 43$ represents unique Datastream level-six industries in the sample, and $t = 1995 \ldots 2011$ represents 17 sample years.[5] If a signal is lower (higher) than $\gamma_{k,t}^{20}$ ($\gamma_{k,t}^{80}$), we consider it too low (high). We apply this procedure to all individual signals that require an industry-specific benchmark to construct. Next, we construct ESCORE using individual signals under four broad categories.

4.1.2 Incentives to manage earnings

The first category covers the incentives for managing earnings, including equity issue, debt issue, share-for-share merger and acquisition, and stock overvaluation. Prior evidence suggests that firms inflate earnings prior to equity issues (Teoh et al. 1998b; Cohen and Zarowin 2010; DuCharme et al. 2004; Siew Hong and Wong 2002; Rangan 1998; Shivakumar 2000; Iqbal et al. 2009; Iqbal and Strong 2010, Kothari et al. 2016). This study defines the indicator of equity issue, denoted as ESEO, as a dummy that takes a value of one if (i) a firm's outstanding shares at the end of the current fiscal year (year *t*) increase by at least 5% compared to last year's (year *t*-1) and (ii) there are positive proceeds from issuing common/preferred stocks in the current year, zero otherwise.

Managers may also like to 'decorate' financial statements prior to a major debt issue to negotiate the cost of debt down. Athanasakou and Olsson (2012) find a positive relation between an indicator of debt issue and earnings management. To capture debt-issue-related incentives to inflate earnings, we define the indicator of debt issue, EDDEBT, as a dummy that takes a value of one if DDEBT is 5% or higher, where DDEBT is calculated as the percentage change of total of short- and long-term debt[6] at the end of the current year (year *t*) compared to last year's (year *t*-1) total debt, zero otherwise. The 5% benchmark is employed to ensure that the issue is large enough for managers to consider managing earnings.

---

[4] We do the ranking within each industry to put a firm's characteristics in the relevant context of its corresponding industry. For example, it would be problematic to compare tangible fixed assets of a firm in the oil and gas industry with those from the service industry. We acknowledge, however, that ranking within each industry would require the sample to be large, which might impede the applicability of our model to smaller markets. Besides, our approach requires the benchmarks to be constructed every year, which might have some issues with practicability as one would need to wait for all firms in the industry to publish their financial statements before a benchmark could be constructed. To mitigate these concerns and to avoid discouraging future studies to apply ESCORE in smaller markets, we replicate all of the main analyses using a version of the ESCORE which exclude all of the signals which require an industry-year benchmark to construct. Unreported results show that the main conclusions of the paper do not change qualitatively using the compact version of ESCORE.

[5] The cut-off points are set at 20th and 80th percentiles to ensure the resulting individual signals would not flag up too many firms as 'suspicious'. However, it is admittedly an arbitrary choice. Although it is needed for the research to go ahead, all main analyses in the paper are replicated using the cut-off points at 33rd and 67th, 25th and 75th percentiles as well as 10th and 90th percentiles. To further mitigate the concern that transforming continuous variables into binary signals using arbitrary cut-off points might fail to flag up earnings management if the managers deliberately avoid reporting extremely high or low signals, we also remove all signals requiring cut-off points (including EOV, EDEBT, ESIZE, EBLOAT, ECAP and EBT) to calculate a suppressed version of the ESCORE and redo all of the main test. Untabulated results show that none of the main conclusions are sensitive to the choice of the cut-off points to construct ESCORE.

[6] A value of zero is assigned to observations with missing Datastream's data for short- and long-term debt.

Firms also have strong incentives to inflate earnings prior to share-for-share mergers and acquisitions (M&A) in an attempt to temporarily push stock price up to minimise the number of shares paid (Erickson and Wang 1999; Botsari and Meeks 2008, 2018; Louis 2004). We define the indicator of share-financed M&A, denoted as EMA, as a dummy that takes a value of one if a firm announces an M&A deal within the current financial year for which shares are proposed as (part of) the payment method, zero otherwise.

Recent literature also considers the effect of stock market overvaluation on earnings management. Jensen (2005) conjectures that overvaluation creates a pressure on firms to inflate earnings to maintain their high market valuation. Empirical evidence also supports the premise that overvaluation induces income-increasing earnings management (Chi and Gupta 2009; Houmes and Skantz 2010; Badertscher 2011; Duong and Pescetto 2019). To capture this signal, we define the indicator of overvaluation, denoted as EOV, as a dummy that takes a value of one if a firm's beginning of year t market-to-book ratio (MTB), calculated as market value to book value of equity, is higher than the corresponding $MTB_{k,t}^{80}$, zero otherwise.

4.1.3 Pressures to manage earnings

We use various proxies in the model to capture the pressures to manage earnings, including meeting or just beating earnings benchmarks, financial distress, debt level, firm size, and business life cycle stage. Burgstahler and Dichev (1997) document a discontinuity of earnings around two important benchmarks, namely zero earnings and last year's earnings. A similar pattern has also been documented in the UK (Gore et al. 2007; Al-Shattarat et al. 2018). The indicator employed to capture the pressure to meet or beat zero earnings benchmark, denoted as EROA, is defined as a dummy that takes a value of one if a firm's returns-on-assets ratio (ROA), calculated as earnings before extraordinary items in year t scaled by beginning (year t-1) total assets, is equal to or larger than zero but smaller than 0.01, zero otherwise. EDROA, the indicator employed to capture the pressure to avoid reporting earnings decreases, is defined as a dummy that takes a value of one if a firm's DROA, calculated as the change in earnings before extraordinary items in year t compared to that in year t-1 scaled by beginning total assets, is between zero and 0.005, zero otherwise.

Prior research shows that firms would engage in earnings management if the unmanaged earnings fall short of the expected dividends by a small amount (Daniel et al. 2008; Atieh and Hussain 2012). EDIV captures this pressure. It is a dummy that takes a value of one if a firm's dividend deficit, denoted as DIVDEF and calculated as the difference between net income and total cash dividends in year t scaled by beginning total assets, is between zero and 0.01, zero otherwise.

Financially distressed firms are understandably under pressure to inflate earnings. Garcia Lara et al. (2009) show that such firms manage earnings upwards. Beneish (1997) reports that financial distress is a factor that leads to GAAP violation. To capture the presence of these pressures, we estimate the UK-based ZSCORE (Taffler 1983). Taffler (1983) and Agarwal and Taffler (2007) show that UK firms with negative ZSCORE are more likely to become bankrupt. Following this evidence, we define EDISTRESS (the indicator of financial distress) as a dummy that takes a value of one if a firm's ZSCORE in year t is negative, zero otherwise.

The use of debt also has implications for earnings management. Watts and Zimmerman (1986) suggest that debt contracts have a vital influence on a firms' accounting policy. On one hand, higher debts induce pressures on firms to inflate earnings. Indeed, debts

usually come with some covenants which firms need to comply with. Violating debt covenants leads to firms being penalised by lenders by means of higher cost of debt (Dichev and Skinner 2002; Dyreng et al. 2020). Therefore, firms with more debt have a greater pressure to manage earnings to avoid violation of debt covenants. DeFond and Jiambalvo (1994) find that abnormal accruals are significantly higher in the years preceding debt covenant violations. Ghosh and Moon (2010) find that firms with high debt would have strong incentive to manage earnings. On the other hand, however, the literature also suggests that firms with low level of debt are also likely to engage in earnings management (Astami and Tower 2006). In addition, the evidence that financial leverage is positively related to accounting conservatism (for example, Watts 2003a, b; Pae 2007) implies that firms with little debt are less bound contractually and their reported earnings are less subject to scrutiny from lenders, hence there could be more scope for earnings management. In brief, the literature suggests that firms which have either too high or too low debts are suspicious of earnings management. The ZSCORE, as explained earlier, has already captured firms with high debts. The indicator of firms with too little debts, denoted as EDEBT, is defined as a dummy that takes a value of one if a firms' beginning of year t DEBT, measured as the total of short- and long-term debt scaled by year t total assets, is lower than the corresponding $DEBT_{k,t}^{20}$, zero otherwise. EDEBT captures the context in which firms are subject to less scrutiny from lenders, hence have more room for managing earnings, in both directions.

It is also more difficult for large firms to manage earnings due to their high public visibility (Lang and Lundholm 1993; Dechow and Dichev 2002; Zhang et al. 2019). Smaller firms, on the contrary, usually face less public attention and they struggle to perform under various financial constraints. Hence small firms are often more likely to engage in earnings management, especially if the managers believe the struggles are just transitory. Indeed, various studies in the earnings management literature use firm size as a control variable and the evidence shows that firm size is related to discretionary accruals. ESIZE, the indicator of small firms, is a dummy that takes a value of one if a firm's beginning of year t market value of equity (MVE hereafter) is lower than the corresponding $MVE_{k,t}^{20}$, zero otherwise. ESIZE captures the context in which firms are subject to less scrutiny from the public, hence could have more room to manage earnings, in both directions.

The last variable in this group, ECYCLE, captures firms which are in the introduction and growth stage in their business life cycle. Young listed firms, most of which use funds from the capital markets for the first time, are usually under pressure to perform and grow. Accounting manipulation could be a way for these young listed firms to respond to such pressures (Beneish 1997; Dopuch et al. 1987). Growth firms usually face strong investment opportunities and are expected to deliver strong growth and financial performance. Fama and French (1995) show that growth firms typically report higher earnings. Lakonishok et al. (1994) suggest that the market generally places too much expectation on growth stocks which results in market overreaction. Under such pressure, firms might have to resort to earnings management should their underlying economic performance fall short of the expectation to avoid market penalty. Such prediction has been substantiated by empirical evidence (Skinner and Sloan 2002). Following Dickinson (2011), ECYCLE is defined as a dummy that takes a value of one if a firm's operating cash flows are negative, financing cash flows are positive and investing cash flows are negative (introduction stage), or its operating and financing cash flows are positive while its investing cash flows are negative (growth stage), and zero otherwise. All cash flows are measured at year t-1.

4.1.4 Constraints on earnings management

We include external auditor and balance sheet bloat in the model to represent constraints on earnings management. Prior literature shows that external audit quality plays a major role in constraining accruals management (Becker et al. 1998; Francis et al. 1999; Alzoubi 2016). Krishnan (2003) finds that firms whose external auditors have more industry experience, on average, have less discretionary accruals. Following this evidence, several studies use an indicator of firms being audited by the Big-5 auditors as a control variable in regression where the dependent variable is discretionary accruals and in general these studies report a significant negative relationship (Zang 2012; Athanasakou and Olsson 2012; Choi et al. 2018). Nevertheless, the existing evidence about the constraining role of external auditors is mixed regarding the sign of the manipulation. For example, Becker et al. (1998) predict that the presence of the Big-5 auditors is negatively related to the signed discretionary accruals, while Francis et al. (1999) only present evidence about the relationship between Big-5 external auditors and the absolute value of discretionary accruals. Overall, the absence of Big-5 auditors could give room for firms to manage earnings more easily, in both directions. We, therefore, use the absence of a Big-5 external auditor as a signal of earnings management, but do not predict the sign of the manipulation. EAUDIT, the indicator of the absence of Big-5 auditor, is defined as a dummy that takes a value of one if a firm is not audited by the Big-5 accountancy firms in year t. The Big-5 is defined as the following firms and their affiliates: Arthur Andersen, Deloitte Touche Tohmatsu, Ernst & Young, KPMG, PriceWaterhouseCoopers. Audit firms which are later merged with one of the Big-5 are also considered as part of the Big 5 (e.g. Coopers and Lybrand is deemed as PriceWaterhouseCoopers). If data on the auditor is missing from Bloomberg for a firm in a year, it is assumed that the firm is not audited by a Big-5 auditor.

Due to the self-reversal nature of accruals, we consider past use of accruals management to act as a constraint on further engagement (Barton and Simko 2002). Net operating asset (NOA) can proxy for the 'balance sheet bloat', which captures the constraint induced by past engagement in accruals management (Houmes and Skantz 2010, Beuselinckc et al. 2019). Firms with high NOA have engaged extensively in income-increasing accruals management in the past, which in turn constrains the firm's ability to further manage accruals. Following this, we calculate NOA as the sum of net book value of equity and total debt minus cash and cash equivalents, all scaled by total assets. The indicator of low balance sheet bloat, denoted as EBLOAT, is a dummy that takes a value of one if a firm's beginning of year t NOA is lower than the corresponding $NOA_{k,t}^{20}$, zero otherwise.

4.1.5 Innate characteristics

Earnings management is engaged not only because of managerial motives, but also due to some firm's innate factors (Dechow and Dichev 2002; Francis et al. 2005, 2004; Athanasakou and Olsson 2012). Dechow and Dichev (2002) suggest some important innate factors which could imply earnings management, for example, the variability in some fundamentals such as sales or cash flows, firm size, operating cycle and incident of losses. Several of the innate factors identified in the extant literature as signals of earnings management, such as firm size, operating cycle and incident of losses, have been covered earlier. We do not include some of the innate factors, such as the variability of sales and cash flows, which require long history of data to calculate. Requiring long history of data would eliminate young firms from the sample, a practice that may introduce bias in the main analysis since

some earnings management signals (e.g. ECYCLE) are designed to capture young firms. We also do not consider the intensity of intangible assets due to insufficient data to establish plausible industry benchmarks. However, to capture the intensity of tangible assets, we estimate CAP as the ratio of property, plant, and equipment divided by total assets. Prior literature shows that smaller CAP is associated with poor earnings quality, hence such firms are suspicious of earnings management (Athanasakou and Olsson 2012; Francis et al. 2004). ECAP, the indicator of low intensity of tangible assets, is a dummy that takes a value of one if a firm's beginning of year t CAP is smaller than the corresponding $CAP_{k,t}^{20}$, zero otherwise. For firms which have the signal ECAP of one, we do not predict the sign of the manipulation.

Lastly, some studies document the effect of book-tax conformity on earnings management (Hanlon and Heitzman 2010; Athanasakou and Olsson 2012; Sundvik 2017). If one agrees that taxable profits are difficult and costly to manipulate, then the more accounting earnings diverge from taxable profits, the more likely it is that such accounting earnings have been manipulated. Generally, the evidence supports such intuition (Desai 2005). Following the literature, we calculate the book-tax difference, denoted as BOOKTAX, as the absolute value of the difference between year t's reported pre-tax income and an estimate of total taxable profits, denoted by TTP, all scaled by sales in year t. We estimate TTP using the lower and upper limit for marginal tax relief (denoted LL and UL, respectively), small profit tax rate (SR) and main tax rate (MR) applicable at the time in conjunction with the reported income tax expenses (TXT). We source LL, UL, SR and MR in each sample year from HM Revenue & Customs (2013). With only published information, it is almost impossible to estimate TTP. Therefore, we make some assumptions to simplify the estimation. First, we assume that the reported tax expenses represent solely the amount of income tax levied in the considered period (i.e. no extraordinary penalty or retrospective payment or anything else of that nature). Second, for the profits that fall between the LL and UL, we assume that the tax rate is the average of SR and MR, denoted as AR, to avoid complex calculation. With these assumptions, TTP is worked back from the tax expenses as follows:

$$If\ TXT\ \leq\ 0,\ then\ TTP\ =\ 0$$

$$If\ 0\ <\ TXT\ \leq\ LL{\times}SR,\ then\ TTP\ =\ \frac{TXT}{SR}$$

$$If\ LL{\times}SR\ \leq\ TXT\ \leq\ (UL-LL){\times}AR,\ then\ TTP\ =\ \frac{TXT - (LL \times SR)}{AR} + LL$$

$$If\ TXT\ \geq\ (UL-LL){\times}AR,\ then\ TTP\ =\ \frac{TXT - (LL \times SR) - [(UL - LL) \times AR]}{MR} + UL$$

$$(1)$$

We define EBT as a dummy that takes a value of one if a firm's BOOKTAX is higher than the corresponding $BOOKTAX_{k,t}^{20}$, zero otherwise. EBT, therefore, captures firms which have reported accounting earnings that are different from taxable profits, an indication that accounting earnings may have been managed, in both directions.

4.1.6 The ESCORE

We define the composite ESCORE as the sum of all 15 individual binary signals as follows:

Table 1 Eigen values of the correlation matrix from principal components analysis

Principal components	Eigenvalue	Difference	Proportion	Cumulative
1	1.9246	0.5198	0.1283	0.1283
2	1.4048	0.0552	0.0937	0.2220
3	1.3496	0.0909	0.0900	0.3119
4	1.2587	0.1562	0.0839	0.3959
5	1.1025	0.1143	0.0735	0.4694
6	0.9882	0.0511	0.0659	0.5352
7	0.9371	0.0084	0.0625	0.5977
8	0.9286	0.0288	0.0619	0.6596
9	0.8999	0.0930	0.0600	0.7196
10	0.8068	0.0561	0.0538	0.7734
11	0.7507	0.0328	0.0500	0.8234
12	0.7179	0.0493	0.0479	0.8713
13	0.6686	0.0012	0.0446	0.9159
14	0.6673	0.0726	0.0445	0.9604
15	0.5947	–	0.0396	1.0000

The table reports the Eigen values of the correlation matrix resulted from principal component analyses on 15 individual components of ESCORE. Definitions of variables are in the Appendix

$$
\begin{aligned}
ESCORE = {} & ESEO + EDDEBT + EMA + EOV + EROA + EDROA \\
& + EDIV + EDISTRESS + EDEBT + ESIZE + ECYCLE \\
& + EAUDIT + EBLOAT + ECAP + EBT
\end{aligned}
\tag{2}
$$

As designed, ESCORE is an integer which can theoretically range from 0 to 15. The smaller (larger) the ESCORE, the less (more) suspicious the context of earnings management surrounding a firm is. Being aggregated from 15 individual signals, an immediate question is whether those signals are correlated and thus could be reduced to a more parsimonious model through, for example, principal component analysis. To respond to this possibility, we calculate and look at the Eigen values from principal components analysis. The results are reported in Table 1. The first principal component, which has the largest variance of any linear combination of the individual scores, could explain only 12.83% of the total variance. Subsequent principal components contribute even less than this and range from 9.37% to 3.96%. Looking at the Eigen vectors, we could not find too high loading on any particular variables, which suggests that none of the individual scores plays a dominant role in the variance of the composite ESCORE. Overall, it seems unlikely that variable reduction through principal component analysis would significantly affect the ESCORE compared to the simple sum-of-binary-variable approach.

Table 2 shows the distribution of firm-year observations across ESCORE portfolios. Although ESCORE could theoretically range from 0 to 15, no firm in our sample accumulates more than 9 signals. In subsequent analyses, we pay particular attention to the portfolios of low and high ESCORE stocks. For this purpose, we arbitrarily group stocks with ESCORE of zero into the low-ESCORE, those with ESCORE of six and above into the High-ESCORE and the rest to the medium-ESCORE group. Since there are fewer number of stocks with larger ESCORE, we group all stocks with ESCORE of 6 and above into the High ESCORE group (865 observations). The purpose is to ensure that the High ESCORE

Table 2 Distribution of firms across ESCORE groups

ESCORE	N	ESCORE GROUP	N
0	862	Low (0)	862
1	2218		
2	2925		
3	2381		
4	1675		
5	994	Medium (1–5)	10,193
6	519		
7	232		
8	88		
9	26	High (6–9)	865

The table reports the distribution of observations across groups sorted by ESCORE. Low(0), Medium (1–5), and High (6–9) represent firms with ESCORE of 0, firms with ESCORE of 1 to 5, and firms with ESCORE of 6 to 9, respectively. Definitions of variables are in the Appendix

portfolio has comparable number of observations to the Low ESCORE counterpart (which comprises 862 stocks with ESCORE of zero). Intuitively, our grouping scheme is equivalent to considering that the context surrounding a stock which has accumulated six or more signals is highly susceptible of earnings management.[7]

4.2 Measures of accruals and real earnings management

To test how the context of earnings management is effective in predicting future stock returns, we first need to test if the ESCORE could indeed capture such context. We illustrate the effectiveness of ESCORE to detect the context of earnings management by looking at how other traditional measures of earnings management (e.g. discretionary accruals and real earnings management proxies) vary as the context (captured by ESCORE) changes. For this, we consider six proxies of earnings management.

To begin, we employ the cross-sectional version of the modified-Jones model (Jones 1991; Dechow et al. 1995) to estimate discretionary accruals (DAC).[8] In this model, total accruals are calculated as the difference between income before extraordinary items and net operating cash flows. The calculation of total accruals follows the cash flows approach to avoid the potential measurement errors identified by Hribar and Collins (2002).[9] To

[7] The choice of the cut-off at six is arbitrary since we cannot say a stock with ESCORE of five is qualitatively less 'susceptible' than another one with ESCORE of six. While we cannot proceed without an arbitrarily-determined cut-off point, unreported results show that the main conclusions of the paper do not change qualitatively if we group stocks with ESCORE of zero and one into the Low ESCORE portfolio (3,080 observations) and those with ESCORE of four and above into the High ESCORE portfolio (3,534 observations).

[8] In unreported results, we also employ the cross-sectional version of the original Jones model. The main results remain unchanged.

[9] Other authors, such as Gore et al. (2007), argue that the cash flow approach is also problematic. We have, therefore, replicated the test using the balance sheet approach. The results are not reported for brevity, but none of the main conclusions change qualitatively.

obtain DAC, we run regressions in each (Datastream level-six) industry-year with at least fifteen observations.[10]

Although there are other competing models to estimate discretionary accruals (Dechow et al. 1995; Guay et al. 1996; Bernard and Skinner 1996; Young 1999; Thomas and Zhang 2000; Peasnell et al. 2000; Fields et al. 2001), the existing literature generally suggests that there is no other model that clearly outperforms the modified-Jones model (Peasnell et al. 2000; Botsari and Meeks 2008). Nevertheless, many UK studies focus only on working capital accruals arguing that depreciation is not a suitable means to manage earnings since it is highly visible and if earnings are managed through depreciation, the effects could be unwound quite easily by financial statement users (Young 1999; Peasnell et al. 2000; Gore et al. 2007). To account for this argument, we also estimate discretionary working capital accruals (DWAC) using the 'margin model' as described in Peasnell et al. (2000), which has been shown to work well in the UK context.

We next consider real earnings management by following Roychowdhury (2006) to estimate three measures of real earnings management, namely the abnormal cash flow (DCF), abnormal production cost (DPROD) and abnormal discretionary expense (DDISEXP). Regressions are estimated for each (Datastream level-six) industry-year with at least fifteen observations. DPROD is exactly as described in Roychowdhury (2006). Nevertheless, to make their sign consistent with other measures of earnings management used in this paper, we multiply Roychowdhury's (2006) measures of abnormal cash flow (DCF) and abnormal discretionary expenses (DDISEXP) by -1. As a result, a positive value of DCF and DDIS-EXP would imply income-increasing earnings management and vice versa.

DCF, DPROD and DDISEXP capture three dimensions of real earnings management, namely the manipulation of sales activities, production activities, and discretionary expenses. These three ways of managing earnings could be used as substitutes, i.e. a manager would manipulate earnings through changing real operation decisions in one or two areas out of the three, and not necessarily all of them at the same time. As a result, for example, when the context suggests a firm is inflating earnings and the firm decides to do it through sales manipulation, DPROD and DDISEXP are not necessarily high. It is, hence, important to look at the overall real earnings management strategy rather than just the individual ones. To facilitate this, we also construct a composite measure that pools together the three measures of real earnings management as follows:

$$TOTALRM_{i,t} = \left[\frac{DCF_{i,t} - \overline{DCF_{t,k}}}{\sigma(DCF)_{t,k}} + \frac{DPROD_{i,t} - \overline{DPROD_{t,k}}}{\sigma(DPROD)_{t,k}} + \frac{DDISEXP_{i,t} - \overline{DDISEXP_{t,k}}}{\sigma(DISEXP)_{t,k}} \right] / 3 (i \in k) \quad (3)$$

where $TOTALRM_{i,t}$ is the composite measure of real earnings management of firm i in year t; $\overline{DCF_{t,k}}$, $\overline{DPROD_{t,k}}$, $\overline{DDISEXP_{t,k}}$ [$\sigma(DCF)_{t,k}$, $\sigma(DPROD)_{t,k}$, $\sigma(DISEXP)_{t,k}$] are, respectively, the mean [standard deviation] of DCF, DPROD, DDISEXP of all firms in industry k in year t; and $k = 1 \ldots 43$ are 43 unique Datastream level-six industries.

---

[10] Many UK studies (e.g. Gore et al. 2007; Botsari and Meeks 2008) require 6 observations when estimating the proxies for earnings management. This approach would yield larger sample size, but at the cost of the concern over the reliability of regressions with too few observations. We follow another strand of the literature (including, for example, Roychowdhury 2006) and require 15 observations to address that concern.

The above procedure converts DCF, DPROD and DDISEXP into standardised variables before averaging them. $TOTALRM_{i,t}$, therefore, captures the combined effects of Roychowdhury's (2006) real earnings management strategies. Compared to other studies which simply add DCF, DPROD and DDISEXP together (e.g. Cohen and Zarowin 2010), our approach is advantageous because our standardization process could mitigate the concerns regarding adding variables with different distributions.

The six measures of earnings management as described above are then employed to test the efficacy of ESCORE in capturing the context of earnings management. ESCORE is primarily designed to capture the context in which earnings management is more likely to occur, not the *sign* of such manipulation. Some components of ESCORE, including ESEO, EDDEBT, EMA, EOV, EBLOAT, EROA, EDROA, EDIV, EDISTRESS, ECYCLE, predict inflationary (i.e. aggressive) earnings management, while others, including EAUDIT, EBT, ECAP, EDEBT, ESIZE, only suggest the presence of earnings management behaviour regardless of the sign. We, therefore, test the effectiveness of ESCORE in two ways. First, we examine if ESCORE is able to indicate the presence of earnings management, in both directions, by looking at how the absolute values of DAC, DWAC, DCF, DPROD, DDISEXP and TOTALRM (denoted by ADAC, ADWAC, ADCF, ADPROD, ADDISEXP and ATOTALRM, respectively) vary across ESCORE groups. Second, as most of the components of ESCORE suggest an inflation of earnings, we also expect that ESCORE could identify the context in which the most aggressive earnings management occurs. For investors, aggressive earnings management is arguably more harmful, hence it is important to see if ESCORE can indicate those circumstances. For this purpose, we examine the association of ESCORE with the indicators of aggressive earnings management, denoted by HDAC, HDWAC, HDCF, HDPROD, HDDISEXP and HTOTALRM. We define these as the dummy variables that take a value of one if the stock is in the top quintile ranked in each industry-year by DAC, DWAC, DCF, DPROD, DDISEXP and TOTALRM, respectively.

4.3 Calculation of returns

We calculate ESCORE for calendar year t ($t=1995,…, 2011$) for all stocks with fiscal year ending in any month of the year. For each year, we then sort firms by their ESCORE. Based on the ESCORE of year t, we form portfolios at the beginning of June of year $t+1$ and hold them until the end of May of year $t+2$. For each month, we estimate buy-and-hold raw returns for each stock as the percentage change in Datastream's Return Index, assuming dividend reinvestment, and is denoted by $BHRR_{i,j}^m$. If a stock delists during the holding period, we treat the delisting returns as follows. If a stock does not have a monthly return for June (the first month after portfolio formation), we exclude the firm-year observation from the sample (equivalent to assuming that investors cannot consider the stock for trading due to non-existence). If a stock has a return for June, but then delists before the end of the holding period due to non-performance-related reasons, we assume that the investors earn the returns from portfolio forming date to delisting date, and then reinvest the proceeds in the size-matched portfolio which assumingly bears similar risk compared to the delisted firm. Prior studies (for example, Soares and Stark, 2009; Desai et al., 2004) use this approach to reflect the reality that the returns in most M&A-related delisting cases are positive. We estimate returns on the size-matched portfolio using similar procedure to calculate size-adjusted returns described below. If the delisting is performance-related, we assume that the whole initial investment is lost, hence a delisting return of –100% is used.

To test the effectiveness of ESCORE-based trading strategies, the study uses various measures of buy-and-hold abnormal returns. First of all, we estimate firm-specific monthly buy-and-hold size-adjusted returns as follows. Each year, we sort all stocks with available data from Datastream into deciles based on market capitalization at the end of the previous fiscal year. We then estimate returns for each size decile portfolio d ($d = 1\dots 10$), $SDR_{d,j}^m$, as the average $BHRR_{i,j}^m$ of all stocks which belong to decile d. For each stock, its corresponding size decile and size decile returns are identified. Finally, we estimate the buy-and-hold size-adjusted return for stock i in month j, denoted by $BHSAR_{i,j}^m$, as the difference between the raw return and the return on the corresponding size decile portfolio.

From the above firm-specific returns, the raw and size-adjusted returns of portfolio p, denoted by $BHRR_{p,j}^m$ and $BHSAR_{p,j}^m$, are respectively the equally-weighted $BHRR_{i,j}^m$ and $BHSAR_{i,j}^m$ of all stocks in portfolio p. Following Desai et al. (2004), to avoid the potential inflation of t-statistics when assessing the abnormal portfolio returns over time, we calculate $BHSAR_{p,j}^m$ for each month and treat it as one observation. The t-statistics used to test if $BHSAR_p^m$ and $BHMAR_p^m$ are significantly different from zero are calculated from 204 time-series monthly observations (across 17 sample years).

We calculate $BHSAR_{p,j}^m$ using reference portfolios, an approach which could bias the test statistics (Barber and Lyon 1997; Kothari and Warner 1997). In addition, size-adjusted returns are not capable of capturing some other known dimensions of risk, such as market-to-book and momentum factors. To mitigate these concerns, we also estimate the Fama–French model augmented by the momentum factor (Carhart 1997) as follows:

$$BHRR_{p,j}^m - Rf_j = \alpha + \beta_1\left(Rm_j - Rf_j\right) + \beta_2 SMB_j + \beta_3 HML_j + \beta_4 UMD_j + \varepsilon \qquad (4)$$

where $BHRR_{p,j}^m$ is equally-weighted raw return of portfolio p for month j; Rf_j, Rm_j, SMB_j, HML_j, UMD_j are, respectively, the monthly risk-free rate, returns on the market portfolio, size, market-to-book and momentum factors, all as described and downloaded from the publicly available database of Gregory et al. (2013). We then calculate the monthly buy-and-hold portfolio abnormal returns using the estimated coefficients obtained from regression (4), denoted by $BHAR4F_{p,j}^m$. Similar to the t-test employed for size-adjusted returns, the t-statistic used to test if $BHAR4F_{p,j}^m$ is significantly different from zero is calculated from 204 time-series monthly observations.

Carhart's (1997) approach to estimate abnormal returns is also not flawless, especially in the UK context (Lee et al., 2007; Bauer et al., 2010). Nevertheless, since we use both the reference and regression-based approaches, it would reasonably guard the results against any possible significant biases due to the way abnormal returns are calculated.[11]

The monthly returns as calculated above are used in portfolio tests. For multivariate regressions, we compound monthly returns into annual buy-and-hold returns, denoted by an 'a' superscript in place of the 'm' after each measure of returns, to match with the annual update of the explanatory variables.[12]

---

[11] We also use market-adjusted returns, the CAPM and the Fama–French three-factor model. None of the main results change qualitatively using those abnormal returns metrics.

[12] We also use cumulative returns as robustness check. Untabulated results show that the main conclusions are not affected.

5 Results and discussions

5.1 Descriptive statistics and correlations

Table 3 presents some descriptive statistics of the main variables used in this study. Mean market value of equity, MVE, (£390 million) is larger than the median (£44 million) which suggests the existence of some very large observations. Those large firms could significantly influence the returns of value-weighted portfolios. The paper, therefore, reports the results from applying the equally-weighted scheme in the main portfolio tests.[13] The mean of ROA is –0.0072 while the median is 0.0451, which shows the existence of some very large negative values. This could be a sign of the presence of firms which 'take a bath' since such practice typically involves booking very large losses.

Table 4 presents correlations between the main variables. The correlations between individual signals are quite low (ranging from only 38.1% between EDISTRESS and EBT to −20.5% between EDEBT and EDDEBT) and insignificant in many cases. It suggests that the individual signals capture different and uncorrelated dimensions of the context of earnings management which further reinforces the construction of ESCORE as the sum of all factors. ESCORE also shows significant negative correlation with all measures of returns. This initial evidence suggests that ESCORE could predict stock returns.

5.2 ESCORE and the context of earnings management

5.2.1 Univariate analysis

To test the effectiveness of ESCORE in capturing the context of earnings management, we employ three tests. In the first test, we examine how the six measures of earnings management (as explained in Sect. 4.2) vary as the context of earnings management (i.e. the ESCORE) changes. Table 5 presents the mean of ADAC, ADWAC, ADCF, ADPROD, ADDISEXP, ATOTALRM (the absolute values) and HDAC, HDWAC, HDCF, HDPROD, HDDISEXP, HTOTALRM (the indicators of aggressive earnings management) across ESCORE groups, together with the t-test comparing the means of the High ESCORE group (ESCORE of six and above) with those of the Low ESCORE group (ESCORE of zero). The results show that as ESCORE increases, all of the 12 measures of earnings management also increase monotonically and consistently. The differences across all measures between the High ESCORE and Low ESCORE group are positive, economically large, and statistically significant. The results, therefore, suggest that ESCORE, despite being constructed using a completely different methodology, is consistent with other more traditional proxies of earnings management. An important feature of the ESCORE which could potentially be a useful tool for future research is that ESCORE could proxy for both accruals and real earnings management. In general, the evidence implies that ESCORE is highly effective in capturing the context of earnings management as when the context is more susceptible (higher ESCORE), firms indeed manage earnings in larger magnitudes and are more likely to be aggressors.

---

[13] We also employ the value-weighted scheme as a robustness check. Unreported results show that the main conclusions of the paper do not change qualitatively.

Table 3 Descriptive statistics ($n = 11,920$)

	Mean	25th percentile	Median	75th percentile	Standard deviation
AT (£mil)	402	16	54	200	1204
SALE (£mil)	409	14	55	232	1149
NI (£mil)	19	−1	2	10	73
MVE (£mil)	390	12	44	188	1246
MTB	3.3217	1.0471	1.8317	3.3463	5.0650
ROA	−0.0072	−0.0288	0.0451	0.0965	0.2201
DAC	0.0066	−0.0447	0.0096	0.0621	0.1251
DWAC	0.0017	−0.0383	0.0026	0.0441	0.0926
DCF	0.0033	−0.0793	−0.0060	0.0661	0.2176
DPROD	0.0112	−0.0996	0.0199	0.1423	0.2627
DDISEXP	0.0059	−0.1021	0.0266	0.1486	0.3196
TOTALRM	−0.0202	−0.3648	0.0092	0.3610	0.6446
ESEO	0.2107	0	0	0	0.4078
EDDEBT	0.3790	0	0	1	0.4852
EMA	0.0498	0	0	0	0.2176
EOV	0.2161	0	0	0	0.4116
EROA	0.0344	0	0	0	0.1823
EDROA	0.0496	0	0	0	0.2171
EDIV	0.0553	0	0	0	0.2285
EDISTRESS	0.1573	0	0	0	0.3641
EDEBT	0.2436	0	0	0	0.4293
ESIZE	0.2163	0	0	0	0.4117
ECYCLE	0.0273	0	0	0	0.1631
EAUDIT	0.4453	0	0	1	0.4970
EBLOAT	0.2159	0	0	0	0.4115
ECAP	0.2157	0	0	0	0.4113
EBT	0.2149	0	0	0	0.4108
ESCORE	2.7313	1	2	4	1.7346
BHRR[a]	0.0704	−0.2843	0.0114	0.3245	0.5467
BHSAR[a]	0.0079	−0.3003	−0.0420	0.2297	0.5047
BHAR4F[a]	−0.0096	−0.2968	−0.0568	0.2020	0.4976

The table reports the mean, 25th, 50th (the median), 75th percentiles and standard deviation of selected variables. Definitions of variables are in the Appendix

Table 4 Correlations

	ESEO	EDDEBT	EMA	EOV	EROA	EDROA	EDIV	EDIS-TRESS	EDEBT	ESIZE	ECYCLE	EAUDIT	EBLOAT	ECAP	EBT	ESCORE
BHRR[a]	−0.123	−0.043	−0.069	−0.021	0.016	0.024	0.026	−0.055	−0.016	−0.014	−0.033	−0.056	−0.010	−0.031	−0.077	−0.112
BHSAR[a]	−0.097	−0.043	−0.052	−0.025	0.008	0.019	0.021	−0.062	−0.004	−0.006	−0.030	−0.022	−0.010	−0.028	−0.074	−0.092
BHAR4F[a]	−0.098	−0.028	−0.046	−0.011	0.000	0.023	0.013	−0.060	−0.011	−0.012	−0.029	−0.036	−0.009	−0.025	−0.076	−0.092
ESEO	1.000															
EDDEBT	0.078	1.000														
EMA	0.218	0.103	1.000													
EOV	0.053	0.013	0.034	1.000												
EROA	−0.025	−0.008	−0.014	−0.043	1.000											
EDROA	−0.049	0.022	−0.017	−0.026	0.021	1.000										
EDIV	−0.048	−0.011	−0.010	−0.056	0.333	0.024	1.000									
EDIS-TRESS	0.134	−0.042	−0.007	0.035	−0.049	−0.055	−0.058	1.000								
EDEBT	−0.003	−0.205	0.008	0.067	−0.043	−0.019	−0.044	0.041	1.000							
ESIZE	0.101	−0.056	0.001	−0.117	0.024	−0.026	0.000	0.153	0.058	1.000						
ECYCLE	0.101	0.029	0.009	−0.003	−0.012	−0.019	−0.023	0.117	−0.003	0.072	1.000					
EAUDIT	0.108	−0.030	0.021	−0.003	−0.017	−0.015	−0.037	0.061	0.089	0.223	0.029	1.000				
EBLOAT	0.025	−0.071	0.026	0.224	−0.015	−0.035	−0.035	0.119	0.233	0.023	0.010	0.020	1.000			
ECAP	0.089	−0.025	0.031	0.053	−0.003	−0.027	−0.005	0.096	0.118	0.047	0.023	0.064	0.173	1.000		
EBT	0.148	−0.056	0.011	0.007	−0.071	−0.069	−0.103	0.381	0.082	0.149	0.108	0.100	0.063	0.126	1.000	
ESCORE	0.435	0.200	0.232	0.304	0.089	0.057	0.071	0.434	0.344	0.386	0.202	0.431	0.420	0.415	0.448	1.000

The table reports Pearson correlation coefficients between selected variables. Definitions of variables are in the Appendix. Values reported in *italic* indicate the corresponding coefficients are *not significant at 5% level*

Table 5 Measures of accruals management and real earnings management across ESCORE groups

Panel A: Accruals management

ESCORE	DAC			DWAC		
	ADAC	HDAC		ADWAC	HDWAC	
0	0.0528	0.1334		0.0389	0.1092	
1	0.0580	0.1479		0.0466	0.1695	
2	0.0692	0.2051		0.0549	0.1996	
3	0.0859	0.2373		0.0658	0.2294	
4	0.0995	0.2591		0.0738	0.2575	
5	0.1198	0.2807		0.0870	0.2981	
6	0.1506	0.2832		0.0971	0.2807	
7	0.1657	0.3017		0.1116	0.3030	
8	0.1639	0.3636		0.1153	0.3256	
9	0.1935	0.3846		0.1105	0.2800	
Low (0)	0.0528	0.1334		0.0389	0.1092	
High (6–9)	0.1573	0.2994		0.1032	0.2912	
High—Low	0.1045	0.1660		0.0643	0.1821	
t-stat	20.007***	8.549***		17.779***	9.665***	

Table 5 (continued)

Panel B: Real earnings management

ESCORE	DCF		DPROD		DDISEXP		TOTALRM	
	ADCF	HDCF	ADPROD	HDPROD	ADDISEXP	HDDISEXP	ATOTALRM	HTOTALRM
0	0.0951	0.0841	0.1427	0.1606	0.1537	0.1722	0.3672	0.1252
1	0.0967	0.1075	0.1608	0.1765	0.1672	0.1916	0.4426	0.1727
2	0.1033	0.1646	0.1769	0.1971	0.1800	0.2117	0.4663	0.1956
3	0.1215	0.2202	0.1876	0.2321	0.2031	0.2117	0.4863	0.2238
4	0.1464	0.2753	0.1988	0.2381	0.2338	0.2290	0.4825	0.2313
5	0.1872	0.3702	0.2084	0.2763	0.2702	0.2848	0.5400	0.3309
6	0.2328	0.4386	0.2255	0.2933	0.3088	0.2444	0.5527	0.3423
7	0.2790	0.4787	0.2285	0.3086	0.3694	0.2450	0.6438	0.3473
8	0.2930	0.5750	0.2388	0.2958	0.3280	0.3286	0.6461	0.3279
9	0.3745	0.6364	0.2853	0.3684	0.4112	0.1500	0.8847	0.3889
Low (0)	0.0951	0.0841	0.1427	0.1606	0.1537	0.1722	0.3672	0.1252
High (6–9)	0.2558	0.4694	0.2296	0.3000	0.3311	0.2504	0.5991	0.3437
High—Low	0.1607	0.3854	0.0869	0.1394	0.1774	0.0781	0.2319	0.2185
t-stat	14.495***	18.713***	7.823***	6.205***	11.425***	3.5***	8.283***	9.181***

Panel A (B) reports the mean of ADAC, HDAC, ADWAC, HDWAC (ADCF, HDCF, ADPROD, HDPROD, ADDISEXP, HDDISEXP, ATOTALRM, HTOTALRM) in each group sorted by ESCORE. Low(0) and High (6–9) represent firms with ESCORE of 0 and firms with ESCORE of 6 to 9, respectively. High-Low is the difference between High (6–9) and Low(0) ESCORE firms. Definitions of variables are in the Appendix. *, **, *** indicate significance at 10%, 5%, 1% levels, respectively

5.2.2 Multivariate regression

The univariate analysis discussed above has shown that ESCORE is able to capture both accruals and real EM. However, the univariate analysis suffers from possible problems of omitted variables. Particularly, in selecting the individual signals to include in the ESCORE, we deliberately select only those which could be easily constructed using financial statement information. Hence, some dimensions of the context of earnings management may have been omitted, most notably compensation and corporate governance. It has been shown that larger and more independent boards, and especially the audit committees, play a more effective monitoring role and hence constrain earnings management (e.g. Beasley 1996; Dechow et al. 1996; Klein 2002; Xie et al. 2003; Peasnell et al. 2005). With regards to compensation, the existing evidence generally suggests that where managers' compensation package is linked to performance, they would have stronger motives to inflate earnings (e.g. Dechow and Sloan 1991; Holthausen et al. 1995; Guidry et al. 1999; Bergstresser and Philippon 2006; Burns and Kedia 2006; Efendi et al. 2007). It is important to determine if ESCORE is still related to other measures of earnings management after controlling for these omitted variables and the incremental magnitude of such relationship. In this section, we control for these omitted variables by considering the size of the board, audit committees, the independence of the boards, and the performance-linked components of executives' compensation packages. In particular, we first estimate the following regressions:

$$
\begin{aligned}
AEM_{i,t} = {} & \alpha + \beta_1 BOSIZE_{i,t} + \beta_2 BOIND_{i,t} + \beta_3 AUSIZE_{i,t} \\
& + \beta_4 DUALITY_{i,t} + \beta_5 PLCOM_{i,t} + \beta_6 ESCORE_{i,t} \\
& + YearFixedEffects + IndustryFixedEffects + \varepsilon
\end{aligned}
\tag{5}
$$

where AEM is replaced in each regression by ADAC, ADWAC, ADCF, ADPROD, ADDISEXP, ATOTALRM; BOSIZE is the number of board directors; BOIND is the percentage of non-executive directors on board to proxy for board independence; AUSIZE is the number of directors on the audit committee (equals to zero if a firm does not have an audit committee); DUALITY is a dummy which is zero if a firm's CEO is also the chairman; PLCOM is the average performance-linked compensation of all executive directors scaled by sales, where performance-linked compensation is defined as the total of bonus, shares, options and other long-term incentive pay awarded during the year.

In a similar fashion, we also estimate the following logistic regressions of the indicators of aggressive earnings management on ESCORE and the above-mentioned control variables:

$$
\begin{aligned}
Logit(HEM_{i,t}) = {} & \alpha + \beta_1 BOSIZE_{i,t} + \beta_2 BOIND_{i,t} + \beta_3 AUSIZE_{i,t} + \beta_4 DUALITY_{i,t} \\
& + \beta_5 PLCOM_{i,t} + \beta_6 ESCORE_{i,t} + YearFixedEffects \\
& + IndustryFixedEffects + \varepsilon
\end{aligned}
\tag{6}
$$

where HEM is replaced in each regression by HDAC, HDWAC, HDCF, HDPROD, HDDISEXP, and HTOTALRM.

The following procedure is followed to prepare the sample for the above multivariate regressions, which is a subsample of the main sample. First, we restrict the sample to the period from 2005 to 2011 only because going further backwards would make manual collection of data on compensation and corporate governance very difficult as firm's annual reports are no longer available online. Second, for all firm-years which remain in the

main sample, we manually acquire data on corporate governance and compensation (as described above) from Bloomberg. Third, those firm years which do not have the additional data from Bloomberg, we retrieve their annual reports from Key Note platform and manually collect the relevant data. Finally, those which still have missing data after the above steps, are excluded from the subsample for multivariate regression. This procedure yields a subsample of 2059 observations, smaller than the main sample due to the constraint over availability of corporate governance and compensation data, but still large enough for statistical inferences. All continuous variables are winsorised at the 1st and 99th percentiles to mitigate the influence of outliers.

Table 6 presents the results of the multivariate regression test. The control variables generally have the predicted signs, i.e. measures of earnings management are negatively (positively) related to BOSIZE, BOIND and AUSIZE (DUALITY and PLCOM, respectively). The main focus is on ESCORE, which is shown to be significantly positively related to all measures of earnings management. After controlling for compensation and corporate governance, one unit increase of ESCORE results in an increase of 1.24% (2.54%) in ADAC (ATOTALRM), which is statistically significant at 1% level. Similar conclusions about the positive relationship between ESCORE and the indicators of aggressive earnings management could be drawn by looking at the results reported in Panel B of Table 6 with the coefficient on ESCORE being positive and significant across all regressions (except only for the coefficient of HDDISEXP on ESCORE). The evidence reinforces our earlier conclusion that ESCORE, although estimated differently from those of Beneish (1997, 1999) and Dechow et al. (2011), is still consistent with other traditional measures of earnings management.

5.2.3 Other measures of earnings management

We have so far shown that ESCORE is consistent with other traditional measures of earnings management, including discretionary accruals and real earnings management proxies. Although these measures are the most popular ones in the earnings management literature, they are increasingly being subjected to criticism. For example, Dechow et al. (2010, p. 348) observe that 'the majority of the studies… are about the determinants and consequences of *abnormal accruals* derived from accrual models, with the idea that abnormal accruals, whether they represent errors or bias, erode decision usefulness'. In other words, the literature over-relies on models, such as the accruals models, to disentangle the component of earnings subject to managers' discretion from the 'normal' level of performance without fully appreciating that discretionary accrual is a 'noisy' measure of earnings management (for example, Holthausen et al., 1995; Fields et al., 2001; Ball, 2013; Owens et al., 2017). With the lack of a comprehensive theory on the accrual generating process (i.e. what accrual would be if there is no manipulation), as a profession we are using (allegedly) mis-specified models trying to measure the 'immeasurable' (McNichols, 2000; Dechow et al., 2010; Owens et al., 2017). In addition, some researchers raise a concern about the implausibly large magnitude and high frequency of earnings management documented in the extant literature using accruals models (Ball, 2013; Gerakos and Kovrijnykh, 2013). Ball (2013) 'worries' that the current practice that considers a positive (negative) discretionary accrual seems to create 'the most incorrect belief' that earnings management is 'rife' because technically 'no observation sits exactly on the regression line'. Walker (2013) describes the existing statistical approaches of detecting earnings management as 'good for rejecting a null that nobody believes is true'.

Table 6 Measures of accruals management and real earnings management regressed on ESCORE and control variables

Panel A: OLS regressions of absolute values of earnings management proxies on ESCORE and control variables

Variables	ADAC Coef	t-stat	ADWAC Coef	t-stat	ADCF Coef	t-stat	ADPROD Coef	t-stat	ADDISEXP Coef	t-stat	ATOTALRM Coef	t-stat
INTERCEPT	0.0957	6.051***	0.077	6.718***	0.1966	4.889***	0.0695	2.088**	0.2566	6.522***	0.4424	6.296***
BOSIZE	-0.0038	-3.471***	-0.0002	-0.296	-0.0011	-0.401	0.0021	0.9	0.001	0.361	0.0076	1.571
BOIND	-0.0203	-1.418	-0.0286	-2.765***	-0.0262	-0.72	-0.0294	-0.975	-0.0818	-2.303**	-0.1321	-2.081**
AUSIZE	0.0002	0.097	-0.0013	-1.078	0.0009	0.215	0.0016	0.451	-0.0054	-1.283	0.0013	0.168
DUALITY	0.0175	2.246**	0.0081	1.433	-0.0014	-0.071	0.004	0.242	-0.0022	-0.114	-0.0336	-0.972
PLCOM	0.0536	2.261**	0.0065	0.38	0.032	0.53	-0.0579	-1.158	0.0457	0.775	-0.0545	-0.517
ESCORE	0.0124	10.432***	0.0083	9.663***	0.0129	4.266***	0.0131	5.217***	0.0188	6.357***	0.0254	4.818***
Adj. R^2	0.0513		0.0834		0.1203		0.0328		0.0603		0.1032	

Panel B: Logistic regressions of indicators of aggressive earnings management on ESCORE and control variables

Variables	HDAC Coef	z-stat	HDWAC Coef	z-stat	HDCF Coef	z-stat	HDPROD Coef	z-stat	HDDISEXP Coef	z-stat	HTOTALRM Coef	z-stat
INTERCEPT	-0.8654	3.355*	-1.4544	9.57***	-1.9348	14.909***	-0.0892	0.042	0.3289	0.578	-0.5196	1.42
BOSIZE	-0.053	2.405	-0.0273	0.707	-0.1	6.976***	-0.0613	2.975*	-0.0803	5.552**	-0.0887	6.686***
BOIND	-0.7611	3.334*	-0.4054	0.996	0.2024	0.21	-0.9002	4.285**	-0.9664	5.566**	-0.3435	0.684
AUSIZE	-0.0729	1.906	-0.0338	0.461	-0.0209	0.147	-0.0086	0.027	-0.0802	2.399	0.0053	0.011
DUALITY	0.2992	2.031	0.1681	0.632	0.1355	0.332	-0.0834	0.117	0.1957	0.857	-0.1719	0.533
PLCOM	-0.9954	1.851	0.1426	0.052	-0.1104	0.031	0.4305	0.491	0.0298	0.002	0.3981	0.433
ESCORE	0.1839	30.19***	0.1578	22.991***	0.3696	104.954***	0.1177	11.321***	0.0259	0.59	0.1318	15.638***
Adj. R^2	0.0671		0.0453		0.1067		0.0962		0.0630		0.1102	

Panel A (B) reports the results of OLS (logistic) regressions of ADAC, ADWAC, ADCF, ADPROD, ADDISEXP, ATOTALRM (HDAC, HDWAC, HDCF, HDPROD, HDDISEXP, HTOTALRM) on ESCORE and control variables. Year and industry fixed effects are included in all regressions. Definitions of variables are in the Appendix. *, **, *** indicate significance at 10%, 5%, 1% levels, respectively

Given these criticisms, a valid concern would be that our evidence on the relationship between ESCORE and discretionary accruals and real earnings management proxies might be attributable to the mis-specification and measurement errors of the established models rather than a reflection of the association of ESCORE with actual earnings management. We address this concern and further test the effectiveness of ESCORE in capturing the context of earnings management by looking at how ESCORE is associated with two other measures of earnings management in the next two sub-sections.

Ex-post measure of earnings management

We first employ an *ex-post* measure of earnings management. In the UK, FRRP is responsible for ensuring financial statements of public companies, the main input to our ESCORE model, comply with applicable laws and financial reporting standards. FRRP selects firms for review based on some published criteria, including firms from specific sectors in the economy which are under particular stress, firms involved with special accounting issues which give rise to judgement, subjectivity and risk of misstatements as well as from complaints from the public, press or the accounting and financial community. As such, similar to the AAER and GAO samples of restatements in the US, the FRRP sample too is not free from selection bias. However, as each institution has a different sampling scheme, the evidence could reinforce each other and the limitations of each source could be mitigated.

If a firm is selected by FRRP for review, several steps are taken, including an initial review, formal and informal discussions before a Review Group being set up if necessary, then a thorough investigation followed by a recommendation to the FRRP chairman. A review may investigate one or more annual reports of the selected firm. At the end of the process, FRRP might decide whether it is suitable for a press notice or not. It is most likely that a press notice is issued in case the directors have agreed that the financial statements are defective and proposed corrective actions have been taken and that FRRP is satisfied with those actions.

From the above description, we define firm-years which are investigated by FRRP followed by a press notice as instances of earnings management. As shown in Panel A of Table 7, there are 70 annual reports with fiscal year ending between 1/1/1995 and 31/12/2012 which are subjected to FRRP press notices. We remove 37 firm-years which are in the financial and utility industries and do not have sufficient data to calculate ESCORE. The 33 remaining cases spread across 22 Datastream level-six industries.

If our ESCORE does capture the context of earnings management, we would expect to see ESCORE of the 33 firms, being investigated by FRRP and subsequently having a press notice (FRRP firms henceforth), is significantly larger in the year subjected to the investigation compared to other years. To test this conjecture, we extract ESCORE (calculated using the whole sample as described in Sect. 3) of 33 FRRP firms for the period from 1995 to 2012 to create a subsample of 576 firm-year observations. As shown in Panel B of Table 7, the size of FRRP firms is generally larger compared to average firms in the main sample (see Table 3), e.g. mean MVE of FRRP firms is £2708 million compared to £390 million in the main sample. It suggests the FRRP's sampling method is quite biased towards larger firms, which typically play an important role in the economy and if a misstatement exists, it would have more pronounced effects on investors. The mean (standard deviation) of ESCORE in this subsample is 2.17 (1.47). We define the year for which the annual reports are investigated by FRRP as restatement year. Panel C of Table 7 shows that the mean ESCORE of FRRP firms in restatement years (3.24) is significantly larger than that of the rest of the sample (2.10) at 1% level. The magnitude of the difference (1.14) is

Table 7 ESCORE of FRRP firms

Panel A: the FRRP sample selection

FRRP press notices related to annual reports with fiscal year ended between 1/1/1995 and 31/12/2012	70
Less: Financial and utility firms and those with insufficient data to calculate ESCORE	37
Final sample for FRRP restatement	33

Panel B: Descriptive statistics ($n = 576$ firm-year observations)

	Mean	25th percentile	Median	75th percentile	Std. Dev
AT	2230	39	118	508	8895
SALE	1404	36	133	702	4298
IB	152	0	4	29	706
BVE	899	17	56	162	3748
MVE	2708	20	105	419	9846
ESCORE	2.1667	1.0000	2.0000	3.0000	1.4685

Panel C: Mean ESCORE in FRRP years versus the rest of the sample

	Restated firm-years	Rest of sample	Difference	t-statistic
FRRP firm-years ($n = 33$) versus the rest of the sample ($n = 543$)	3.2424	2.1013	1.1411	4.4***
FRRP firm-years with IS code ($n = 12$) versus the rest of the sample ($n = 564$)	3.2	2.1464	1.0536	2.26**

Panel D: Logistic regression of the indicator of FRRP restatements on ESCORE

ESCORE coefficient	0.4521	0.532
z-stat	4.1***	3.85***
Marginal effect	0.0232	0.0202
Pseudo R^2	6.57%	13.37%
Year fixed effects	No	Included
Industry fixed effects	No	Included
Adj. R^2	0.0232	0.0432

Panel A reports the FRRPsample selection. Panel B presents descriptive statistics of the FRRP sample, including 33 FRRP firms across 1995 to 2012 where data is sufficient to calculate ESCORE. Panel C compares the mean ESCORE of FRRP firms with that of the rest of the sample using a two-tailed t-test. Panel D reports the results of the logistic regression of the indicator of FRRP restatement (equals to one if the firm-year is subjected to a FRRP restatement, zero otherwise) on a constant and ESCORE. Definitions of variables are in the Appendix. *, **, *** indicate significance at 10%, 5%, 1% levels, respectively

also large, considering that the standard deviation of ESCORE in the subsample is only 1.47 and that of the main sample (see Table 3) is 1.73.

Ideally, we would have included ESCORE together with some control variables which are potentially related to restatements but not included in ESCORE, such as corporate governance and compensation variables, to the right hand side of a logistic regression where the indicator of restatements is on the left hand side. However, further constraining the sample requiring the availability of corporate governance and compensation data would mean that the sample is too small for any reliable statistical inferences. Instead, we estimate a logistic regression of the indicator of restatement (equals to one for firm-years

which are investigated by FRRP followed by a release of press notice, zero otherwise) on ESCORE only and with year and industry fixed effects included. The coefficient on ESCORE, as shown in Panel D of Table 7, is 0.5320 (significant at 1% level). In terms of economic significance, one unit increase of ESCORE would increase the probability of a restatement by 2.02%. Compared to the unconditional probability of 5.73% (33/576), the effect of ESCORE on the likelihood of restatement is economically large. Overall, the evidence suggests that firms which are required to restate their financial statements, especially the income statements, generally have higher ESCORE. It further supports our claim that ESCORE captures the context of earnings management.

The 33 selected cases analysed above might involve different types of restatements which result in different effects on financial statements. For each case, we read through FRRP press notice to determine the nature of the cases and the effects on the firm's financial statements. We apply the following codes to classify the effects on the firm's financial statements: IS (if it involves a restatement which affects the income statement) and OT (if the restatement does not affect the income statement). Of the 33 cases selected, FRRP requires restatements from only 12 cases (36%), involving items on the income statement (IS code). Unreported results show that the mean ESCORE of FRRP firms with IS code is 3.20, which is 1.05 unit higher than that of the rest of the sample (significant at 5% level).

Measure of earnings management based on Benford's law

The ex-post measure of earnings management as used in the previous section has the advantage of having low Type I error. However, as FRRP does not select firms for investigation randomly, it suffers from the issue of generalization as discussed by Dechow et al. (2010). We, therefore, reinforce our evidence of the association between ESCORE and the context of earnings management by employing one more measure of earnings management which does not have the same pitfalls as the *ex-post* indicator of earnings restatement. In this section, we use an empirical measure of earnings management which is constructed based on the Benford's law. Benford's law refers to the observation that many real-life numerical datasets, accounting figures included, are distributed in such a way that the first digits are likely to be small. The probability for the first digit of an accounting item (e.g. an asset or liability reported on the balance sheet or an income or expense on the income statement) being one, according to Benford's law, is highest (30.10%), followed by the probability for it being two (17.61%), three (12.49), four (9.69%), five (7.92%), six (6.70%), seven (5.80%), eight (5.12%) and nine (4.58%). If a firm, in its financial statements, reports accounting figures which are not in conformity to this distribution, then that is 'abnormal'. Amiram et al. (2015) shows that the more the reported figures on financial statements deviate from the distribution expected by Benford's law, the more likely the financial statements actually contain errors, regardless of whether it is a pure unintentional error or a deliberate earnings management attempt. The advantage of this approach of capturing earnings management is that it does not suffer from any statistical biases as experienced by the traditional models of discretionary accruals, nor does it rely on the problematic endogenous relationship between earnings management and firm's fundamental characteristics.

We, therefore, follow Amiram et al. (2015) to estimate an empirical measure of the deviation of the distribution of the first digits of financial statement items from the distribution expected by Benford's law as follows. For this test, we collect additional data from Bloomberg. As a result, we have to further constrain the main sample to include only firms which are still listed on London Stock Exchange first as at the end of May 2017. From this

Table 8 FSD_SCORE regressed on ESCORE and control variables ($n = 2373$)

Variables	Coefficient	t-statistic
INTERCEPT	0.0352	35.74***
BOSIZE	−0.0008	−7.91***
BOIND	−0.0015	−1.09
AUSIZE	0.0002	0.12
DUALITY	0.0004	0.48
PLCOM	0.0009	2.58***
ESCORE	0.0012	9.74***
Adj. R^2	0.1025	

The table reports the results of OLS regressions of FSD_SCORE on ESCORE and control variables. Year and industry fixed effects are included. Definitions of variables are in the Appendix. *, **, *** indicate significance at 10%, 5%, 1% levels, respectively

initial list of live firms, we keep only the firm-year observations with sufficient data to calculate ESCORE. For each of those observations, we then download from Bloomberg all items on the balance sheet, income statement and cash flow statements. Missing items are replaced by zeros. In line with prior research (e.g. Amiram et al., 2015), firm-years with less than 50 items are dropped to avoid measurement errors. For each item, we keep only the first digit (ignoring the negative sign if an item has a negative value), except for items with the absolute value being less than one for which the first non-zero digit is kept.

The above process results in a sample of 2373 firm-year observations, each having a pool of at least 50 integers ranging from one to nine representing the first digits of items reported on the firm's balance sheet, income statement and cash flow statement in that year. We then follow Amiram et al. (2015) to calculate FSD_SCORE as follows:

$$FSD\_SCORE_{i,t} = \frac{\sum_{d=1}^{9} \left| AD_{d,i,t} - ED_d \right|}{9} (d = 1, 2, \dots, 9) \qquad (7)$$

where $FSD\_SCORE_{i,t}$ is the deviation of the observed distribution of financial statement figures from the distribution expected by Benford's law of firm i in year t; $AD_{d,i,t}$ ($d = 1, 2, \dots, 9$) is the actual distribution of digit d, measured as the number of times d appears as the first digit of items reported on firm i's balance sheet, income statement and cash flow statement in year t divided by the total number of items reported on those statements; ED_d is the expected distribution of digit d under Benford's law (i.e. $ED_1 = 0.3010$, $ED_2 = 0.1761$, $ED_3 = 0.1249$, $ED_4 = 0.0969$, $ED_5 = 0.0792$, $ED_6 = 0.0669$, $ED_7 = 0.0580$, $ED_8 = 0.0512$, $ED_9 = 0.0458$).

We use FSD_SCORE as a proxy for earnings management and examine how ESCORE is associated with FSD_SCORE by estimating the following regression:

$$\begin{aligned} FSD\_SCORE_{i,t} = {} & \alpha + \beta_1 BOSIZE_{i,t} + \beta_2 BOIND_{i,t} + \beta_3 AUSIZE_{i,t} \\ & + \beta_4 DUALITY_{i,t} + \beta_5 PLCOM_{i,t} + \beta_6 ESCORE_{i,t} \\ & + YearFixedEffects + IndustryFixedEffects + \varepsilon \end{aligned} \qquad (8)$$

The results are presented in Table 8. ESCORE is positively and significantly (at 1% level) related to FSD_SCORE. The evidence, therefore, suggests that firms with higher ESCORE tend to have financial statement figures being distributed more anomalously given the expectation under Benford's law. It further validates that ESCORE is a good measure of the context of earnings management.

Table 9 Stock returns across ESCORE groups

ESCORE	BHRR^m (%)		BHSAR^m (%)		BHAR4F^m (%)	
	Returns	t-statistic	Returns	t-statistic	Returns	t-statistic
0	1.10	n/a	0.48	3.694***	0.33	2.086**
1	0.91	n/a	0.32	4.041***	0.18	1.553
2	0.80	n/a	0.23	3.678***	0.08	0.636
3	0.55	n/a	0.01	0.134	−0.16	−1.243
4	0.20	n/a	−0.27	−2.709***	−0.46	−2.752***
5	−0.19	n/a	−0.61	−4.178***	−0.80	−3.483***
6	−0.35	n/a	−0.72	−3.146***	−1.09	−3.586***
7	−0.11	n/a	−0.52	−1.512	−0.80	−2.075**
8	−2.02	n/a	−2.31	−3.562***	−2.49	−3.565***
9	−0.84	n/a	−0.83	−0.883	−1.56	−1.534
Low (0)	1.10	n/a	0.48	3.694***	0.33	2.086**
Medium (1–5)	0.55	n/a	0.02	1.121	−0.14	−1.239
High (6–9)	−0.31	n/a	−0.69	−3.662***	−1.04	−3.923***
Low—High	1.41	5.156***	1.17	4.584***	1.37	5.102***

The table reports the returns on different portfolios formed on the basis of ESCORE. Low(0), Medium (1–5), and High (6–9) represent firms with ESCORE of 0, ESCORE of 1 to 5, and ESCORE of 6 to 9, respectively. Low-High is the difference between Low (0) and High (6–9) firms. Definitions of variables are in the Appendix. *, **, *** indicate significance at 10%, 5%, 1% levels, respectively

5.3 ESCORE and stock returns

We have established that ESCORE is able to capture the context of earnings management. The next question is if investors would misprice the context of earnings management. Table 9 reports the buy-and-hold returns on each ESCORE portfolio (0–9), the low, medium and high ESCORE portfolios as well as the hedge portfolio. The t-statistics are reported under the null hypothesis that the corresponding return is zero. The results are easy to summarise. First, as ESCORE increases, all measures of stock returns decrease monotonically. Secondly, low ESCORE stocks earn abnormally high and high ESCORE stocks earn abnormally low returns. Third, the hedge portfolio earns positive abnormal returns.

Since the results are consistent across different return metrics, we discuss the results based only on the abnormal returns estimated using the four-factor model. The portfolio of stocks with ESCORE of zero earns an abnormal return of 0.33% per month (significant at 5% level). As ESCORE increases, abnormal returns decrease monotonically. The High ESCORE portfolio (includes all stocks with ESCORE of six or higher) earns an abnormal return of −1.04% per month (significant at 1% level). The hedge portfolio that takes long position in Low ESCORE stocks and short position in High ESCORE stocks earns 1.37% abnormal return per month. To put the results in perspective, we compare our findings to other similar return anomalies documented in the literature. For example, Sloan (1996) documents an annual size-adjusted return of 10.4% on a hedge portfolio which takes long position in stocks with low and short position in those with high accruals. Soares and Stark (2009) provide similar results showing that the accruals anomaly exists in the UK with the hedge portfolio earning an abnormal return (adjusted for size and book-to-market factors

but without controlling for transaction costs) of 18.7% per year. The annualised return on the hedge portfolio based on our ESCORE is 17.74% (1.0137^12 – 1), which is non-trivial in economic terms. Overall, the result cannot reject our hypothesis suggesting that the market misprices the information contained in the ESCORE, which is designed to capture the context of earnings management.

5.4 Is there another 'market anomaly' in disguise?

The results from the portfolio analysis strongly suggest that ESCORE could predict future stock returns. However, there may be some other known 'market anomalies' associated with ESCORE which could partly explain the predictive power of ESCORE. This section addresses such concerns.

To see if ESCORE is indeed related to other known patterns in realised returns, Table 10 presents fundamental characteristics of stocks across ESCORE groups. Firm size, measured by either total assets (AT) or market capitalization (MVE), is inversely related with ESCORE. Firms with higher ESCORE are also more likely to issue seasoned equity and debt and have lower NOA. High ESCORE firms are also highly valued by the market evidenced by the monotonic increase of the market-to-book ratio across the ESCORE groups. The decrease in ROA and DROA as ESCORE increases also suggests that high ESCORE stocks are typically less profitable. High ESCORE stocks are also more financially distressed as measured by the ZSCORE.

The above patterns raise a concern whether ESCORE could predict future returns beyond the known return effects embedded in it. To start with, we construct ESCORE based on prior literature on earnings management, and not from market anomalies' literature. Therefore, the signals embedded in ESCORE do not necessarily include only those factors which are known as stock return predictors. We argue that the predictive power of ESCORE comes from the context of earnings management which is revealed *collectively* by the composite ESCORE, and not by the predictive power of the individual signals separately. In fact, the established literature even suggests that some signals, including ESIZE and EBLOAT, would predict future returns in the opposite direction. Particularly, based on the established evidence of the size effect (e.g. Banz 1981) and the irrational market reaction to balance sheet bloat (e.g. Hirshleifer et al. 2004; Gray et al. 2018), stocks with ESIZE and EBLOAT of one (smaller stocks and those which have smaller NOA) are expected to earn higher (not lower) future returns. Meanwhile, the literature is silent about whether other signals, namely EROA, EDROA, EDIV, EDEBT, EDDEBT, EMA, ECYCLE, EAUDIT and EBT, could predict future returns or not. The concern lies, therefore, mainly with the high market-to-book ratio, high likelihood of issuing seasoned equity, more financial distress and low profitability of high ESCORE stocks. Prior research widely documents that abnormally low returns are associated with high market-to-book firms (e.g. Fama and French 1992; Lakonishok et al. 1994), seasoned equity offers (e.g. Loughran and Ritter 1995; Spiess and Affleck-Graves 1995; Chen et al. 2019), firms with negative ZSCORE (e.g. Agarwal and Taffler 2008), and firms with lower profitability (e.g. Ou and Penman 1989a, b; Piotroski 2000, Fama and French 2006). These known patterns of returns are embedded in ESCORE through EOV, ESEO and EDISTRESS. In addition, as ESCORE is designed to capture the context of

Table 10 Fundamental characteristics across ESCORE groups

	ESCORE									
	0	1	2	3	4	5	6	7	8	9
AT (£mil)	733	615	458	348	264	130	92	41	33	11
SALE (£mil)	703	611	472	362	274	140	97	65	67	9
NI (£mil)	41	32	23	16	10	2	−1	−1	−3	−1
DIV (£mil)	18	15	12	9	6	2	1	1	1	0
MVE (£mil)	689	576	443	349	270	134	94	61	25	17
DSHARE	0.0054	0.0224	0.0445	0.0904	0.1763	0.2677	0.3677	0.4911	0.6565	0.9111
DDEBT	−0.2765	0.1760	0.5336	1.6761	2.3236	3.6770	4.2057	5.4630	7.7978	8.8561
MTB	2.2047	2.3857	2.8669	3.2708	4.1842	4.3591	5.2342	6.4006	6.0342	5.9664
ROA	0.0589	0.0552	0.0476	0.0169	−0.0360	−0.1153	−0.2325	−0.3791	−0.4571	−0.5874
DROA	0.0335	0.0219	0.0185	0.0151	0.0074	−0.0018	−0.0071	−0.0346	−0.0026	−0.1953
DIVDEF	0.0343	0.0286	0.0208	−0.0086	−0.0598	−0.1320	−0.2451	−0.3863	−0.4589	−0.5908
ZSCORE	17.9433	16.6083	15.9328	13.6591	9.5750	7.0799	−1.1646	−4.3763	−8.2776	−12.5474
DEBT	0.1668	0.1755	0.1676	0.1572	0.1439	0.1289	0.1225	0.1038	0.1010	0.0760
NOA	0.5701	0.5603	0.5314	0.4906	0.4458	0.4340	0.3992	0.3501	0.3683	0.3702
CAP	0.5874	0.5562	0.5030	0.4316	0.3667	0.3127	0.2840	0.2453	0.2187	0.1969
BOOKTAX	0.0587	0.0831	0.1952	0.4172	1.0355	2.8103	3.8965	4.3297	4.0088	4.5332
DAC	0.0070	0.0054	0.0127	0.0085	0.0037	0.0079	−0.0101	−0.0239	−0.0059	0.0125
ESEO	0.0000	0.0437	0.1084	0.2230	0.3290	0.4608	0.5934	0.6810	0.7614	0.9615
EDDEBT	0.0000	0.2674	0.4017	0.4746	0.4412	0.4718	0.4721	0.4526	0.5568	0.5000
EMA	0.0000	0.0063	0.0191	0.0470	0.0818	0.1207	0.1503	0.1681	0.2841	0.5000
ECYCLE	0.0000	0.0032	0.0062	0.0189	0.0400	0.0915	0.0886	0.1293	0.1591	0.3077
EAUDIT	0.0000	0.2029	0.3921	0.5254	0.6209	0.6982	0.8054	0.8491	0.9773	0.9615

The table reports the mean of selected variables across groups sorted by ESCORE. Definitions of variables are in the Appendix

earnings management, it is also important to control for the documented market mispricing of discretionary accruals (Xie 2001).

While previous studies (such as Jiang et al. 2016) demonstrate a number of other risk factors, we primarily focus only on factors which might have already been embedded in ESCORE as identified from the preceding analyses. To demonstrate that ESCORE is still significantly associated with future returns after controlling for the above five anomalies, we estimate the following regression using Fama–MacBeth methodology with the t-statistics calculated using the Newey-West corrected standard errors[14]:

$$
\begin{aligned}
RET^a_{i,t+1} =& \alpha + \beta_1 Ln(MVE_{i,t}) + \beta_2 MTB_{i,t} + \beta_3 ROA_{i,t} + \beta_4 ESEO_{i,t} \\
& + \beta_5 EDISTRESS_{i,t} + \beta_6 NOA_{i,t} + \beta_7 DAC_{i,t} + \gamma ESCORE_{i,t} + \varepsilon
\end{aligned}
\tag{9}
$$

where $RET^a_{i,t+1}$ is annual buy-and-hold return measured from June of year $t+1$ to May of year $t+2$ and is replaced by $BHRR^a_{i,t+1}$, $BHSAR^a_{i,t+1}$ and $BHAR4F^a_{i,t+1}$.

Table 11 presents the results of estimating Eq. (9) along with four other specifications where we exclude ESCORE and DAC one-by-one and together as well as the last specification where we only keep ESCORE and DAC as explanatory variables. Each panel reports the results of a return metric.

In Table 11, all control variables have the predicted signs. DAC is always negative and significant, which is in line with the existing literature (e.g. Xie 2001). The focus of the paper is the coefficient on ESCORE, which is always negative and significant in all specifications. We, therefore, argue that ESCORE can predict stock returns beyond the existing anomalies. From specification (4) in Panel C of Table 11, one unit increase in ESCORE pulls annual four-factor risk-adjusted returns down by 1.40%. As a comparison with the portfolio analysis where we do not control for other market 'anomalies', the annualised buy-and-hold four-factor risk-adjusted returns of the hedge portfolio reported in Table 9 is 17.74% (1.0137^12 – 1). The average ESCORE of the low ESCORE portfolio is 0 and that of the high ESCORE portfolio is 6.56 [(519×6 + 232×7 + 88×8 + 26×9) / 879], yielding a difference of –6.56. Therefore, after adjusting for other known market anomalies the four-factor risk-adjusted returns on the hedge portfolio shrink from 17.74% to 9.18% per year (1.40×6.56), which is still significant in economic terms.

One issue with the above multivariate regression is the correlation between the control variables and ESCORE, as highlighted in Table 10. We respond to this issue in two ways. First, we drop the control variables in Eq. (9) one at a time, one pair at a time, and all together. For brevity, we only report the result when all control variables are dropped (specification (5) in Table 11). In all of those specifications, the main conclusions of the paper remain unchanged.

Another way to deal with this issue is to exclude ESEO, EDISTRESS and EOV from the construction of ESCORE. We calculate four compressed versions of ESCORE in which ESEO, EDISTRESS and EOV are dropped one by one from the construction of ESCORE, and all together. We then redo the returns analysis and perform multivariate

---

[14] We also used pooled regressions and the results (unreported) remain qualitatively similar.

Table 11 Stock returns regressed on DAC, ESCORE and control variables

Variables	Predicted sign	Specification (1)		Specification (2)		Specification (3)		Specification (4)		Specification (5)	
		Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Panel A: BHRR[a]											
INTERCEPT		0.1810	2.21**	0.2756	3.98***	0.1947	2.44**	0.2794	4.14***	0.1587	4.23***
Ln(MVE)	(−)	−0.0007	−0.11	−0.0049	−0.86	−0.0021	−0.33	−0.0059	−1.05		
MTB	(−)	−0.0032	−1.8*	−0.0027	−1.4	−0.0030	−1.69	−0.0025	−1.34		
ROA	(+)	0.1383	3.81***	0.1255	3.09***	0.1806	4.26***	0.1645	3.54***		
ESEO	(−)	−0.0884	−3.65***	−0.0605	−3.41***	−0.0831	−3.67***	−0.0584	−3.56***		
EDISTRESS	(−)	−0.0428	−1.54	−0.0211	−1.03	−0.0484	−1.71	−0.0282	−1.35		
EBLOAT	(−)	−0.1162	−3.71***	−0.1359	−4.94***	−0.1107	−3.55***	−0.1291	−4.77***		
DAC	(−)			−0.0184	−2.24**			−0.1802	−3.47***	−0.0869	−2.21**
ESCORE	(−)					−0.2013	−3.47***	−0.0166	−2.1*	−0.0294	−2.84**
Adj. R²		0.0227		0.0257		0.0244		0.0273		0.0132	
Panel B: BHSAR[a]											
INTERCEPT		0.1938	5.24***	0.2918	4.39***	0.2069	5.81***	0.2955	4.66***	0.0786	3.33***
Ln(MVE)	(−)	−0.0084	−2.69**	−0.0127	−2.84**	−0.0097	−3.14***	−0.0135	−3.12***		
MTB	(−)	−0.0031	−1.8*	−0.0025	−1.36	−0.0028	−1.69	−0.0023	−1.3		
ROA	(+)	0.1436	3.84***	0.1315	3.13***	0.1824	4.26***	0.1669	3.55***		
ESEO	(−)	−0.0838	−3.43***	−0.0548	−3.08***	−0.0788	−3.45***	−0.0529	−3.22***		
EDISTRESS	(−)	−0.0442	−1.57	−0.0215	−1.01	−0.0493	−1.72	−0.0281	−1.29		
EBLOAT	(−)	−0.1153	−3.54***	−0.1358	−4.58***	−0.1106	−3.4***	−0.1300	−4.43***		
DAC	(−)			−0.0192	−2.5**			−0.1627	−3.08***	−0.0709	−1.83*
ESCORE	(−)					−0.1846	−3.15***	−0.0175	−2.38**	−0.0251	−3.15***
Adj. R²		0.0180		0.0204		0.0199		0.0220		0.0089	
Panel C: BHAR4F[a]											
INTERCEPT		0.1012	1.42	0.1827	3.03***	0.1143	1.64	0.1858	3.17***	0.0559	2.27**
Ln(MVE)	(−)	−0.0024	−0.37	−0.0060	−1.07	−0.0038	−0.6	−0.0069	−1.23		

Table 11 (continued)

Variables	Predicted sign	Specification (1)		Specification (2)		Specification (3)		Specification (4)		Specification (5)	
		Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
MTB	(−)	−0.0015	−1.13	−0.0010	−0.7	−0.0012	−0.96	−0.0009	−0.6		
ROA	(+)	0.1178	3.87***	0.1059	3.18***	0.1635	4.48***	0.1486	3.84***		
ESEO	(−)	−0.0781	−3.17***	−0.0544	−3.31***	−0.0724	−3.13***	−0.0519	−3.38***		
EDISTRESS	(−)	−0.0538	−2.93***	−0.0350	−3.37***	−0.0590	−3.11***	−0.0419	−3.78***		
EBLOAT	(−)	−0.1144	−4.15***	−0.1312	−5.56***	−0.1085	−3.98***	−0.1240	−5.43***		
DAC	(−)					−0.2119	−3.83***	−0.1917	−3.9***	−0.1030	−2.9**
ESCORE	(−)			−0.0158	−2.11*			−0.0140	−1.97*	−0.0237	−2.36**
Adj. R²		0.0156		0.0176		0.0179		0.0195		0.0094	

Specification (4) reports the results of estimating the following equation (sample size = 11,920): $RET^a_{i,t+1} = \alpha + \beta_1 Ln(MVE_{i,t}) + \beta_2 MTB_{i,t} + \beta_3 ROA_{i,t} + \beta_4 ESEO_{i,t} + \beta_5 EDISTRESS_{i,t} + \beta_6 NOA_{i,t} + \beta_7 DAC_{i,t} + \gamma ESCORE_{i,t} + \varepsilon$. From full Eq. (4), we drop DAC and ESCORE in specification (1); DAC in specification (2); ESCORE in specification (3); and Ln(MVE), MTB, ROA, ESEO, EDISTRESS, EBLOAT in specification (5) from the explanatory variables. Panels A, B, and C report results where raw return, size-adjusted abnormal return, and four-factor abnormal return, are the dependent variables, respectively. The table reports the results from Fama–Macbeth regressions in which the standard errors are adjusted for serial correlation using Newey-West procedure. Definitions of variables are in the Appendix. *, **, *** indicate significance at 10%, 5%, 1% levels, respectively

regressions. Untabulated results confirm that none of the main results change qualitatively. The hedge portfolio, using ESCORE without ESEO, EDISTRESS and EOV, yields an average $BHSAR^m$ ($BHAR4F^m$, respectively) of 0.92% (1.01%, respectively) per month, all statistically significant at conventional levels. Using the compressed ESCORE without ESEO, EDISTRESS and EOV to estimate Eq. (9), the coefficient on ESCORE is –0.0172 (–0.0138) when $BHSAR^a$ ($BHAR4F^a$, respectively) is the dependent variable, all being statistically significant at conventional levels. We therefore conclude that the power of the ESCORE to predict future returns goes beyond the known patterns of returns related to other known market anomalies.

5.5 Generalization of the results to other markets

Although our use of the UK market as the setting is well justified and makes a significant contribution to the advancement of knowledge in this strand of research, a valid concern is whether the results of this study are generalizable to other markets, especially the US where many previous studies in this area focus on. To address this concern, we replicate some of the main analyses in the paper using a sample of US (live and dead) listed stocks during the period from 1987 to 2013. The data collection and sampling procedures are the same as used for the main tests. The US sample contains 87,645 observations. We winsorise all continuous variables at the 1st and 99th percentiles to mitigate the influence of outliers.

We first replicate the results as reported in Table 5. Untabulated results show that all measures of earnings management increase monotonically when ESCORE increases. The means of ADAC and HDAC of high-ESCORE stocks (7757 observations) are 0.1526 and 0.2858, respectively, while the corresponding means of low-ESCORE stocks (7986 observations) are 0.0670 and 0.0995. The differences are both economically large and statistically significant at 1% level. When looking at real earnings management, the means of ATOTALRM and HTOTALRM of high-ESCORE stocks (0.6219 and 3.3019, respectively) are also significantly higher than those of the low-ESCORE counterparts (at 1% level). We then conduct a replication of Table 9 using raw and market-adjusted returns. The low-ESCORE portfolio earns an annualised buy-and-hold market-adjusted return of 8.78%, which is significantly higher than that of the high ESCORE of -9.8%. The hedge portfolio which takes long position in low-ESCORE and short position in high-ESCORE stocks earn a return of 18.58% per year after controlling for the market factor, which is significant at 1% level and economically large. We also replicate Table 11 using both Fama–MacBeth and pooled regressions of raw and market-adjusted returns on ESCORE and the control variables. Again, all main conclusions from the main tests are generalizable to the US market. After controlling for size, book-to-market, profitability, seasoned equity, financial distress, balance sheet bloat and discretionary accruals, one unit increase of ESCORE pulls the annual buy-and-hold market-adjusted returns by 2.6%, which is statistically significant at 1% level. In general, while we invite future research to replicate other aspects of our analyses by pooling in more data, e.g. compensation, corporate governance, restatements in the US or in another market, a restricted replication as reported in this section provides some reassurance that our main findings are generalizable to other developed markets, especially the US.

6 Conclusions

This study demonstrates that an index, named as ESCORE which accumulates 15 individual financial-statement-based signals, can capture the context of earnings management and reliably predict future stock returns. We show that ESCORE is related to both accruals-based and real earnings management measured using various ways which are popularly used in the extant literature. Firms which are required to restate their financial statements by the Financial Reporting Review Panel in the UK also exhibit a significantly higher ESCORE in the restatement year. Higher ESCORE firms also have the distribution of financial statement figures being more divergent from what is expected under Benford's law. Having established that ESCORE is effective in capturing the context of earnings management, we use ESCORE to form portfolios and find that after adjusting for market, size, book-to-market and momentum factors, low ESCORE stocks outperform high ESCORE stocks by 1.37% per month, which is both statistically and economically significant. We also report that the main findings of the paper are generalizable to the US market.

The paper makes several contributions to the literature. First, ESCORE could be used as an alternative empirical proxy for earnings management. The appeal of ESCORE is that it allows financial statement users to assess the reliability of reported earnings by looking at the surrounding context rather than the magnitude of the actual earnings and its components. In addition, ESCORE captures both accruals-based and real earnings management as well as financial reporting violations which require restatements. The ESCORE is, hence, particularly advantageous to use in subsequent studies as it only focuses on earnings management and makes no prediction regarding which methods have been used to manage earnings.

The second and more important contribution of the paper is that ESCORE can be applied by investors to screen out the information about the context of earnings management which is mispriced by the market, and hence earn (avoid) economically large abnormal returns (losses). We also add to the literature on fundamentals-based strategies by showing that the dimension of financial statement information which suggests earnings management exists is also mispriced. ESCORE has undeniably not been designed to capture *all* signals of suspicious earnings management. We deliberately focus on the context which could be easily extracted from annual financial statements, hence the exclusion of areas such as performance-linked compensation, institutional holding, and corporate governance. The reason is twofold. First, we propose a parsimonious model which covers a broad range of signals for which data can be easily obtained in practice. Second, we want to avoid the constraint of data availability which could severely depress the sample size if compensation, institutional holding, and corporate governance variables are included. Dechow et al. (2011) argue that the inclusion of such variables would introduce biases into the sample due to data unavailability. Nevertheless, we feel that these omissions do not affect the main conclusions of the paper and invite future research to expand our model to cover these aspects of the context of earnings management. Besides, the ESCORE is constructed as the sum of 15 signals each carrying an equal weight. This approach is suitable to create a composite score that accumulates various aspects of the general context of earnings management and the resulting ESCORE is relatively easy to construct and understand. Nevertheless, although all selected signals are valid predictors of earnings management, the equal-weighted approach does not assess the relative importance of each signal. An alternative index construction approach which could deal with this issue is through a regression, but only if one could identify a suitable left-hand-side variable. Moreover, the

conversion of some of the components of ESCORE from continuous variables to binary signals could potentially be problematic. For example, managers could still manage earnings while keeping those values not look extremely high or low, hence not being flagged up by the ESCORE. While we still assert that the ESCORE as designed in this paper is suitable for the intended purposes, future research which seeks to develop a weighted and continuous model to capture the context of earnings management seems an interesting idea.

Appendix

Variable definitions [here t (t-1) refers to the current (previous) financial year].

$\gamma_{k,t}^{20}$ and $\gamma_{k,t}^{80}$ are, respectively, the lower and upper benchmarks of industry k in year t, determined as the 20th and 80th percentiles of γ in each industry-year where: γ is substituted by DEBT, MTB, MVE, NOA, CAP, BOOKTAX (definitions of these variables are below); $k = 1 \dots 43$ are 43 unique Datastream level 6 industries; $t = 1995 \dots 2011$ represent the sample years.

EDDEBT is defined as a dummy that takes the value of one if DDEBT is 5% or higher, zero otherwise. DDEBT is the percentage change in total of short- and long-term debt at the end of year t compared to that in year t-1, and zero otherwise.

ESEO is one if CSHO increases by 5% compared to last year and PROISSUE is positive, zero otherwise. CSHO is number of outstanding shares at the end of year t. PROISSUE is the proceeds from issuing common/preferred stocks in year t.

EMA is one if a firm announces a share-financed M&A deal in year t, zero otherwise.

EOV is one if beginning of year t MTB is higher than the corresponding $MTB_{k,t}^{80}$, zero otherwise. MTB is the ratio of market value to book value of equity, both measured at the same time.

EROA is one if ROA is equal to or larger than zero but smaller than 0.01, zero otherwise. ROA is earnings before extraordinary items in year t scaled by year t-1 total assets.

EDROA is one if DROA is equal to or larger than zero but smaller than 0.005, zero otherwise. DROA is the change in ROA in year t as compared to that in year t-1.

EDIV is one if DIVDEF is equal to or larger than zero but smaller than 0.01, zero otherwise. DIVDEF is the difference between net income and total cash dividends paid in year t scaled by total assets in year t-1.

EDISTRESS is one if ZSCORE for year t is negative, zero otherwise. $ZSCORE = 3.2 + 12.8x_1 + 2.5x_2 - 10.68x_3 + 0.029x_4$, where: x_1 is pre-tax income divided by current liabilities; x_2 is current assets divided by total liabilities; x_3 is current liabilities divided by total assets; x_4 is quick assets minus current liabilities divided by daily operating expense, where daily operating expense is sales minus pre-tax income minus depreciation expense divided by 365. x_1 to x_4 are all measured in the same year.

EDEBT is one if beginning of year t DEBT is lower than the corresponding $DEBT_{k,t}^{20}$, zero otherwise. DEBT is the ratio of total of short- and long-term debts to total assets, both measured at the same time.

ESIZE is one if beginning of year t MVE is smaller than the corresponding $MVE_{k,t}^{20}$, zero otherwise. MVE is market value of equity.

ECYCLE is one if (i) CFO is negative, CFF is positive, CFI is negative, or (ii) CFO is positive, CFF is positive, CFI is negative, zero otherwise, where CFO is operating

cash flows, CFF is financing cash flows, CFI is investing cash flows. All cash flows are beginning of year t cash flows.

EAUDIT is one if current year financial statements are not audited by one of the Big-5 audit firms, zero otherwise.

EBLOAT is one if beginning of year t NOA (net operating assets) is smaller than the corresponding $NOA_{k,t}^{20}$, zero otherwise. NOA $=$ (BVE + TDEBT − CHE)/TA, where BVE is net book value of equity, TDEBT is total of short- and long-term debt, CHE is cash and cash equivalents, and TA is total assets, all measured at the same time.

ECAP is one if beginning of year t CAP is smaller than the corresponding $CAP_{k,t}^{20}$, zero otherwise. CAP is property, plant, and equipment divided by total assets, both measured at the same time.

EBT is one if BOOKTAX is higher than the corresponding $BOOKTAX_{k,t}^{80}$, zero otherwise. BOOKTAX is the absolute value of the difference between year t pre-tax income and TTP scaled by year t sales. The total taxable profit, TTP for year t, is estimated as follows:

- If $TXT \leq 0$, then $TTP = 0$
- If $0 < TXT \leq LL \times SR$, then $TTP = \frac{TXT}{SR}$
- If $LL \times SR \leq TXT \leq (UL - LL) \times AR$, then $TTP = \frac{TXT - (LL \times SR)}{AR} + LL$
- If $TXT \leq (UL - LL) \times AR$, then $TTP = \frac{TXT - (LL \times SR) - [(UL - LL) \times AR]}{MR} + UL$

where TXT is the reported income tax expense, LL is the lower limit for marginal tax relief, UL is the upper limit for marginal tax relief, SR is the small profit tax rate, MR is the main tax rate, and AR $=$ (SR + MR)/2 (all measured at the same time).

ESCORE $=$ ESEO + EDDEBT + EMA + EOV + EROA + EDROA + EDIV + EDISTRESS + EDEBT + ESIZE + ECYCLE + EAUDIT + EBLOAT + ECAP + EBT.

$\mathbf{BHRR_{i,j}^{m}}$ is monthly buy-and-hold raw returns of stock i in month j, calculated as the percentage change in the Returns Index downloaded from Datastream at the end of each month.

$\mathbf{BHRR_{p,j}^{m}}$ is monthly buy-and-hold raw returns of portfolio p in month j, calculated as the equally-weighted $BHRR_{i,j}^{m}$ of all stocks belong to portfolio p.

$\mathbf{BHRR_{i,t}^{a}}$ is annual buy-and-hold raw returns of stock i in year t, calculated as $BHRR_{i,t}^{a} = \prod_{j=1}^{12} \left(1 + BHRR_{i,j}^{m} \right) - 1 (j = \text{June of year t} \dots \text{May of year t} + 1)$.

$\mathbf{CRR_{i,t}^{a}}$ is annual cumulative raw returns of stock i in year t, calculated as $CRR_{i,t}^{a} = \sum_{j=1}^{12} BHRR_{i,j}^{m} (j = \text{June of year t} \dots \text{May of year t} + 1)$.

$\mathbf{SDR_{d,j}^{m}}$ is the average monthly $BHRR_{i,j}^{m}$ of all stocks in size decile d in month j ($j = \text{June of year t} \dots \text{May of year t} + 1$), where the deciles are determined by sorting stocks by market value of equity at the end of year t-1.

$\mathbf{BHSAR_{i,j}^{m}}$ is monthly buy-and-hold size-adjusted returns of stock i in month j, calculated as the difference between $BHRR_{i,j}^{m}$ and the $SDR_{d,j}^{m}$ of the corresponding size decile to which stock i belongs.

$\mathbf{BHSAR_{p,j}^{m}}$ is monthly buy-and-hold size-adjusted returns of portfolio p in month j, calculated as the equally-weighted $BHSAR_{i,j}^{m}$ of all stocks belong to portfolio p.

$\mathbf{BHSAR_{i,t}^{a}}$ is annual buy-and-hold size-adjusted returns of stock i in year t, calculated as $BHSAR_{i,t}^{a} = \prod_{j=1}^{12} \left(1 + BHRR_{i,j}^{m} \right) - \prod_{j=1}^{12} \left(1 + SDR_{d,j}^{m} \right) (i \in d, j = \text{June of year t} \dots \text{May of year t} + 1)$.

$CSAR_{i,t}^a$ is annual cumulative size-adjusted returns of stock i in year t, calculated as $CSAR_{i,t}^a = \sum_{j=1}^{12}\left(BHRR_{i,j}^m - SDR_{d,j}^m\right)$ ($i \in d$, $j =$ June of year t … May of year t + 1).

$BHAR4F_{p,j}^m = BHRR_{p,j}^m - \left[Rf_j + \hat{\beta}_{1,p}^{4F}\left(Rm_j - Rf_j\right) + \hat{\beta}_{2,p}^{4F}SMB_j + \hat{\beta}_{3,p}^{4F}HML_j + \hat{\beta}_{4,p}^{4F}UMD_j\right]$ is monthly buy-and-hold abnormal returns of portfolio p in month j adjusted for the market, size, book-to-market and momentum factors; where: $\hat{\beta}_{1,p}^{4F}$, $\hat{\beta}_{2,p}^{4F}$, $\hat{\beta}_{3,p}^{4F}$, $\hat{\beta}_{4,p}^{4F}$ is the estimated intercept from the regression $BHRR_{p,j}^m - Rf_j = \alpha + \beta_1\left(Rm_j - Rf_j\right) + \beta_2 SMB_j + \beta_3 HML_j + \beta_4 UMD_j + \varepsilon$; Rf_j, Rm_j, SMB_j, HML_j, UMD_j are, respectively, the risk-free rate, returns on the market portfolio, size factor, book-to-market factor, momentum factor in month j, all taken from Gregory et al. (2013).

$E(R4F_{i,j}^m) = Rf_j + \hat{\beta}_{1,i}^{4F}\left(Rm_j - Rf_j\right) + \hat{\beta}_{2,i}^{4F}SMB_j + \hat{\beta}_{3,i}^{4F}HML_j + \hat{\beta}_{4,i}^{4F}UMD_j$ is monthly buy-and-hold expected returns of stock i in month j adjusted for the market, size, book-to-market and momentum factors; where: $\hat{\beta}_{1,i}^{4F}$, $\hat{\beta}_{2,i}^{4F}$, $\hat{\beta}_{3,i}^{4F}$, $\hat{\beta}_{4,i}^{4F}$ is the estimated coefficient from the regression $BHRR_{i,j}^m - Rf_j = \alpha + \beta_1\left(Rm_j - Rf_j\right) + \beta_2 SMB_j + \beta_3 HML_j + \beta_4 UMD_j + \varepsilon$; Rf_j, Rm_j, SMB_j, HML_j, UMD_j are, respectively, the risk-free rate, returns on the market portfolio, size factor, book-to-market factor, momentum factor in month j, all taken from Gregory et al. (2013).

$BHAR4F_{i,t}^a$ is annual buy-and-hold abnormal returns of stock i in year t adjusted for the market, size, book-to-market and momentum factors, calculated as $BHAR4F_{i,t}^a = \prod_{j=1}^{12}\left(1 + BHRR_{i,j}^m\right) - \prod_{j=1}^{12}\left[1 + E(R4F_{i,j}^m)\right]$ ($j =$ June of year t … May of year t + 1).

$CAR4F_{i,t}^a$ is annual cumulative abnormal returns of stock i in year t adjusted for the market, size, book-to-market and momentum factors, calculated as $CAR4F_{i,t}^a = \sum_{j=1}^{12}\left[BHRR_{i,j}^m - E(R4F_{i,j}^m)\right]$ ($j =$ June of year t … May of year t + 1).

$DAC_{i,t} = \frac{AC_{i,t}}{TA_{i,t-1}} - \left[\hat{\alpha} + \hat{\beta}_1\left(\frac{1}{TA_{i,t-1}}\right) + \hat{\beta}_2\left(\frac{\Delta REV_{i,t} - \Delta REC_{i,t}}{TA_{i,t-1}}\right) + \hat{\beta}_3\left(\frac{PPE_{i,t}}{TA_{i,t-1}}\right)\right]$, is discretionary accruals of stock i in year t. $\hat{\alpha}$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are the estimated coefficients from the following regression, which is run in each industry-year with at least 15 observations: $\frac{AC_{i,t}}{TA_{i,t-1}} = \alpha + \beta_1\left(\frac{1}{TA_{i,t-1}}\right) + \beta_2\left(\frac{\Delta REV_{i,t}}{TA_{i,t-1}}\right) + \beta_3\left(\frac{PPE_{i,t}}{TA_{i,t-1}}\right) + \varepsilon_{i,t}$, where $AC_{i,t}$ is total accruals of firm i in year t, which is calculated as the difference between income before extraordinary items and net operating cash flows; $TA_{i,t-1}$ is total assets of firm i at the end of year $t-1$; $\Delta REV_{i,t}$ and $\Delta REC_{i,t}$ are the changes in sales and receivables from year $t-1$ to year t of firm i, respectively; and $PPE_{i,t}$ is gross plant, property and equipment of firm i at the end of year t.

ADAC is the absolute value of DAC.

HDAC is one if DAC is equal to or higher than the 80th percentile of the corresponding industry-year ranked by DAC, zero otherwise.

$DWAC_{i,t} = \frac{WAC_{i,t}}{TA_{i,t-1}} - \left[\hat{\alpha} + \hat{\beta}_1\left(\frac{REV_{i,t}}{TA_{i,t-1}}\right) + \hat{\beta}_2\left(\frac{REV_{i,t} - \Delta REC}{TA_{i,t-1}}\right)\right]$, is discretionary working capital accruals of stock i in year t. $\hat{\alpha}$, $\hat{\beta}_1$, $\hat{\beta}_2$ are the estimated coefficients from the following regression, which is run in each industry-year with at least 15 observations: $\frac{WAC_{i,t}}{TA_{i,t-1}} = \alpha + \beta_1\left(\frac{REV_{i,t}}{TA_{i,t-1}}\right) + \beta_2\left(\frac{REV_{i,t} - \Delta REC}{TA_{i,t-1}}\right) + \varepsilon_{i,t}$, where: $WAC_{i,t}$ is working capital accruals of stock i in year t, which is calculated as $WAC = (\Delta CA - \Delta CHE) - (\Delta CL - \Delta STD)$ [ΔCA is change in current assets; ΔCHE is change in cash and cash equivalents; ΔCL is change in current liabilities; ΔSTD is change in short-term debts].

ADWAC is the absolute value of DWAC.

HDWAC is one if DWAC is equal to or higher than the 80th percentile of the corresponding industry-year ranked by DWAC, zero otherwise.

$\mathbf{DCF_{i,t}} = -1 \times \left\{ \frac{CFO_{i,t}}{TA_{i,t-1}} - \left[\hat{\alpha} + \hat{\beta}_1 \left(\frac{1}{TA_{i,t-1}} \right) + \hat{\beta}_2 \left(\frac{REV_{i,t}}{TA_{i,t-1}} \right) + \hat{\beta}_3 \left(\frac{\Delta REV_{i,t}}{TA_{i,t-1}} \right) \right] \right\}$, is abnormal cash flows of stock i in year t. $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ are the estimated coefficients from the following regression, which is run in each industry-year with at least 15 observations: $\frac{CFO_{i,t}}{TA_{i,t-1}} = \alpha + \beta_1 \left(\frac{1}{TA_{i,t-1}} \right) + \beta_2 \left(\frac{REV_{i,t}}{TA_{i,t-1}} \right) + \beta_3 \left(\frac{\Delta REV_{i,t}}{TA_{i,t-1}} \right) + \varepsilon_{i,t}$, where: $CFO_{i,t}$ is net cash flows from operation of firm i in year t.

ADCF is the absolute value of DCF.

HDCF is one if DCF is equal to or higher than the 80th percentile of the corresponding industry-year ranked by DCF, zero otherwise.

$\mathbf{DPROD_{i,t}} = \frac{PROD_{i,t}}{TA_{i,t-1}} - \left[\hat{\alpha} + \hat{\beta}_1 \left(\frac{1}{TA_{i,t-1}} \right) + \hat{\beta}_2 \left(\frac{REV_{i,t}}{TA_{i,t-1}} \right) + \hat{\beta}_3 \left(\frac{\Delta REV_{i,t}}{TA_{i,t-1}} \right) + \hat{\beta}_4 \left(\frac{\Delta REV_{i,t-1}}{TA_{i,t-1}} \right) \right]$, is abnormal production costs of stock i in year t. $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ are the estimated coefficients from the following regression, which is run in each industry-year with at least 15 observations: $\frac{PROD_{i,t}}{TA_{i,t-1}} = \alpha + \beta_1 \left(\frac{1}{TA_{i,t-1}} \right) + \beta_2 \left(\frac{REV_{i,t}}{TA_{i,t-1}} \right) + \beta_3 \left(\frac{\Delta REV_{i,t}}{TA_{i,t-1}} \right) + \beta_4 \left(\frac{\Delta REV_{i,t-1}}{TA_{i,t-1}} \right) + \varepsilon_{i,t}$, where: $PROD_{i,t}$ is production cost, calculated as the sum of cost of goods sold and change in inventory, of firm i in year t.

ADPROD is the absolute value of DPROD.

HDPROD is one if DPROD is equal to or higher than the 80th percentile of the corresponding industry-year ranked by DPROD, zero otherwise.

$\mathbf{DDISEXP_{i,t}} = -1 \times \left\{ \frac{DISEXP_{i,t}}{TA_{i,t-1}} - \left[\hat{\alpha} + \hat{\beta}_1 \left(\frac{1}{TA_{i,t-1}} \right) + \hat{\beta}_2 \left(\frac{REV_{i,t-1}}{TA_{i,t-1}} \right) \right] \right\}$, is abnormal discretionary expenses of stock i in year t. $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$ are the estimated coefficients from the following regression, which is run in each industry-year with at least 15 observations: $\frac{DISEXP_{i,t}}{TA_{i,t-1}} = \alpha + \beta_1 \left(\frac{1}{TA_{i,t-1}} \right) + \beta_2 \left(\frac{REV_{i,t-1}}{TA_{i,t-1}} \right) + \varepsilon_{i,t}$, where: $DISEXP_{i,t}$ is discretionary expenses, calculated as selling and general administrative expenses plus research and development expenses, of firm i in year t.

ADDISEXP is the absolute value of DDISEXP.

HDDISEXP is one if DDISEXP is equal to or higher than the 80th percentile of the corresponding industry-year ranked by DDISEXP, zero otherwise.

$\mathbf{TOTALRM_{i,t}} = \left[\frac{DCF_{i,t} - \overline{DCF_{t,k}}}{\sigma(DCF)_{t,k}} + \frac{DPROD_{i,t} - \overline{DPROD_{t,k}}}{\sigma(DPROD)_{t,k}} + \frac{DDISEXP_{i,t} - \overline{DDISEXP_{t,k}}}{\sigma(DISEXP)_{t,k}} \right] / 3 (i \in k)$ is total real earnings management, where: $\overline{DCF_{t,k}}$, $\overline{DPROD_{t,k}}$, $\overline{DDISEXP_{t,k}}$ [$\sigma(DCF)_{t,k}$, $\sigma(DPROD)_{t,k}$, $\sigma(DISEXP)_{t,k}$] is, respectively, the mean [standard deviation] of DCF, DPROD, DDISEXP of all firms in industry k in year t; $k = 1 \ldots 43$ are 43 unique Datastream level 6 industries.

BOSIZE is the number of board directors.

BOIND is the percentage of non-executive directors on board.

AUSIZE is the number of directors on the audit committee (equals to zero if a firm does not have an audit committee).

DUALITY is a dummy which is zero if a firm's CEO is also the chairman.

PLCOM is the average performance-linked compensation of all executive directors scaled by sales, where performance-linked compensation is defined as the total of bonus, shares, options and other long-term incentive pay awarded during the year.

$\mathbf{FSD\_SCORE_{i,t}} = \frac{\sum_{d=1}^{9} |AD_{d,i,t} - ED_d|}{9}$ ($d = 1, 2, \ldots, 9$) is the deviation of the observed distribution of financial statement figures from the distribution expected by Benford's law of firm i in year t, where: $AD_{d,i,t}$ ($d = 1, 2, \ldots, 9$) is the actual distribution of digit d, measured as the number of times d appears as the first digit of items reported on firm i's balance

sheet, income statement and cash flow statement in year t divided by the total number of items reported on those statements; ED_d is the expected distribution of digit d under Benford's law (i.e. $ED_1 = 0.3010$, $ED_2 = 0.1761$, $ED_3 = 0.1249$, $ED_4 = 0.0969$, $ED_5 = 0.0792$, $ED_6 = 0.0669$, $ED_7 = 0.0580$, $ED_8 = 0.0512$, $ED_9 = 0.0458$).

Data availability The sources of data used in the research are identified in the study.

Code availability Commonly used regressions within Stata.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Agarwal V, Taffler R (2007) Twenty-five years of the Taffler z-score model: does it really have predictive ability? Account Bus Res 37:285–300

Agarwal V, Taffler R (2008) Does financial distress risk drive the momentum anomaly? Financ Manage 37:461–484

Al-Shattarat B, Hussainey K, Al-Shattarat W (2018) The impact of abnormal real earnings management to meet earnings benchmarks on future operating performance. Int Rev Financ Anal. https://doi.org/10.1016/j.irfa.2018.10.001

Allen EJ, Larson CR, Sloan RG (2013) Accrual reversals, earnings and stock returns. J Account Econ 56:113–129

Alzoubi ESS (2016) Audit quality and earnings management: Evidence from Jordan. J Appl Acc Res 17:170–189

Amiram D, Bozanic Z, Rouen E (2015) Financial statement errors: evidence from the distributional properties of financial statement numbers. Rev Acc Stud 20:1540–1593

Astami EW, Tower G (2006) Accounting-policy choice and firm characteristics in the Asia Pacific region: an international empirical test of costly contracting theory. Int J Account 41:1–21

Athanasakou V, Strong NC, Walker M (2009) Earnings management or forecast guidance to meet analyst expectations? Account Bus Res 39:3–35

Athanasakou V, Strong NC, Walker M (2011) The market reward for achieving analyst earnings expectations: does managing expectations or earnings matter? J Bus Financ Acc 38:58–94

Athanasakou VE, Olsson P (2012) Earnings quality: firm fundamentals versus managerial discretion. *Available at:* http://www.hec.unil.ch/documents/seminars/dcc/758.pdf.

Atieh A, Hussain S (2012) Do UK firms manage earnings to meet dividend thresholds? Account Bus Res 42:77–94

Badertscher BA (2011) Overvaluation and the choice of alternative earnings management mechanisms. Account Rev 86:1491–1518

Ball R (2013) Accounting informs investors and earnings management is rife: two questionable beliefs. Account Horiz 27:847–853

Banz RW (1981) The relationship between return and market value of common stocks. J Financ Econ 9:3–18

Barber BM, Lyon JD (1997) Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. J Financ Econ 43:341–372

Barton J, Simko PJ (2002) The balance sheet as an earnings management constraint. Account Rev 77:1–27

Bauer R, Cosemans M, Schotman PC (2010) Conditional asset pricing and stock market anomalies in Europe. Eur Financ Manag 16:165–190

Beasley MS (1996) An empirical analysis of the relation between the board of director composition and financial statement fraud. Account Rev 71:443–465

Becker CL, Defond ML, Jiambalvo J, Subramanyam KR (1998) The effect of audit quality on earnings management. Contemp Account Res 15:1–24

Beneish MD (1997) Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance. J Account Public Policy 16:271–309

Beneish MD (1999) The detection of earnings manipulation. Financ Anal J 55:24

Beneish MD (2001) Earnings management: a perspective. Manag Financ 27:3–17

Beneish MD, Lee CMC, Nichols DC (2013) Earnings manipulation and expected returns. Financ Anal J 69:57–82

Bergstresser D, Philippon T (2006) CEO incentives and earnings management. J Financ Econ 80:511–529

Bernard VL, Skinner DJ (1996) What motivates managers' choice of discretionary accruals? J Account Econ 22:313–325

Beuselinck C, Cascino S, Deloof M, Vanstraelen A (2019) Earnings management within multinational corporations. Account Rev 94:45–76

Bond S (2000) UK investment and the capital market. *Paper presented at HM Treasury's Economic Growth and Government Policy Seminar at 11 Downing Street on 12th October 2000.*

Botsari A, Meeks G (2008) Do acquirers manage earnings prior to a share for share bid? J Bus Financ Acc 35:633–670

Botsari A, Meeks G (2018) Acquirers' earnings management ahead of stock-for-stock bids in 'hot'and 'cold'markets. J Account Public Policy 37:355–375

Burgstahler D, Dichev I (1997) Earnings management to avoid earnings decreases and losses. J Account Econ 24:99–126

Burns N, Kedia S (2006) The impact of performance-based compensation on misreporting. J Financ Econ 79:35–67

Carhart M (1997) On persistence in mutual fund performance. J Financ 52:57–82

Chen YW, Chou RK, Lin CB (2019) Investor sentiment, SEO market timing, and stock price performance. J Empir Financ 51:28–43

Cheng CSA, Thomas WB (2006) Evidence of the abnormal accrual anomaly incremental to operating cash flows. Account Rev 81:1151

Chi J, Gupta M (2009) Overvaluation and earnings management. J Bank Finance 33:1652–1663

Choi A, Choi JH, Sohn BC (2018) The joint effect of audit quality and legal regimes on the use of real earnings management: International evidence. Contemp Account Res 35:2225–2257

Choy SK, Lobo GJ, Tan Y (2021) Testing the accruals anomaly based on the speed of price adjustment. The European Journal of Finance. https://doi.org/10.1080/1351847X.2021.1998175

Chung R, Firth M, Kim J-B (2002) Institutional monitoring and opportunistic earnings management. J Corp Finan 8:29–48

Cohen DA, Dey A, Lys TZ (2008) Real and accrual-based earnings management in the pre- and post-sarbanes-oxley periods. Account Rev 83:757–787

Cohen DA, Zarowin P (2010) Accrual-based and real earnings management activities around seasoned equity offerings. J Account Econ 50:2–19

Daniel ND, Denis DJ, Naveen L (2008) Do firms manage earnings to meet dividend thresholds? J Account Econ 45:2–26

Dechow P, Ge W, schrand, C. (2010) Understanding earnings quality: a review of the proxies, their determinants and their consequences. J Account Econ 50:344–401

Dechow PM, Dichev ID (2002) The quality of accruals and earnings: the role of accrual estimation errors. Account Rev 77:35

Dechow PM, Ge W, Larson CR, Sloan RG (2011) Predicting material accounting misstatements. Contemp Account Res 28:17–82

Dechow PM, Sloan RG (1991) Executive incentives and the horizon problem. J Account Econ 14:51–89

Dechow PM, Sloan RG, Sweeney AP (1995) Detecting earnings management. Account Rev 70:193–225

Dechow PM, Sloan RG, Sweeney AP (1996) Causes and consequences of earnings manipulation: an analysis of firms subject to enforcement actions by the SEC. Contemp Account Res 13:1–36

Defond ML, Jiambalvo J (1994) Debt covenant violation and manipulation of accruals. J Account Econ 17:145–176

Desai H, Rajgopal S, Venkatachalam M (2004) Value-glamour and accruals mispricing: one anomaly or two? Account Rev 79:355

Desai MA (2005) The degradation of reported corporate profits. J Econom Perspect 19:171–192

Dichev ID, Skinner DJ (2002) Large-sample evidence on the debt covenant hypothesis. J Account Res 40:1091–1123

Dickinson V (2011) Cash flow patterns as a proxy for firm life cycle. Account Rev 86:1964–1994

Dopuch N, Holthausen RW, Leftwich RW (1987) Predicting audit qualifications with financial and market variables. Account Rev 62:431

Ducharme LL, Malatesta PH, Sefcik SE (2004) Earnings management, stock issues, and shareholder lawsuits. J Financ Econ 71:27

Duong C, Pescetto G (2019) Overvaluation and earnings management: does the degree of overvaluation matter? Account Bus Res 49:121–146

Dyreng SD, Hillegeist SA, Penalva F (2020) Earnings management to avoid debt covenant violations and future performance. Eur Account Rev. https://doi.org/10.1080/09638180.2020.1826337

Efendi J, Srivastava A, Swanson EP (2007) Why do corporate managers misstate financial statements? The role of option compensation and other factors. J Financ Econ 85:667

Erickson M, Wang S-W (1999) Earnings management by acquiring firms in stock for stock mergers. J Account Econ 27:149–176

Fama EF, French KR (1992) The cross-section of expected stock returns. J Finance 47:427–465

Fama EF, French KR (1995) Size and book-to-market factors in earnings and returns. J Finance 50:131–155

Fama EF, French KR (2006) Profitability, investment and average returns. J Financ Econ 82:491–518

Fields TD, Lys TZ, Vincent L (2001) Empirical research on accounting choice. J Account Econ 31:255–307

Francis J, Lafond R, Olsson P, Schipper K (2005) The market pricing of accruals quality. J Account Econ 39:295–327

Francis J, Lafond R, Olsson PM, Schipper K (2004) Costs of equity and earnings attributes. Account Rev 79:967–1010

Francis JR, Maydew EL, Sparks HC (1999) The role of big 6 auditors in the credible reporting of accruals. Auditing 18:17

Garcia Lara JM, Garcia Osma B, Neophytou E (2009) Earnings quality in ex-post failed firms. Account Bus Res 39:119–138

Gerakos J, Kovrijnykh A (2013) Performance shocks and misreporting. J Account Econ 56:57–72

Ghosh A, Moon D (2010) Corporate debt financing and earnings quality. J Bus Financ Acc 37:538–559

Gore P, Pope PF, Singh AK (2007) Earnings management and the distribution of earnings relative to targets: UK evidence. Account Bus Res 37:123–149

Gray P, Liao IS, Strydom M (2018) The profitability of trading NOA and accruals: one effect or two? Int Rev Financ Anal 58:211–224

Gregory A, Tharyan R, Christidis A (2013) Constructing and testing alternative versions of the fama-french and carhart models in the UK. J Bus Financ Acc 40:172–214

Guay W, Kothari SP, Watts RL (1996) A market-based evaluation of discretionary accrual models. J Account Res 34:83–105

Guidry F, Leone AJ, Rock S (1999) Earnings-based bonus plans and earnings management by business-unit managers. J Account Econ 26:113–142

Gunny K (2010) The relation between earnings management using real activities manipulation and future performance: evidence from meeting earnings benchmarks. Contemp Account Res 27:855–888

Hanlon M, Heitzman S (2010) A review of tax research. J Account Econ 50:127–178

Healy PM, Wahlen JM (1999) A review of the earnings management literature and its implications for standard setting. Account Horiz 13:365–383

Hirshleifer D, Hou K, Teoh SH (2012) The accrual anomaly: risk or mispricing? Manage Sci 58:320–335

Hirshleifer D, Hou K, Teoh SH, Zhang Y (2004) Do investors overvalue firms with bloated balance sheets? J Account Econ 38:297

HM Revenue & Customs (2013) Corporation Tax Statistics. HM Revenue & Customs, London

Holthausen RW, Larcker DF, Sloan RG (1995) Annual bonus schemes and the manipulation of earnings. J Account Econ 19:29–74

Houmes RE, Skantz TR (2010) Highly valued equity and discretionary accruals. J Bus Financ Acc 37:60–92

Hribar P, Collins DW (2002) Errors in estimating accruals: implications for empirical research. J Account Res 40:105–134

Iqbal A, Espenlaub S, Strong N (2009) Earnings management around UK open offers. Eur J Financ 15:29–51

Iqbal A, Strong N (2010) The effect of corporate governance on earnings management around UK rights issues. Int J Manag Finance 6:168–189

Jensen MC (2005) Agency costs of overvalued equity. Financ Manage 34:5–19

Jiang W, Soares N, Stark AW (2016) Loss persistence and returns in the UK. Account Bus Res 46:221–242

Jones JJ (1991) Earnings management during import relief investigations. J Account Res 29:193–228

Khan M (2008) Are accruals mispriced? Evidence from tests of an intertemporal capital asset pricing model. J Account Econ 45:55

Klein A (2002) Audit committee, board of director characteristics, and earnings management. J Account Econ 33:375

Kothari SP, Warner JB (1997) Measuring long-horizon security price performance. J Financ Econ 43(3):301–339

Kothari SP, Mizik N, Roychowdhury S (2016) Managing for the moment: the role of earnings management via real activities versus accruals in SEO valuation. Account Rev 91:559–586

Krishnan GV (2003) Does big 6 auditor industry expertise constrain earnings management? Account Horiz 17:1–16

Lakonishok J, Shleifer A, Vishny RW (1994) Contrarian investment, extrapolation, and risk. J Finance 49:1541–1578

Lang M, Lundholm R (1993) Cross-sectional determinants of analyst ratings of corporate disclosures. J Account Res 31:246–271

Lee E, Liu W, Strong N (2007) UK evidence on the characteristics versus covariance debate. Eur Financ Manag 13:742–756

Loughran T, Ritter JR (1995) The new issues puzzle. J Finance 50:23–51

Louis H (2004) Earnings management and the market performance of acquiring firms. J Financ Econ 74:121–148

Mashruwala C, Rajgopal S, Shevlin T (2006) Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. J Account Econ 42:3

Mcnichols MF (2000) Research design issues in earnings management studies. J Account Public Policy 19:313–345

Mohanram SP (2005) Separating winners from losers among low book-to-market stocks using financial statement. Rev Acc Stud 10:133–170

Nguyen TT, Duong CM, Narendran S (2021) CEO profile and earnings quality. Rev Quant Financ Acc 56:987–1025

Ou JA, Penman SH (1989a) Accounting measurement, price-earnings ratio, and the information content of security prices. J Account Res 27:111–144

Ou JA, Penman SH (1989b) Financial statement analysis and the prediction of stock returns. J Account Econ 11:295–329

Owens EL, Wu JS, Zimmerman J (2017) Idiosyncratic shocks to firm underlying economics and abnormal accruals. Account Rev 92:183–219

Pae J (2007) Unexpected accruals and conditional accounting conservatism. J Bus Financ Acc 34:681–704

Papanastasopoulos G (2020) Percent accruals and the accrual anomaly: Evidence from the UK. Account Forum 44:287–310

Peasnell KV, Pope PF, Young S (2000) Detecting earnings management using cross-sectional abnormal accruals models. Account Bus Res 30:313–326

Peasnell KV, Pope PF, Young S (2005) Board monitoring and earnings management: do outside directors influence abnormal accruals? J Bus Financ Acc 32:1311–1346

Piotroski JD (2000) Value investing: the use of historical financial statement information to separate winners from losers. J Account Res 38:1–41

Rangan S (1998) Earnings management and the performance of seasoned equity offerings. J Financ Econ 50:101–122

Roychowdhury S (2006) Earnings management through real activities manipulation. J Account Econ 42:335–370

Sakaki H, Jackson D, Jory S (2017) Institutional ownership stability and real earnings management. Rev Quant Financ Acc 49:227–244

Shivakumar L (2000) Do firms mislead investors by overstating earnings before seasoned equity offerings? J Account Econ 29:339–371

Siew Hong T, Wong TJ (2002) Why new issues and high-accrual firms underperform: the role of analysts' credulity. Review of Financial Studies 15:869–900

Simlai P (2021) Accrual mispricing, value-at-risk, and expected stock returns. Rev Quant Financ Acc 57:1487–1517

Skinner DJ, Sloan RG (2002) Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Rev Acc Stud 7:289–312

Sloan RG (1996) Do stock prices fully reflect information in accruals and cash flows about future earnings? Account Rev 71:289–315

Soares N, Stark AW (2009) The accruals anomaly – can implementable portfolio strategies be developed that are profitable net of transactions costs in the UK? Account Bus Res 39:321–345

Spiess DK, Affleck-Graves J (1995) Underperformance in long-run stock returns following seasoned equity offerings. J Financ Econ 38:243–267

Srivastava A (2019) Improving the measures of real earnings management. Rev Acc Stud 24:1277–1316

Sundvik D (2017) Book-tax conformity and earnings management in response to tax rate cuts. J Int Account Audit Tax 28:31–42

Taffler R (1983) The assessment of company solvency and performance using a statistical model. Account Bus Res 13:295–307

Teoh SH, Welch I, Wong TJ (1998a) Earnings management and the long-run market performance of initial public offerings. J Finance 53:1935–1974

Teoh SH, Welch I, Wong TJ (1998b) Earnings management and the underperformance of seasoned equity offerings. J Financ Econ 50:63–99

Thomas J, Zhang X-J (2000) Identifying unexpected accruals: a comparison of current approaches. J Account Public Policy 19:347–376

Twedt B (2016) Spreading the word: price discovery and newswire dissemination of management earnings guidance. Account Rev 91:317–346

Walker M (2013) How far can we trust earnings numbers? What research tells us about earnings management. Account Bus Res 43:445–481

Watts RL (2003a) Conservatism in accounting part i: explanations and implications. Account Horiz 17:207–221

Watts RL (2003b) Conservatism in accounting part ii: evidence and research opportunities. Account Horiz 17:287–301

Watts RL, Zimmerman JL (1986) Positive accounting theory. Prentice-Hall, London

Wu J, Zhang L, Zhang XF (2010) Theq-theory approach to understanding the accrual anomaly. J Account Res 48:177–223

Xie B, Davidson WN III, Dadalt PJ (2003) Earnings management and corporate governance: the role of the board and the audit committee. J Corp Finan 9:295

Xie H (2001) The mispricing of abnormal accruals. Account Rev 76:357

Young S (1999) Systematic measurement error in the estimation of discretionary accruals: an evaluation of alternative modelling procedures. J Bus Financ Acc 26:833

Zang AY (2012) Evidence on the trade-off between real activities manipulation and accrual-based earnings management. Account Rev 87:675–703

Zhang Y, Chong G, Jia R (2019) Fair value, corporate governance, social responsibility disclosure and banks' performance. Rev Acc Financ 19:30–47