ARTICLE IN PRESS

Advances in Accounting xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Advances in Accounting

journal homepage: www.elsevier.com/locate/adiac

Does benchmark-beating detect earnings management? Evidence from accounting irregularities[☆]

David G. Harris[a], Linna Shi[b], Hong Xie[c,*]

[a] Whitman School of Management, Syracuse University, Syracuse, NY 13244-2450, United States
[b] Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, United States
[c] Gatton College of Business and Economics, University of Kentucky, Lexington, KY 40506, United States

ARTICLE INFO

Keywords:
Accounting irregularities
Accounting fraud
Earnings management
Earnings benchmarks
Benchmark-beating

ABSTRACT

We examine whether meeting or slightly beating an earnings benchmark (benchmark-beating) is (1) associated with accounting irregularities, an extreme and certain case of earnings management, (2) useful for detecting accounting irregularities both incremental and relative to discretionary accruals and to F-scores (Dechow, Ge, Larson, & Sloan, 2011), and (3) more useful for detecting *opportunistic* accounting irregularities, a more harmful form of earnings manipulation identified in Badertscher, Collins, and Lys (2012), than accounting irregularities in general. We identify an accounting irregularity sample where earnings are restated due to intentional misreporting and construct a control sample where earnings are not restated. We find that benchmark-beating is significantly positively associated with the probability of accounting irregularities after controlling for other determinants of accounting irregularities. In addition, benchmark-beating is useful for detecting accounting irregularities incremental to discretionary accruals and F-scores; benchmark-beating ties with and sometimes outperforms discretionary accruals for detecting accounting irregularities in a one-on-one horse race but is dominated by F-scores. Finally, benchmark-beating is more useful for detecting opportunistic accounting irregularities than accounting irregularities in general. Overall, we contribute to the literature by validating benchmark-beating as a proxy for earnings management.

1. Introduction

We examine whether meeting or slightly beating an earnings benchmark (hereafter, benchmark-beating) is (1) associated with accounting irregularities, an extreme and certain case of earnings management, (2) useful for detecting accounting irregularities both incremental and relative to discretionary accruals and to F-scores (Dechow, Ge, Larson, & Sloan, 2011), and (3) more useful for detecting *opportunistic* accounting irregularities, a more harmful form of earnings manipulation identified in Badertscher, Collins, and Lys (2012), than accounting irregularities in general. The literature documents three earnings benchmarks and measures benchmark-beating by identifying firms whose earnings slightly increase from last year's earnings (the earnings change benchmark), whose earnings are slightly positive (the earnings level benchmark), and whose earnings are equal to or slightly above analyst earnings forecasts (the earnings forecast benchmark).

Our research questions are important for several reasons. First, a large and growing volume of studies in the accounting literature use benchmark-beating as a proxy for earnings management while evidence that links benchmark-beating to *actual* earnings management is limited (see more detailed discussion in the next section). Dechow, Ge, and Schrand (2010), p.365) make the above point clear when they conclude, after reviewing the vast literature of earnings quality and earnings management, that "[t]the totality of the evidence indicates that the use of small profits as a proxy for earnings management more generally is *unsubstantiated* (emphasis added)." We seek to provide evidence on a link between benchmark-beating and earnings management in this paper.

Second, benchmark-beating and discretionary accruals are arguably the two most widely used proxies for earnings management (Kothari,

☆ We thank Laurel Franzen, Brian Bratten, Monika Causholli, Joseph Comprix, David Hulse, Badr Ismail, Danny Lanier, Lihong Liang, Jeff Payne, Bob Ramsay, Dan Stone, David Weinbaum, David Ziebart and workshop participants at University of Kentucky, Syracuse University, the 2010 New York Accounting and Finance Forum, and the 2013 American Accounting Association Annual Meeting for helpful comments and suggestions. We also thank Karen Hennes, Andrew Leone, and Brian Miller for generously sharing their financial reporting irregularity dataset and Thomson Financial for providing analyst forecast data through the Institutional Brokers Estimate System (I/B/E/S). Hong Xie gratefully acknowledges financial support from the Von Allmen Research Support endowment and the PWC Fellowship endowment at University of Kentucky.

* Corresponding author.
E-mail addresses: dgharris@syr.edu (D.G. Harris), shiln@ucmail.uc.edu (L. Shi), hongxie98@uky.edu (H. Xie).

https://doi.org/10.1016/j.adiac.2018.04.001
Received 15 October 2017; Received in revised form 6 April 2018; Accepted 6 April 2018
0882-6110/

Please cite this article as: Harris, D.G., Advances in Accounting (2018), https://doi.org/10.1016/j.adiac.2018.04.001

2001) and F-scores are arguably the most powerful detector of earnings misstatements (Dechow et al., 2011). However, little is known about whether benchmark-beating, discretionary accruals, and F-scores capture the same or different aspects of earnings management, and how these three measures compare with one another in terms of detecting earnings management.[1] If benchmark-beating captures the same aspects of earnings management as discretionary accruals and F-scores, the coefficient on benchmark-beating could become insignificant, in an earnings management detection model where benchmark-beating is an explanatory variable, after including discretionary accruals and F-scores as additional explanatory variables. On the other hand, if benchmark-beating captures aspects of earnings management different from discretionary accruals and F-scores, the coefficient on benchmark-beating will remain significant after including discretionary accruals and F-scores as additional explanatory variables. In such a case, benchmark-beating has detective power for earnings management *incremental* to discretionary accruals and F-scores. Relatedly, it is also of interest to examine which measure, out of these three, is *relatively* superior in a one-on-one horse race to detect earnings management. Dechow et al. (2011, p. 23) highlight the need to compare different measures of earnings management by calling for future research "to analyze the role of governance, compensation, insider trading, short selling, incentives to *meet and beat* analyst forecasts, and so on and to determine the *relative* importance of these variables (emphases added) over financial statement information in detecting overstatements of earnings." We answer this call in Dechow et al. (2011).

Third, understanding the incremental and relative ability of benchmark-beating in detecting earnings manipulation with respect to discretionary accruals and F-scores is of importance in its own right. Benchmark-beating is a parsimonious and *timely* metric, which can be determined by even naïve investors as soon as a firm's earnings are announced without relying on earnings of any other firm in the industry. In sharp contrast, one must wait weeks or even months after a firm's earnings announcement until the release of that firm's financial statements *and* the releases of financial statements of *all* other firms in the same industry to estimate that firm's discretionary accruals and F-score because prior literature commonly estimate discretionary accruals and F-scores in the cross-section in each year and industry.[2] Moreover, many "average" investors may not have the resources, time, and skill to estimate discretionary accruals and F-scores. Given benchmark-beating's lead in timeliness and ease of implementation, discretionary accruals and F-scores must dominate benchmark-beating in detecting accounting irregularities for them to remain viable contenders for detectors of earnings management. Thus, it is important to compare the efficacy of benchmark-beating, discretionary accruals, and F-scores in detecting accounting irregularities.

Fourth, accounting literature shows that some earnings management is for opportunistic reasons. For example, Badertscher et al. (2012) find that originally reported (or manipulated) earnings and accrual components are less predictive of future cash flows than the restated (or non-manipulated) counterparts for their opportunistic manipulation subsample whereas the opposite is true for their non-opportunistic manipulation subsample. These findings suggest that opportunistic manipulation is more harmful than non-opportunistic manipulation. Therefore, the detection of opportunistic manipulation is potentially more valuable to investors, auditors, creditors, financial analysts, regulators, and other stakeholders. The extant literature, however, has not examined whether benchmark-beating, discretionary accruals, and F-scores can detect opportunistic earnings management. We fill this void by investigating whether benchmark-beating is more useful in detecting opportunistic accounting irregularities than accounting irregularities in general.

One important reason for why the extant literature has provided only limited evidence on a link between benchmark-beating and actual earnings management is the difficulty of measuring earnings management, which is largely unobservable. Prior studies that use, for example, discretionary accruals and earnings response coefficients (ERCs) to examine the relation between benchmark-beating and earnings management provide only circumstantial evidence due to the inability of discretionary accruals and ERCs to unequivocally capture earnings management. We overcome this difficulty by using a sample of accounting irregularity firms where earnings are known to be restated due to intentional misreporting.[3] Since we know these firms violated U.S. GAAP and were required to restate their earnings, we can unequivocally identify these firms as earnings manipulators and precisely measure the amounts of their earnings manipulations. This allows us to provide evidence on the link between benchmark-beating and *actual* earnings management.

We construct an irregularity sample, based on Hennes, Leone, and Miller (2008), that consists of firms that restated their earnings during 1999 to 2005 due to intentional misreporting.[4] We also construct a control sample of firms from the COMPUSTAT universe during the same period that did *not* restate their earnings. We conduct several sets of tests. First, we compare the benchmark-beating samples (i.e., firm year observations where earnings meet or slightly beat one of the three benchmarks) with the non-benchmark-beating samples. We find that the percentages of intentional misreporting are higher in the benchmark-beating samples than the corresponding non-benchmark-beating samples. This finding provides univariate evidence that benchmark-beating is positively associated with accounting irregularities, i.e., benchmark-beating firms are more likely to be earnings manipulators than non-benchmark-beating firms. We then follow the research design of Jones, Krishnan, and Melendrez (2008) and Dechow et al. (2011), and provide multivariate evidence on the association between benchmark-beating and accounting irregularities using a logistic model. We find that benchmark-beating is significantly positively associated with irregularities after controlling for common determinants of accounting irregularities, suggesting that benchmark-beating can detect accounting irregularities, similar to discretionary accruals tested in Jones et al. (2008) and F-scores tested in Dechow et al. (2011).

Second, we successively add discretionary accruals and F-scores into our accounting irregularity detection model. We find that benchmark-beating remains significantly positively associated with accounting irregularities after incorporating discretionary accruals, F-scores, or both as additional explanatory variables. This suggests that benchmark-beating has *incremental* detective power for accounting irregularities beyond discretionary accruals and F-scores, which further implies that these three measures capture different aspects of earnings management and that they are complements to one another in detecting accounting irregularities.[5] In addition, we conduct a one-on-one horse race

[1] By construct, these three measures capture different aspects of earnings management. Specifically, benchmark-beating captures incentives to manage earnings to beat benchmarks, discretionary accruals capture manipulated earnings that are not related to cash flows, sales, and other operating activities, while F-scores capture fraudulent misreporting. However, it is ultimately an empirical question whether these three measures are incremental to one another in detecting earnings management.

[2] Francis, Schipper, and Vincent (2002, p. 538) report that only a small percentage (2.6%) of firms disclose detailed statements of cash flows concurrently in their earnings announcement press releases. Researchers thus need to wait weeks after a firm's earnings announcement until the release of that firm's statement of cash flows to estimate that firm's accruals using the statement of cash flow approach (Hribar & Collins, 2002).

[3] Although SAS No. 99 of American Institute of Certified Public Accountants (AICPA, 2002) classify intentional misstatements as "fraud," we follow Hennes et al. (2008) and use the more inclusive term "irregularities."

[4] We thank Karen Hennes, Andrew Leone, and Brian Miller for generously sharing this dataset.

[5] Assume that we have two firms and these two firms have similar discretionary accruals and similar F-scores with each other. Based on discretionary accruals and F-scores only, these two firms have equal probability of committing financial misreporting. Benchmark-beating having *incremental* detective power for accounting irregularities means, if one firm meets or slightly beats an earnings benchmark whereas the other does not, the benchmark beater has a higher probability of committing financial misreporting

between benchmark-beating and discretionary accruals, and find that the former beats (ties with) the latter in one (two) race (races). Thus, benchmark-beating is at least as good as, and sometimes better than, discretionary accruals in detecting accounting irregularities. In the horse race between benchmark-beating and F-scores, the former is dominated by the latter in all three races. This re-confirms the superior ability of F-scores in detecting accounting irregularities (Dechow et al., 2011).[6]

Third, we follow Badertscher et al. (2012) and partition the irregularity sample into the opportunistic irregularity subsample, where earnings are manipulated for opportunistic reasons, and the non-opportunistic irregularity subsample, where earnings are manipulated for non-opportunistic reasons (see Section 4.4 for the construction of the opportunistic and non-opportunistic irregularity subsamples). We find that benchmark-beating is more positively associated with opportunistic accounting irregularities than accounting irregularities in general. In addition, in the horse race between benchmark-beating and discretionary accruals for detecting opportunistic irregularities, the former beats (ties with) the latter in two (one) races (race); in the horse race between benchmark-beating and F-scores, the former ties with (is outperformed by) the latter in two (one) races (race). These findings suggest that benchmark-beating is more useful for detecting a more harmful form of manipulation—opportunistic accounting irregularities than accounting irregularities in general.

We contribute to the literature in several ways. First, we are among the first to provide evidence that benchmark-beating, a widely-used proxy for earnings management in the accounting literature, is associated with an extreme and certain case of earnings management—accounting irregularities. Because less extreme, within GAAP, earnings management that does not lead to restatements is unobservable, it is not possible for us to directly examine whether benchmark-beating is associated with less extreme earnings management. However, our findings suggest that *some* managers are willing to commit fraud in order to meet or slightly beat an earnings benchmark. If so, then it stands to reason that a larger number of other managers are willing to commit less extreme earnings management to achieve the same goal. Thus, our finding that benchmark-beating is associated with extreme earnings management (i.e., accounting irregularities) implies that benchmark-beating is also associated with less extreme earnings management.

Second, to our knowledge, we are the first to provide evidence that three most commonly used earnings management proxies, benchmark-beating, discretionary accruals, and F-scores, capture different aspects of earnings management. As such, these three proxies are incrementally useful to one another in detecting earnings management. Finally, we are the first to show that benchmark-beating is especially useful in detecting a more harmful form of earnings manipulation—opportunistic accounting irregularities. Taken together, our three findings complement one another to validate the use of benchmark-beating as a proxy for earnings management. Our findings also have important implications to investors, auditors, regulators, and other stakeholders for developing procedures to detect earnings management and fraud because benchmark-beating measures are timelier and easier to implement than discretionary accruals and F-scores.

The remainder of the paper is organized as follows. Section 2 reviews the literature and develops research questions. Section 3 describes sample selection and variable measurement. Section 4 presents regression models and reports our findings. We conclude in Section 5.

2. Literature review and research questions

2.1. Benchmark-beating and earnings management

Hayn (1995), Burgstahler and Dichev (1997), and Degeorge, Patel, and Zeckhauser (1999) document discontinuities in earnings distributions around three earnings benchmarks: (1) earnings of last year (the earnings change benchmark), (2) zero earnings (the earnings level benchmark), and (3) consensus analyst earnings forecasts (the earnings forecast benchmark). In particular, the number of firms whose earnings meet or slightly beat an earnings benchmark is unusually high whereas the number of firms whose earnings slightly miss the benchmark is unusually low. Burgstahler and Dichev (1997) and Degeorge et al. (1999) attribute these discontinuities to earnings management.

Durtschi and Easton (2005, 2009) challenge earnings management as an explanation for discontinuities and propose an alternative explanation for discontinuities: discontinuities are due to certain research design choices such as scaling and sample selection. Burgstahler and Chuk (2015) re-examine the scaling and sample selection issues raised by Durtschi and Easton (2005, 2009) and refute scaling or selection as an explanation for discontinuities. In a similar spirit, Donelson, McInnis, and Mergenthaler (2013) examine a sample of firms that settled accounting-related securities litigation and restated earnings for alleged GAAP violations (the litigation sample). They compare frequency distributions of restated earnings with those of originally-reported earnings for the litigation sample. They find that discontinuities in earnings distributions around zero analyst forecast errors, zero earnings changes, and zero earnings levels are *not* present or are attenuated in restated earnings but *are* present in originally-reported earnings. Since the only difference between restated earnings and originally-reported earnings is that the former is non-manipulated earnings and the latter is manipulated earnings of the *same* firm and *same* year, the discontinuities in originally-reported earnings are due to earnings management and cannot be attributed to scaling, sample selection biases, and other research design issues raised in Durtschi and Easton (2005, 2009).

There are only a relatively small number of papers that use discontinuities around benchmarks, detected by the standardized difference statistic proposed in Burgstahler and Dichev (1997) and calculated from a sample *distribution*, as a proxy for earnings management (e.g., Beatty, Ke, & Petroni, 2002; Beaver, McNichols, & Nelson, 2003; Brown & Caylor, 2005). In sharp contrast, a large and growing body of studies use benchmark-beating as a proxy for earnings management. Studies using that proxy include, just to mention a few, Frankel, Johnson, and Nelson (2002), Phillips, Pincus, and Rego (2003), Cheng and Warfield (2005), Vafeas (2005), Davis, Soo, and Trompeter (2009), Gunny (2010), Kanagaretnam, Lim, and Lobo (2010), Reichelt and Wang (2010), Cahan, Zhang, and Veenman (2011), McInnis and Collins (2011), Boone, Khurana, and Raman (2012), Srinidhi, Gul, and Tsui (2011), Minutti-Meza (2013), and Carcello and Li (2013). These studies all use a dummy variable to proxy for earnings management. In contrast to the standardized difference statistic for detecting discontinuities around benchmarks, the dummy variable does not depend on a sample distribution, and is set to one if a firm's earnings meet or slightly beat an earnings benchmark and zero otherwise.

Although benchmark-beating is widely used as a proxy for earnings management, there is limited evidence linking benchmark-beating to *unequivocal* earnings management in the extant literature. Several studies provide *circumstantial* evidence that meeting or slightly beating benchmarks represents earnings management. For example, Caramanis and Lennox (2008) examine the effect of audit effort on earnings management and find that when audit hours are lower clients are more likely to manage earnings upwards in order to meet or beat the zero earnings benchmark. This suggests that the benchmark-beating

(*footnote continued*)
than the non-benchmark beater.

[6] We note that the horse race between benchmark-beating (or discretionary accruals) and F-scores for detecting accounting misstatement may not be a fair one because F-scores are calibrated from the accounting irregularity prediction model (see Eq. (3)). Naturally, F-scores so calibrated are quite strongly positively associated with accounting irregularities. In contrast, benchmark-beating and discretionary accruals are *not* pre-calibrated against accounting irregularities.

behavior is more pronounced when auditor monitoring is low, consistent with benchmark-beating being a form of earnings management that can be constrained by auditor effort (see also Keung, Lin, & Shin, 2010).

Some studies investigate whether benchmark-beating represents earnings management by linking benchmark-beating to proxies for earnings management such as discretionary accruals. A positive association supports the conclusion that benchmark-beating constitutes earnings management. However, discretionary accruals are a *noisy* proxy for earnings management (with large Type I and Type II errors) and are systematically associated with firm performance. To alleviate such a problem, Ayers, Jiang, and Yeung (2006) examine the associate between meeting or slightly beating the three earnings benchmarks and discretionary accruals *and* the association between meeting or slightly beating the *pseudo* earnings targets and discretionary accruals. They conclude that, among three earnings benchmarks, only meeting or slightly beating analyst earnings forecasts likely captures earnings management.

Our paper is related to but distinct from Donelson et al. (2013) for the following reasons. First, we examine the effect of *benchmark-beating* on accounting misreporting whereas they investigate whether *discontinuities* in earnings distributions are present or absent in originally-reported (manipulated) earnings and restated (non-manipulated) earnings. As such, we compare the irregularity sample with the control sample whereas they compare originally-reported earnings with restated earnings of the *same* sample (the litigation sample).[7] More substantively, benchmark-beating is related to but is *not* the same as discontinuities. Specifically, discontinuities can be caused by (i) the frequency of meeting or slightly beating benchmarks being too high, (ii) the frequency of slightly missing benchmarks being too low, or (iii) both.[8] Thus, the link established in Donelson et al. (2013) between discontinuities and earnings management does *not* automatically translate into to a link between benchmark-beating and earnings management. This is because discontinuities can be driven entirely by the frequencies of slightly missing benchmarks being too low when the frequencies of meeting or slightly beating benchmarks are perfectly normal. Second, discontinuities are a property of a sample *distribution* whereas benchmark-beating is unrelated to a sample distribution or to any other firms. A researcher must utilize a large enough sample to calculate test statics (e.g., standardized differences) that detect whether there is a discontinuity around a benchmark. That is, discontinuities are a sample-level variable. In sharp contrast, a researcher can determine whether a single *firm* in a year is a benchmark beater without regard to any other firms. That is, benchmark-beating is a firm-year variable. That is why benchmark-beating measure are more widely used as a firm-year proxy for earnings management in the accounting literature and why benchmark-beating is comparable to discretionary accruals or F-scores (because they are also firm-year variables). Third, benchmark-beating is timelier. Investors can immediately determine whether a firm is a benchmark beater after its earnings announcement. In contrast, investors must wait weeks or even months after a firm's earnings announcement until all other firms across ideally all industries announce their earnings to determine whether there is a discontinuity in an earnings distribution. Lastly, we conduct many other analyses that Donelson et al. (2013) could not do using their research design. In fact, there is no overlap in analyses conducted in these two papers except the proportion tests of Table 5 in Donelson et al. (2013, p. 265), which overlaps with our descriptive statistics in Panel A of Table 3.

To summarize, the extant literature provides mostly circumstantial evidence supporting a link between benchmark-beating and earnings management. One of the objectives of our paper is to provide more systematic evidence on the link between benchmark-beating and earnings management, using accounting irregularities (intentional misreporting) as a measure of *actual* earnings management.

2.2. Research questions

We use accounting irregularities (i.e., intentional misreporting) as our measure of earnings management and compare an irregularity sample where earnings are restated due to intentional misreporting with a control sample where earnings are not restated. If benchmark-beating captures earnings management, then we should observe a positive association between benchmark-beating and accounting irregularities. On the other hand, if benchmark-beating firms do nothing wrong (i.e., these firms are not more likely to manipulate earnings than non-benchmark-beating firms), then we should not observe any association between benchmark-beating and accounting irregularities. Thus, our first research question is:

RQ1. Is benchmark-beating positively associated with accounting irregularities?

Benchmark-beating and discretionary accruals are two commonly used proxies for earnings management in the literature. Jones et al. (2008) evaluate the ability of various measures of discretionary accruals to detect the extreme cases of earnings management—earnings restatements. They find that various measures of discretionary accruals are positively related to actual cases of fraud and earnings restatements and thus provide a link between discretionary accruals and actual earnings management. Recently, Dechow et al. (2011) extend the detective model of misstatements to encompass a variety of relevant variables and have developed, arguably, the most powerful detector of accounting misstatements, F-scores. However, little is known about how benchmark-beating, discretionary accruals, and F-scores perform *incremental* to one another (i.e., whether one variable is still significant in predicting accounting misstatements in the presence of another variable) or *relative* to one another (i.e., whether one variable dominates another variable in a one-on-one horse race for predicting accounting misstatements). Our second research question fills this void and is stated below:

RQ2. Is benchmark-beating useful for detecting accounting irregularities both incremental and relative to discretionary accruals and F-scores?

Accounting literature contains two competing views regarding managerial discretion or earnings management. One view holds that earnings management conveys managers' private information and hence enhances earnings' informativeness (Subramanyam, 1996; Watts & Zimmerman, 1986). The other view maintains that, due to conflicts of interest between managers and shareholders, managers may manage earnings opportunistically (Subramanyam, 1996; Watts & Zimmerman, 1986). Badertscher et al. (2012) test the above and other views of managerial discretion by dividing their restatement sample into the opportunistic manipulation and non-opportunistic manipulation subsamples. They find that first-reported (or manipulated) earnings and accrual components are less predictive of future cash flows relative to restated (or non-manipulated) counterparts for the opportunistic

[7] The only exception is their Table 5 where they compare the litigation sample with the broad population. Their proportion tests in Table 5 are analogous to our descriptive statistics in Panel A of Table 3.

[8] This argument can be readily seen from the two measures of discontinuity used in Donelson et al. (2013). These two measures are similar in nature and we take one of them, the standardized difference used in Burgstahler and Dichev (1997), for an illustration. To detect whether there is a discontinuity at Bin 0, Donelson et al. (2013) calculate the standardized difference at Bin -1. The standardized difference at Bin -1 is equal to (1) the frequency of observations in Bin -1 minus (2) the expected frequency in Bin -1 which is approximated by the average frequency in the two adjacent bins (Bin 0 and Bin -2), scaled by the standard deviation of *all* differences at all bins in the sample (see Burgstahler & Dichev, 1997, pp. 102–103). Standardized differences across all bins in a sample are distributed asymptotically standard normal. A significantly negative standardized difference at Bin -1 indicates a discontinuity at Bin 0. Based on the above formula, a significantly negative standardized difference at Bin -1 can be due to (i) the frequency in Bin 0 being too high (i.e., benchmark-beating), (ii) the frequency in Bin -1 being too low, or (iii) both.

Table 1
Sample selection.

	Restatement events	Restated firm-quarters
Non-duplicate irregularities announced during January 2000 to June 2006 in the HLM dataset	464	
Irregularities for which restatement periods cannot be identified on the Audit Analytics database or the EDGAR database	(14)	
Duplicate irregularities since referring to the same restatement period	(5)	
Irregularities for which restated firm-quarter observations cannot be identified on the COMPUSTAT Fundamental Quarterly database	(28)	
Subtotal	417	4048
Irregularities with missing values to calculate discretionary accruals		(747)
Irregularities that do not result in restating earnings		(1913)
Irregularities in financial industry		(100)
Irregularities with restatement periods before 1999 or after 2005		(73)
Irregularities with missing values to calculate F-scores in Dechow et al. (2011)		(238)
Total irregularity		977

Notes
The HLM dataset is constructed based on the GAO earnings restatement database with all restatements classified as either errors (unintentional misstatements) or irregularities (intentional misstatements) as in Hennes et al. (2008).

manipulation subsample, but the opposite is true for the non-opportunistic manipulation subsample. These findings suggest that managers in the opportunistic manipulation subsample do, indeed, manipulate earnings for opportunistic reasons and consequently reduce the usefulness of earnings and accrual components whereas managers in the non-opportunistic manipulation subsample appear to manipulate earnings to convey private information. Therefore, opportunistic manipulation is more harmful to a firm's stakeholders than non-opportunistic manipulation, and detecting opportunistic manipulation is potentially more valuable. We divide our irregularity sample into the opportunistic accounting irregularity and non-opportunistic accounting irregularity subsamples and examine the association between benchmark-beating and opportunistic irregularities. Our third research question is stated below:

RQ3. Is benchmark-beating more useful for detecting opportunistic accounting irregularities than accounting irregularities in general?

3. Research design

3.1. Sample selection

Our irregularity sample is based on the General Accounting/Government Accountability Office (GAO) restatement database with all restatements classified as either errors (unintentional misstatements) or irregularities (intentional misstatements) as in Hennes et al. (2008). We refer to this data as the HLM dataset hereafter, which contains 2705 restatements announced between January 1997 and June 2006. Hennes et al. (2008) demonstrate the importance of distinguishing errors from irregularities when studying earnings restatements. Since we investigate the link between benchmark-beating and unequivocal earnings management, it is critically important for us to include only restatements due to intentional misreporting (irregularities). The HLM dataset contains the restatement announcement date but not the period for which earnings are restated (the restatement period). We use the Audit Analytics Non-Reliance Restatement database to identify the beginning and ending dates of each irregularity. For the irregularities on the HLM dataset that cannot be matched with the Audit Analytics database, we manually search the online EDGAR database to identify the restatement period.

Table 1 shows the selection process for the irregularity sample. Our sample includes irregularities announced during January 2000 (the starting date of Audit Analytics Non-Reliance Restatement database) to June 2006 (the ending date of the HML dataset). After deleting restatements due to errors and duplicate observations indicated in the HML dataset, we obtain 464 restatements due to irregularities. We can identify the restatement period for 358 irregularities on the Audit Analytics database.[9] For 106 irregularities that cannot be identified on the Audit Analytics database, we hand collect the restatement period from the online EDGAR database. We are able to find the restatement period for 92 irregularities and lose 14 irregularities. Overall, we find the restatement period for 450 (i.e. 358 + 106–14) irregularities after losing 14 irregularities due to missing restatement period on both the Audit Analytics and the EDGAR databases.

Among these 450 irregularities, we delete five duplicate events (i.e. referring to the same restatement period, although they are not indicated as “duplicate” in the HLM dataset). We then merge the remaining 445 irregularities with the COMPUSTAT Fundamental Quarterly database using the firm identifier (GVKEY) and the restatement period. We lose 28 irregularities due to no match with the COMPUSTAT database and thus obtain a preliminary sample of 417 irregularities with 4048 restated firm-quarter observations. We lose 747 restated firm-quarters due to missing values for calculating performance-adjusted discretionary accruals using the Kothari, Leone, and Wasley (2005) model. Because our objective is to examine *earnings* management behavior, we exclude 1913 restated firm-quarters where there is *no* difference between first-reported net income (COMPUSTAT mnemonic: NIQR) and restated net income (NIQ), i.e., we keep only firm-quarters where earnings are restated (see below for a more detailed discussion of how first-reported and restated earnings are measured). We further delete 100 restated firm-quarters in the financial industry (SIC = 6000–6999) and 73 restated firm-quarters because their restatement periods are either before 1999 or after 2005.[10] Lastly, we delete 238 restated firm-quarters due to missing values for calculating F-scores using the Dechow et al. (2011) model. Our final irregularity sample contains 977 restated firm-quarters ranging from 1999 to 2005.

We construct a control sample of non-restating firms from the COMPUSTAT universe. Specifically, we identify all firm-quarters in the COMPUSTAT database from 1999 to 2005. We then delete irregularity firms or firm-quarter observations where first-reported net income is different from restated net income, i.e., we keep only firm-quarters

[9] We merge the HLM dataset with the Audit Analytics database using the CIK number and the restatement announcement date. The announcement date in the HLM dataset and the filing date in the Audit Analytics database are often different because the former refers to the date when the restatement is announced while the latter refers to the date when the restatement is filed with the SEC. We allow a three-month difference in the announcement date (the HLM dataset) and the filing date (the Audit Analytics database) when merging the two databases. To assure that these two dates refer to the same earnings restatement event, we randomly select ten restatements where the announcement date and filing date fall within three months of each other and manually compare these two dates with the information from the EDGAR database. We find that these two dates refer to the same restatement event in all ten cases.

[10] These 73 restated firm-quarters are distributed as follows: 1, 17, 53, and 2 in years 1996, 1997, 1998, and 2006, respectively.

Table 2
Frequency distributions of the irregularity and control samples.

Panel A: Distribution of restated firm-quarter observations across years

Fiscal year	Irregularity sample		Control sample		Irregularity/(Irregularity + Control)
	Number	Percent	Number	Percent	Percent
1999	110	11.26%	16646	16.32%	0.66%
2000	121	12.38%	16645	16.32%	0.72%
2001	134	13.72%	15614	15.31%	0.85%
2002	158	16.17%	14088	13.81%	1.11%
2003	178	18.22%	13205	12.95%	1.33%
2004	196	20.06%	12894	12.64%	1.50%
2005	80	8.19%	12894	12.64%	0.62%
Total (Average)	977	100%	101986	100%	0.95%

Panel B: Distribution of restated firm-quarter observations across industries

Industry	Irregularity sample		Control sample		Irregularity/(Irregularity + Control)
	Number	Percent	Number	Percent	Percent
Business service	191	19.55%	16363	16.04%	1.15%
Computers	80	8.19%	7503	7.36%	1.05%
Electronic equipment	73	7.47%	5461	5.35%	1.32%
Food products	72	7.37%	5819	5.71%	1.22%
Machinery	61	6.24%	3674	3.60%	1.63%
Miscellaneous	60	6.14%	6772	6.64%	0.88%
Pharmaceutical products	58	5.94%	4593	4.50%	1.25%
Retail	43	4.40%	3715	3.64%	1.14%
Telecommunications	32	3.28%	1869	1.83%	1.68%
Wholesale	27	2.76%	1506	1.48%	1.76%
Others	280	28.66%	44711	43.84%	0.62%
Total (Average)	977	100%	101986	100%	0.95%

Notes
The irregularity sample consists of 977 restated firm-quarters during fiscal year 1999 to 2005. The control sample consists of 101986 observations from COMPUSTAT for the same period where earnings are not restated. Industry is classified based on the 48-industry classification scheme in Fama and French (1997). The category of "Others" is the sum of other 38 industries.

where earnings are *not* restated. In order to exclude the effects of outliers of the key variables from extremely small firms, we further exclude all firm-quarter observations with average total assets < 2.5 million dollars or with a closing price-per-share at the fiscal quarter end < 0.1 dollar.[11] Finally, we delete observations in the financial industry, observations with missing values needed for estimating performance-adjusted discretionary accruals and F-scores, or observations with less than ten observations in any two-digit SIC code and year combination. Our control sample consists of 101986 firm-quarter observations. However, the actual sample size used for each test could be slightly smaller due to missing other required variables.

We report the frequency distribution of observations in the irregularity and control samples across years and industries, respectively, in Table 2. Panel A of Table 2 shows a pattern of increasing numbers of earnings restatements from 1999 to 2004. Note that the decrease in the number of restatements in 2005 does not mean a decrease in irregularities in 2005. Rather, it means that our irregularity sample in 2005 (80 firm-quarters) is incomplete because we limit our irregularity sample to irregularities announced before July 2006 as in the HLM dataset.[12] In contrast, our control sample decreases in size over time. On average, the irregularity sample is slightly < 1% of the control sample.

Panel B of Table 2 presents the frequency distribution of irregularity firms and control firms in different industries. We use the 48-industry classification scheme in Fama and French (1997) and report the top ten industries where irregularities occur most frequently with the other 38 industries aggregated in the category of "Others." Business service, computers, electronic equipment, food products, and machinery are the top five industries ranked according to the frequency of restated firm-quarters in the irregularity sample. On the other hand, wholesale, telecommunications, machinery, electronic equipment, and pharmaceutical products are the top five industries ranked according to the relative frequency of restated firm-quarters with respect to the combined irregularity and control samples.

3.2. Variable measurement

For the irregularity sample, we measure earnings as *first-reported* earnings. A firm's first-reported earnings are not reported in standard COMPUSTAT databases and were difficult to obtain in the past. Recently, however, COMPUSTAT, in conjunction with Charter Oak Investment Systems Inc., has developed backtest datasets that contain a firm's originally reported financial data (Hollie, Livnat, & Segal, 2012; Livnat & Mendenhall, 2006). We obtain an irregularity firm's first-reported earnings (*EARN*) from the COMPUSTAT Unrestated Quarterly database (Compustat mnemonic: NIQR).[13] For the control sample, we define earnings (*EARN*) as the bottom-line net income (NIQ) from the COMPUSTAT Fundamental Quarterly database.

Following prior literature (e.g. Burgstahler & Dichev, 1997; Degeorge et al., 1999; Frankel et al., 2002), we measure whether

[11] All firms in our irregularity sample satisfy this requirement.

[12] Untabulated analyses suggest that it takes about two years on average for an intentional misstatement to be restated and filed with SEC. That is, more announcements of irregularities after June 2006 will restate earnings in year 2005.

[13] COMPUSTAT adds a suffix "R" in its mnemonic to indicate that variable is the as-first-reported value. For example, NIQ is the updated or restated net income from the Fundamental Quarterly database and NIQR is the as-first-reported net income from the Unrestated Quarterly database.

reported earnings meet or beat three commonly used earnings benchmarks: (1) the earnings change benchmark or earnings in the same quarter last year, (2) the earnings level benchmark or zero earnings, and (3) the earnings forecast benchmark or consensus analyst earnings forecasts. We first define benchmark-beating for the first two benchmarks. Specifically, we calculate seasonal earnings changes (*CHG*) and earnings levels (*LVL*) as follows: $CHG = (EARN_t - EARN_{t-4})/MKTCAP_{t-1}$, where $MKTCAP$ = PRCCQ (stock price) × CSHOQ (common shares outstanding) at the beginning of quarter *t*, and $LVL = EARN_t/MKTCAP_{t-1}$. We then define BMK^{CHG} and BMK^{LVL} as indicator variables equal to 1 if earnings meet or slightly beat the earnings change benchmark (i.e., $0 \le CHG < 1\%$) and the earnings level benchmark (i.e., $0 \le LVL < 1\%$), and 0 otherwise.

For the third earnings benchmark, we obtain analyst earnings forecasts and actual earnings from the I/B/E/S database, unadjusted for stock splits and dividends to avoid rounding errors and a look-ahead bias in the I/B/E/S split-adjusted data (Payne & Thomas, 2003). We calculate earnings surprises per share (*SURP*) as follows: *SURP* = *EPS_ACT* – *EPS_MED*, where *EPS_ACT* is the actual earnings per share and *EPS_MED* is the median analyst earnings forecast in the month immediately before an earnings announcement. We then define BMK^{FCST} as an indicator variable for meeting or slightly beating the analyst forecast benchmark, i.e., BMK^{FCST} equals one if *SURP* = 0¢ or 1¢ and zero otherwise.

We now describe the estimation of performance-adjusted discretionary accruals. For the control sample, we estimate performance-adjusted discretionary accruals using the following cross-sectional Kothari et al. (2005) model:

$$TACC = \alpha_0 + \alpha_1 Q2 + \alpha_2 Q3 + \alpha_3 Q4 + \alpha_4(\Delta REV - \Delta AR) + \alpha_5 PPE + \alpha_6 ROA + \varepsilon, \tag{1}$$

where *TACC* is the total accruals; *Q2* – *Q4* are quarter dummies for the second, third, and fourth quarter, respectively; *ΔREV* and *ΔAR* are the changes in sales and in accounts receivable, respectively, from the previous quarter to the current quarter; *PPE* is the gross property, plant, and equipment; and *ROA* is return on assets. See Appendix A for the details of variable definitions. We estimate Eq. (1) for each two-digit SIC code and year combination with at least 10 observations using the control sample, because we do not want irregularity firms to distort the regression coefficients. Discretionary accruals for the control sample (*DA*) are defined as the residual of Eq. (1).

For the irregularity sample, we estimate discretionary accruals (*DA*) using first-reported values as follows:

$$DA = TACC - [\hat{\alpha}_0 + \hat{\alpha}_1 \mathrm{Q2} + \hat{\alpha}_2 \mathrm{Q3} + \hat{\alpha}_3 \mathrm{Q4} + \hat{\alpha}_4(\Delta REV - \Delta AR) + \hat{\alpha}_5 \mathrm{PPE} + \hat{\alpha}_6 \mathrm{ROA}], \tag{2}$$

where $\hat{\alpha}_0 - \hat{\alpha}_6$ are the coefficient estimates from estimating Eq. (1) in the cross-section using the control sample, and all accounting variables are defined the same as in Eq. (1) but are measured using the first-reported values from the COMPUSTAT Unrestated Quarterly database.

Following Dechow et al. (2011), who developed F-scores, we first estimate the following model using our combined irregularity and control samples:

$$MISSTMT = \beta_0 + \beta_1 RSST_{ACC} + \beta_2 CH_{REC} + \beta_3 CH_{INV} + \beta_4 SOFT_{ASSET} + \beta_5 CH_{CS} + \beta_6 CH_{ROA} + \beta_7 ISSUE + \varepsilon, \tag{3}$$

where *MISSTMT* = 1 for observations from the irregularity sample and 0 for observations from the control sample. See Appendix A for definitions of other variables. Eq. (3) follows Model 1 of Dechow et al. (2011, p. 55).[14]

For the control sample, we calculate the predicted probability of misstatement using the coefficient estimates ($\hat{\beta}_0$ – $\hat{\beta}_7$) from estimating Eq. (3) as:

$$\text{Probability} = \frac{e^{PredictedValue}}{(1 + e^{PredictedValue})}, \text{ where} \tag{4}$$

$$Predicted\ Value = \hat{\beta}_0 + \hat{\beta}_1 \mathrm{RSST_{ACC}} + \hat{\beta}_2 \mathrm{CH_{REC}} + \hat{\beta}_3 \mathrm{CH_{INV}} + \hat{\beta}_4 SOFT_{ASSET} + \hat{\beta}_5 CH_{CS} + \hat{\beta}_6 CH_{ROA} + \hat{\beta}_7 ISSUE. \tag{5}$$

The F-score (*FSCORE*) is the scaled predicted probability of misstatement defined as:

$$FSCORE = \frac{Probability}{Unconditional\ Expectation\ of\ Misstatement}, \tag{6}$$

where *Unconditional Expectation of Misstatement* is the number of observations in the irregularity sample divided by the total number of observations in the irregularity sample and the control sample. Larger values of F-scores indicate higher probabilities of misstatements.

For the irregularity sample, we calculate the first-reported F-score (*FSCORE*) using the same procedure as above except accounting variables in Eq. (5) are all measured using the first-reported values.

4. Empirical results

4.1. Descriptive statistics

Panel A of Table 3 presents descriptive statistics for the irregularity sample (*MISSTMT* = 1) and control sample (*MISSTMT* = 0), respectively. We first compare BMK^{CHG}, BMK^{LVL}, and BMK^{FCST} between the irregularity sample and the control sample. As shown, the percentage of misreporting firms that meet or slightly beat three earnings benchmarks are 31.43%, 25.57%, and 33.50%, respectively, which are significantly higher than their respective counterparts for non-misreporting firms (25.50%, 17.06%, and 29.15%). This is consistent with the proportion test results in Table 5 of Donelson et al. (2013), suggesting that misreporting firms are more likely to meet or beat earnings benchmarks than non-misreporting firms. The mean (median) *DA* of the misreporting sample is significantly larger than the mean (median) *DA* of the non-misreporting sample, consistent with Jones et al. (2008) who find that discretionary accruals are positively related to misstatements. Similarly, we also find that the mean (median) *FSCORE* of the misreporting sample is significantly larger than the mean (median) *FSCORE* of the non-misreporting sample, consistent with a larger F-score indicating higher probability of earnings misstatements. Also consistent with Jones et al. (2008), the mean and median assets (*AVGAT*) of misreporting firms are significantly larger than those of non-misreporting firms. In contrast to Jones et al. (2008), however, we find that the median *ROA* of misreporting firms is significantly smaller than that of non-misreporting firms and that the mean *BigN* is significantly smaller for the misreporting sample than the non-misreporting sample.

We separate the full sample into the benchmark-beating sample (*BMK* = 1) and non-benchmark-beating sample (*BMK* = 0) for each of our three benchmarks (BMK^{CHG}, BMK^{LVL}, and BMK^{FCST}) and calculate the percent of observations that intentionally misreport earnings for benchmark beaters and non-benchmark beaters, respectively. Panel B of Table 3 reports our findings. For the earnings change benchmark (BMK^{CHG}), 25,656 (74,737) observations are benchmark beaters (non-benchmark beaters). We find that 1.18% of the benchmark beaters intentionally misreport earnings whereas only 0.88% of the non-benchmark beaters intentionally misreport earnings. The difference between these two percentages (0.30%) is significant at the 0.01 level. The above findings suggest that benchmark beaters are 34.09% (=

[14] We choose Model 1, instead of Models 2 and 3, for several reasons. First, Dechow et al. (2011) show that Model 1 consistently outperforms the other models in identifying restating firms. Second, Model 1 requires the least number of variables. Finally, Model 1 is robust to variations in sample or time period (Dechow et al., 2011).

Table 3
Descriptive statistics.

Panel A: Irregularity sample vs. control sample

Variable	Irregularity sample (*MISSTMT* = 1)						Control sample (*MISSTMT* = 0)						Difference in	
	n	Mean	Std Dev	Q1	Median	Q3	n	Mean	StdDev	Q1	Median	Q3	Mean	Median
BMK^{CHG}	964	31.43%	46.45%	0.00%	0.00%	100%	99,429	25.50%	43.59%	0.00%	0.00%	100%	5.93%*	0.00%*
BMK^{LVL}	966	25.57%	43.65%	0.00%	0.00%	100%	99,951	17.06%	37.61%	0.00%	0.00%	0.00%	8.51%*	0.00%*
BMK^{FCST}	588	33.50%	47.24%	0.00%	0.00%	100%	47,618	29.15%	45.45%	0.00%	0.00%	100%	4.35%†	0.00%†
DA	977	1.09%	8.31%	−2.43%	1.05%	4.09%	101,986	0.05%	10.48%	−4.41%	0.01%	4.01%	1.04%*	1.04%*
FSCORE	977	1.22	0.46	0.83	1.20	1.59	101,986	1.00	0.47	0.60	0.92	1.34	0.22*	0.28*
LEV	977	25.54%	21.30%	8.26%	23.22%	37.90%	101,986	23.78%	26.10%	1.59%	18.02%	36.35%	1.77%*	5.20%*
ROA	977	−2.02%	8.82%	−2.10%	0.23%	1.27%	101,986	−2.88%	10.51%	−3.49%	0.38%	1.81%	0.85%†	−0.15%
AVGAT	977	3359.36	7410.98	102.69	464.13	1849.19	101,986	1141.91	4199.84	26.34	114.45	534.59	2217.45*	349.68*
BigN	977	62.44%	48.45%	0.00%	100%	100%	101,986	71.03%	45.36%	0.00%	100%	100%	−8.60%*	0.00%*

Panel B: Benchmark-beating sample vs. non-benchmark-beating sample

Variable	Benchmark-beating sample (*BMK* = 1)		Non-benchmark-beating sample (*BMK* = 0)		Difference in	
	n	Mean *MISSTMT* (1)	n	Mean *MISSTMT* (2)	Mean (1)–(2)	Mean Percent [(1)–(2)]/(2)
BMK^{CHG}	25,656	1.18%	74,737	0.88%	0.30%*	34.09%
BMK^{LVL}	17,296	1.43%	83,621	0.86%	0.57%*	66.28%
BMK^{FCST}	14,078	1.40%	34,128	1.15%	0.25%†	21.74%

Notes
See Appendix A for variable definitions. The irregularity sample consists of 977 firm-quarter observations in fiscal years 1999 to 2005 while the control sample consists of 101,986 observations from the COMPUSTAT Fundamental Quarterly database for the same period where earnings are not restated. The sample size for some variables is smaller due to additional missing values.
‡, †, and * denote significance at the 10%, 5% and 1% level, respectively, in a two-tailed *t*-test (for difference in mean) or a two-tailed Wilcoxon z-test (for difference in median).

$\frac{1.18\,\% - 0.88\,\%}{0.88\,\%}$) more likely to intentionally misreport earnings than non-benchmark beaters. For the earnings level benchmark (BMK^{LVL}) and earnings forecast benchmark (BMK^{FCST}), benchmark beaters are 66.28% and 21.74%, respectively, more likely to intentionally misreport earnings than non-benchmark beaters. Overall, Panel B of Table 3 provides univariate evidence that benchmark-beating is positively associated with the probability of intentional misreporting.

Table 4 presents correlations between the key variables for the combined irregularity and control samples. As shown, *MISSTMT* is positively correlated with meeting or slightly beating all three benchmarks, consistent with the finding in Panel B of Table 3 that benchmark beaters are more likely to intentionally misreport than non-benchmark beaters. Similarly, *MISSTMT* is significantly positively correlated with both discretionary accruals (*DA*) and with F-scores (*FSCORE*). It is interesting to note that the Pearson correlation between *MISSTMT* and BMK^{CHG} is larger than that between *MISSTMT* and *DA* but is smaller than that between *MISSTMT* and *FSCORE*. The same pattern is true for BMK^{LVL} and BMK^{FCST}. These correlations seem to suggest that benchmark-beating may outperform discretionary accruals but may be dominated by F-scores for predicting intentional misreporting (*MISSTMT*).

4.2. Benchmark-beating and accounting irregularities

To investigate our first research question (RQ1) of whether benchmark-beating is positively associated with accounting irregularities and our second research question (RQ2) of whether benchmark-beating is useful for detecting accounting irregularities both incremental and relative to discretionary accruals and F-scores, we estimate the following logistic model using the combined irregularity and control samples:

$$MISSTMT = \gamma_0 + \gamma_1 BMK + \gamma_2 DA + \gamma_3 FSCORE + \gamma_4 LEV + \gamma_5 ROA + \gamma_6 AVGAT + \gamma_7 BigN + \varepsilon, \quad (7)$$

where *MISSTMT* is a dummy variable for accounting irregularity as defined earlier; *BMK* is a dummy variable for meeting or slightly beating an earnings benchmark and is equal to BMK^{CHG}, BMK^{LVL}, or BMK^{FCST}; and *DA* and *FSCORE* are discretionary accruals and F-scores, as defined earlier.

Following Jones et al. (2008), we include several control variables in Eq. (7). *LEV* is financial leverage and we control for leverage because it is more likely for a firm to manage earnings when its leverage is high. However, Jones et al. (2008) find no significant relation between misstatements and leverage. *ROA* is return on assets. Jones et al. (2008) find a positive relation between *MISSTMT* and *ROA* in some regressions. *AVGAT* is average assets to control for the size effect. Jones et al. (2008) find that large firms are more likely to misstate earnings. *BigN* is a dummy variable for Big 4, Big 5, or Big 6 auditors. Jones et al. (2008) find a negative relation between misstatements and *BigN*. See Appendix A for definitions of all variables.

Our sample consists of panel data with firm-year observations. Petersen (2009) finds that it is likely that regression residuals using financial and accounting panel data are correlated across years for a given firm (a firm effect) or correlated across firms for a given year (a time effect). Petersen (2009) shows that regressions controlling for two-way clustering effects are the most robust and control well for a time effect, a firm effect, or both. Following Petersen (2009), we estimate our regression models using two-way clustering by firms and years.

We estimate Eq. (7) using the combined irregularity and control samples. Table 5 reports our findings. Panel A shows our findings for the earnings change benchmark (BMK^{CHG}). We first estimate Eq. (7) excluding *DA* and *FSCORE* in order to examine how benchmark-beating, alone, is associated with accounting irregularities. This addresses our first research question (RQ1) of whether benchmark-beating firms are more likely to perpetrate intentional misreporting than non-benchmark-beating firms. The coefficient on BMK^{CHG} is highly significantly positive (0.35, *p*-value = 0.00), suggesting that firms that beat their earnings change benchmark by a small margin, compared to firms that do not beat this benchmark, are 41.91% (= $e^{0.35}$–1) more likely to intentionally misreport after controlling for several common

Table 4
Correlations.

Variable	*MISSTMT*	BMK^{CHG}	BMK^{LVL}	BMK^{FCST}	*DA*	*FSCORE*	*LEV*	*ROA*	*AVGAT*	*BigN*
MISSTMT		0.0133 [< 0.0001]	0.0219 [< 0.0001]	0.0105 [0.0208]	0.0096 [0.0021]	0.0450 [< 0.0001]	0.0066 [0.0349]	0.0079 [0.0116]	0.0506 [< 0.0001]	−0.0184 [< 0.0001]
BMK^{CHG}	0.0133 [< 0.0001]		0.1453 [< 0.0001]	0.1271 [< 0.0001]	−0.0512 [< 0.0001]	0.0707 [< 0.0001]	−0.1037 [< 0.0001]	0.2055 [< 0.0001]	0.0820 [< 0.0001]	0.1256 [< 0.0001]
BMK^{LVL}	0.0219 [< 0.0001]	0.1453 [< 0.0001]		0.1064 [< 0.0001]	−0.0583 [< 0.0001]	0.0209 [< 0.0001]	−0.0977 [< 0.0001]	0.1748 [< 0.0001]	0.0513 [< 0.0001]	0.0735 [< 0.0001]
BMK^{FCST}	0.0105 [0.0208]	0.1271 [< 0.0001]	0.1064 [< 0.0001]		−0.0382 [< 0.0001]	0.0638 [< 0.0001]	−0.0607 [< 0.0001]	0.0815 [< 0.0001]	0.0031 [0.4980]	0.0263 [< 0.0001]
DA	0.0150 [< 0.0001]	−0.0650 [< 0.0001]	−0.0623 [< 0.0001]	−0.0391 [< 0.0001]		−0.0037 [0.2302]	−0.0051 [0.1028]	−0.1540 [< 0.0001]	−0.0262 [< 0.0001]	0.0057 [0.0661]
FSCORE	0.0455 [< 0.0001]	0.0790 [< 0.0001]	0.0305 [< 0.0001]	0.0668 [< 0.0001]	−0.0193 [< 0.0001]		0.0959 [< 0.0001]	0.0914 [< 0.0001]	0.0335 [< 0.0001]	0.0065 [0.0362]
LEV	0.0178 [< 0.0001]	−0.0815 [< 0.0001]	−0.0894 [< 0.0001]	−0.0623 [< 0.0001]	0.0084 [0.0072]	0.1739 [< 0.0001]		−0.0909 [< 0.0001]	0.0514 [< 0.0001]	−0.0575 [< 0.0001]
ROA	−0.0046 [0.1378]	0.3471 [< 0.0001]	0.1709 [< 0.0001]	0.0970 [< 0.0001]	−0.2068 [< 0.0001]	0.1458 [< 0.0001]	−0.0396 [< 0.0001]		0.0948 [< 0.0001]	0.1471 [< 0.0001]
AVGAT	0.0577 [< 0.0001]	0.2265 [< 0.0001]	0.1360 [< 0.0001]	0.0171 [0.0002]	−0.0552 [< 0.0001]	0.1483 [< 0.0001]	0.2290 [< 0.0001]	0.3246 [< 0.0001]		0.1069 [< 0.0001]
BigN	−0.0184 [< 0.0001]	0.1256 [< 0.0001]	0.0735 [< 0.0001]	0.0263 [< 0.0001]	−0.0068 [0.0284]	0.0210 [< 0.0001]	−0.0279 [< 0.0001]	0.1391 [< 0.0001]	0.4476 [< 0.0001]	

Notes
See Appendix A for definitions of all variables. Pearson (Spearman) correlations are above (below) the diagonal. Correlations are calculated using the combined irregularity and control samples. Numbers in brackets are the significance levels of the correlations.

Table 5
Two-way Clustering Logistic Regressions of MISSTMT on Measures of Benchmark-Beating.

Panel A: the earning change benchmark ($BMK = BMK^{CHG}$)

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−4.53 (0.00)	−4.54 (0.00)	−5.46 (0.00)	−5.41 (0.00)	−5.47 (0.00)
BMK^{CHG}	0.35 (0.00)	0.36 (0.00)	0.29 (0.01)		0.30 (0.01)
DA		0.53 (0.00)		0.51 (0.01)	0.54 (0.01)
FSCORE			0.88 (0.00)	0.89 (0.00)	0.88 (0.00)
LEV	0.26 (0.16)	0.26 (0.16)	0.09 (0.66)	0.04 (0.85)	0.09 (0.66)
ROA	−0.03 (0.94)	−0.04 (0.92)	−0.13 (0.62)	−0.07 (0.85)	−0.16 (0.58)
AVGAT/1000	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
BigN	−0.46 (0.01)	−0.45 (0.01)	−0.44 (0.01)	−0.40 (0.01)	−0.44 (0.01)
No. of irregularity obs.	964	964	964	964	964
No. of control obs.	99,429	99,429	99,429	99,429	99,429
Pseudo R^2	0.0105	0.0112	0.0274	0.0266	0.0281

1. R^2 Decomposition
   Incremental R^2 by BMK^{CHG}: $R^2_{Model\ 5} - R^2_{Model\ 4}$ = 0.0281–0.0266 = 0.0015
   Incremental R^2 by *DA*: $R^2_{Model\ 5} - R^2_{Model\ 3}$ = 0.0281–0.0274 = 0.0007
   Incremental R^2 by *FSCORE*: $R^2_{Model\ 5} - R^2_{Model\ 2}$ = 0.0281–0.0112 = 0.0169

2. Vuong (1989) Test
   Model 3 (BMK^{CHG}) vs. Model 4 (*DA*): Vuong's z-statistic = 1.10 (0.27)
   Model 2 (BMK^{CHG}) vs. Model 4 (*FSCORE*): Vuong's z-statistic = −6.00 (0.00)

Panel B: the earnings level benchmark ($BMK = BMK^{LVL}$)

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−4.57 (0.00)	−4.58 (0.00)	−5.52 (0.00)	−5.41 (0.00)	−5.53 (0.00)
BMK^{LVL}	0.57 (0.00)	0.58 (0.00)	0.55 (0.00)		0.56 (0.00)
DA		0.56 (0.00)		0.51 (0.01)	0.57 (0.01)
FSCORE			0.89 (0.00)	0.89 (0.00)	0.88 (0.00)
LEV	0.29 (0.11)	0.29 (0.11)	0.13 (0.53)	0.05 (0.82)	0.13 (0.53)
ROA	−0.05 (0.90)	−0.08 (0.86)	−0.15 (0.55)	−0.05 (0.90)	−0.18 (0.50)
AVGAT/1000	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
BigN	−0.44 (0.01)	−0.44 (0.01)	−0.43 (0.01)	−0.39 (0.02)	−0.42 (0.01)
No. of irregularity obs.	966	966	966	966	966
No. of control obs.	99,951	99,951	99,951	99,951	99,951
Pseudo R^2	0.0131	0.0138	0.0302	0.0264	0.0309

1. R^2 Decomposition
   Incremental R^2 by BMK^{LVL}: $R^2_{Model\ 5} - R^2_{Model\ 4}$ = 0.0309–0.0264 = 0.0045
   Incremental R^2 by *DA*: $R^2_{Model\ 5} - R^2_{Model\ 3}$ = 0.0309–0.0302 = 0.0007
   Incremental R^2 by *FSCORE*: $R^2_{Model\ 5} - R^2_{Model\ 2}$ = 0.0309–0.0138 = 0.0171

2. Vuong (1989) Test
   Model 3 (BMK^{LVL}) vs. Model 4 (*DA*): Vuong's z-statistic = 2.76 (0.01)

(*continued on next page*)

Table 5 (*continued*)

Model 2 (BMK^{LVL}) vs. Model 4 (*FSCORE*):	Vuong's z-statistic = −4.51 (0.00)

Panel C: the earnings forecast benchmark ($BMK = BMK^{FCST}$)

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−3.81 (0.00)	−3.82 (0.00)	−4.76 (0.00)	−4.71 (0.00)	−4.76 (0.00)
BMK^{FCST}	0.27 (0.00)	0.28 (0.00)	0.22 (0.01)		0.23 (0.01)
DA		1.29 (0.05)		1.06 (0.06)	1.11 (0.06)
FSCORE			0.86 (0.00)	0.86 (0.00)	0.85 (0.00)
LEV	0.89 (0.00)	0.88 (0.00)	0.75 (0.01)	0.72 (0.02)	0.74 (0.01)
ROA	−0.59 (0.34)	−0.34 (0.65)	−1.06 (0.05)	−0.86 (0.16)	−0.92 (0.13)
AVGAT/1000	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
BigN	−1.16 (0.00)	−1.15 (0.00)	−1.13 (0.00)	−1.11 (0.00)	−1.12 (0.00)
No. of irregularity obs.	588	588	588	588	588
No. of control obs.	47,618	47,618	47,618	47,618	47,618
Pseudo R^2	0.0296	0.0303	0.0433	0.0423	0.0439

1. R^2 Decomposition	
Incremental R^2 by BMK^{FCST}:	$R^2_{Model\ 5} - R^2_{Model\ 4}$ = 0.0439–0.0423 = 0.0016
Incremental R^2 by *DA*:	$R^2_{Model\ 5} - R^2_{Model\ 3}$ = 0.0439–0.0433 = 0.0006
Incremental R^2 by *FSCORE*:	$R^2_{Model\ 5} - R^2_{Model\ 2}$ = 0.0439–0.0303 = 0.0136
2. Vuong (1989) Test	
Model 3 (BMK^{FCST}) vs. Model 4 (*DA*):	Vuong's z-statistic = 0.42 (0.68)
Model 2 (BMK^{FCST}) vs. Model 4 (*FSCORE*):	Vuong's z-statistic = −4.16 (0.00)

Notes
See Appendix A for definitions of all variables. Eq. (7) is estimated using the combined irregularity and control samples. $BMK = BMK^{CHG}$, BMK^{LVL}, and BMK^{FCST} in Panel A, Panel B, and Panel C, respectively. Numbers in parentheses beside coefficient estimates are *p*-values of a two-tailed test based on the two-way clustered standard errors (Petersen, 2009). Numbers in parentheses under Vuong (1989) Test are *p*-values of a two-tailed test for Vuong's z-statistic.

factors that also affect the probability of misreporting as found in Jones et al. (2008).[15] Panel B of Table 3 shows that firms that that meet or beat their earnings change benchmark (BMK^{CHG}), compared to firms that do not meet or beat this benchmark, are 34.09% more likely to intentionally misreport. So, our multivariate estimate, 41.91%, is comparable to its univariate counterpart, 34.09%.

Panels B (Panel C) of Table 5 shows that the coefficient on BMK^{LVL} (BMK^{FCST}) in Model 1 are 0.57 (0.27), suggesting that firms beating the earnings level (earnings forecast) benchmark are 76.83% (31.00%) more likely to engage in earnings manipulation than firms not beating the earnings level (earnings forecast) benchmark. These two percentages based on multivariate estimation are in line with their univariate counterparts in Panel B of Table 3 (66.28% and 21.74%). In sum, the answer to our first research question (RQ1) is that benchmark-beating firms are substantially more likely to intentionally misreport earnings than non-benchmark-beating firms.

4.3. Incremental and relative detective power of benchmark-beating for accounting irregularities

Our second research question (RQ2) asks whether benchmark-beating is useful for detecting accounting irregularities both incremental and relative to discretionary accruals and F-scores. We first evaluate the incremental contribution of benchmark-beating. Models 2, 3 and 5 in Panels A, B, and C of Table 5 show our findings. In particular, Model 2 examines whether our benchmark-beating variables are useful incremental to discretionary accruals in explaining the probability of accounting irregularities. As shown, the coefficient on BMK^{CHG} in Panel A, that on BMK^{LVL} in Panel B, and that on BMK^{FCST} in Panel C remain highly significantly positive in the presence of *DA*. The coefficients on *DA* in three panels are also highly significantly positive. In addition, the inclusion of *DA* in Model 2 detracts little from the significance of our benchmark-beating variables because the coefficients on BMK^{CHG}, BMK^{LVL}, and BMK^{FCST} in Model 2 of all three panels are not much reduced compared to their respective counterparts in Model 1. Thus, benchmark-beating and discretionary accruals offer explanatory power incremental to each other for explaining the probability of accounting irregularities and they appear to capture different aspects of accounting irregularities. Similarly, Model 3 examines whether our benchmark-beating variables are useful incremental to F-scores in explaining the probability of accounting irregularities. The coefficients on BMK^{CHG}, BMK^{LVL}, and BMK^{FCST} in Panel A, Panel B, and Panel C, respectively, remain highly significantly positive in the presence of powerful F-scores, indicating that benchmark-beating captures additional, important aspects of accounting irregularities over and beyond F-scores.[16] Model 5 examines the incremental explanatory power of benchmarking-beating in the combined presence of *DA* and *FSCORE*. As we can see, the coefficients on the benchmark variables in all three panels are significantly positive (e.g., 0.30 and *p*-value = 0.01 in Panel A of Table 5) and are comparable in magnitude to their respective counterparts in Model 1; the significance levels of the three benchmark variables are hardly reduced, when compared to Model 1, after including *DA* and *FSCORE* in Model 5 in Panel A, Panel B, and Panel C. Moreover, discretionary accruals and F-scores are also highly significantly positive in all three panels. Taking Model 2, Model 3, and Model 5 together, the answer to our second research question (RQ2) is that the three benchmark-beating variables are all useful in detecting accounting irregularities *incremental* to discretionary accruals alone, to F-scores alone, and to discretionary accruals and F-scores combined.

We next examine the explanatory power of benchmark-beating relative to discretionary accruals or relative to F-scores in a one-on-one horse race. We start with an R^2 decomposition of Model 5 with respect to Model 4, Model 3, or Model 2. In Panel A, Table 5, Model 5 contains an extra explanatory variable, BMK^{CHG}, compared with Model 4. Thus, the incremental R^2 contributed by BMK^{CHG} is the R^2 of Model 5 minus the R^2 of Model 4. We can similarly calculate the incremental R^2 of *DA* and *FSCORE*. Panel A reports that the incremental R^2s contributed by BMK^{CHG}, *DA*, and *FSCORE* are 0.0015, 0.0007, and 0.0169, respectively. These incremental R^2s tentatively suggest that BMK^{CHG} is more important than *DA* but less important than *FSCORE* in explaining accounting irregularities.

We conduct the Vuong (1989) test to formally assess the relative importance of our benchmark-beating variables vs. discretionary accruals and vs. F-scores in explaining accounting irregularities. The Vuong (1989) test is also valid in a logistic regression framework (Hillegeist, Keating, Cram, & Lundstedt, 2004). In Table 5, Model 3 contains our benchmark-beating variables but no discretionary accruals whereas Model 4 contains discretionary accruals but no benchmark-beating variables. A comparison between Model 3 and Model 4, thus, represents a one-on-one horse race between the benchmark-beating variables and discretionary accruals. Similarly, a comparison between Model 2 and Model 4 represents a one-on-one horse race between the benchmark-beating variables and F-scores. Panel A, Table 5, shows that the R^2 of Model 3 is 0.0274, larger than the R^2 of Model 4, 0.0266. However, the Vuong (1989) test suggests that the R^2 of Model 3 (0.0274) is not significantly larger than the R^2 of Model 4 (0.0266) because Vuong's z-statistic is 1.10 (*p*-value = 0.27), failing to reject the null hypothesis that BMK^{CHG} has explanatory power equal to *DA* for accounting irregularities. Thus, BMK^{CHG} ties with *DA*. In the comparison between Model 2 and Model 4, Vuong's z-statistic is −6.00 (*p*-value = 0.00). Thus, BMK^{CHG} is outperformed by *FSCORE* for detecting accounting irregularities. Turning to the other two benchmarks, Panel B, Table 5, shows that BMK^{LVL} outperforms *DA* (Vuong's z-statistic = 2.76, *p*-value = 0.01) but is again outperformed by *FSCORE* (Vuong's z-statistic = −4.51, *p*-value = 0.00). Panel C shows that BMK^{FCST} ties with *DA* but is again outperformed by *FSCORE*. In short, the answer to our second research question (RQ2) is that two (one) of our benchmark-beating variables tie with (outperforms) discretionary accruals but all three benchmark-beating variables are outperformed by F-scores in a one-on-one horse race.

We note that the superior performance of *FSCORE* is expected because *FSCORE* is calibrated from an accounting irregularity prediction model (see Eq. (3)) where accounting irregularities (*MISSTMT*), the dependent variable, are regressed on a dozen of explanatory variables. This estimation and the estimation of discretionary accruals may well be beyond the capabilities of average investors. In contrast, benchmark-beating is a very simple metric, which even naïve investors can calculate and use. The fact that benchmark-beating is still significant in the presence of the powerful F-score suggests that it captures different aspects of accounting irregularities than F-scores. Moreover, benchmark-beating is timelier than F-scores. Thus, it is inappropriate to conclude that benchmark-beating is not useful for detecting accounting irregularity given F-scores.

We now briefly discuss the control variables in Table 5. The coefficient on *LEV* is insignificant in some specifications (Panels A and B) but significantly positive in others (Panel C). Jones et al. (2008), Table 6 find an insignificant coefficient on *LEV* in all specifications. Our coefficient on *ROA* is mostly insignificant but significantly negative in Model 3 of Panel C, Table 5. In contrast, the coefficients on *ROA* in Jones et al. (2008) are mostly insignificant but are significantly positive in some specifications. Our coefficients on *AVGAT* (*BigN*) are all significantly positive (negative), consistent with Jones et al. (2008).

[15] Our coefficient on BMK^{CHG} is 0.35 (Table 5, Panel A, Model 1), indicates that the ratio of the odds of accounting irregularities for firms that beat the earnings change benchmark to the odds for firms that do not beat the benchmark is $e^{0.35}$ (= 1.4191); in another word, firms that meet or slightly beat the earnings change benchmark are 41.91% more likely to engage in earnings manipulation than firms that do not beat the earnings change benchmark. See Hosmer and Lemeshow (1989) and Lee, Ingram, and Howard (1999, p. 780) for how to interpret the coefficients in logistic regressions.

[16] This perhaps is expected because benchmark-beating is not one of the inputs for the estimation of F-scores.

Table 6
Construction of the OP-IRREG Subsample and the OP-CTRL Subsample.

Panel A: the earnings change benchmark ($BMK = BMK^{CHG}$)

Bins	(1) Combined irregularity and control samples used in Panel A of Table 5				(2) Combined OP-IRREG and OP-CTRL subsamples used in Panel A of Table 7			
	Control sample (*MISSTMT* = 0)		Irregularity sample (*MISSTMT* = 1)		OP-CTRL subsample (*OP-MISSTMT* = 0)		OP-IRREG subsample (*OP-MISSTMT* = 1)	
	Obs.	Percent	Obs.	Percent	Obs.	Percent	Obs.	Percent
(< −10%)	7548	7.59%	54	5.60%				
[−10%, −9%)	665	0.67%	6	0.62%				
[−9%, −8%)	793	0.80%	7	0.73%				
[−8%, −7%)	900	0.91%	7	0.73%				
[−7%, −8%)	1164	1.17%	9	0.93%				
[−6%, −5%)	1401	1.41%	9	0.93%				
[−5%, −4%)	2005	2.02%	16	1.66%				
[−4%, −3%)	2738	2.75%	20	2.07%				
[−3%, −2%)	4299	4.32%	37	3.84%				
[−2%, −1%)	7262	7.30%	73	7.57%				
[−1%, 0)	16,090	16.18%	183	18.98%				
[0, 1%]	**25,353**	**25.50%**	**303**	**31.43%**	**25,353**	**46.46%**	**164**	**62.12%**
(1%, 2%]	7850	7.90%	84	8.71%	7850	14.39%	44	16.67%
(2%, 3%]	3950	3.97%	33	3.42%	3950	7.24%	10	3.79%
(3%, 4%]	2524	2.54%	26	2.70%	2524	4.63%	12	4.55%
(4%, 5%]	1734	1.74%	11	1.14%	1734	3.18%	5	1.89%
(5%, 6%]	1402	1.41%	7	0.73%	1402	2.57%	2	0.76%
(6%, 7%]	1031	1.04%	10	1.04%	1031	1.89%	6	2.27%
(7%, 8%]	887	0.89%	9	0.93%	887	1.63%	3	1.14%
(8%, 9%]	710	0.71%	4	0.41%	710	1.30%	3	1.14%
(9%, 10%]	611	0.61%	3	0.31%	611	1.12%	0	0.00%
(10%, 11%]	544	0.55%	6	0.62%	544	1.00%	5	1.89%
(> 11%)	7968	8.01%	47	4.88%	7968	14.60%	10	3.79%
Total	99,429	100%	964	100%	54,564	100%	264	100%

Panel B: the earnings level benchmark ($BMK = BMK^{LVL}$)

Bins	(1) Combined irregularity and control samples used in Panel B of Table 5				(2) Combined OP-IRREG and OP-CTRL subsamples used in Panel B of Table 7			
	Control sample (*MISSTMT* = 0)		Irregularity sample (*MISSTMT* = 1)		OP-CTRL subsample (*OP-MISSTMT* = 0)		OP-IRREG subsample (*OP-MISSTMT* = 1)	
	Obs.	Percent	Obs.	Percent	Obs.	Percent	Obs.	Percent
(< −10%)	12,418	12.42%	69	7.14%				
[−10%, −9%)	1103	1.10%	4	0.41%				
[−9%, −8%)	1205	1.21%	11	1.14%				
[−8%, −7%)	1402	1.40%	8	0.83%				
[−7%, −8%)	1757	1.76%	7	0.72%				
[−6%, −5%)	2105	2.11%	18	1.86%				
[−5%, −4%)	2572	2.57%	12	1.24%				
[−4%, −3%)	3383	3.38%	19	1.97%				
[−3%, −2%)	4379	4.38%	37	3.83%				
[−2%, −1%)	5764	5.77%	59	6.11%				

(*continued on next page*)

Table 6 (*continued*)

Panel B: the earnings level benchmark ($BMK = BMK^{LVL}$)

Bins	(1) Combined irregularity and control samples used in Panel B of Table 5				(2) Combined OP-IRREG and OP-CTRL subsamples used in Panel B of Table 7			
	Control sample (*MISSTMT* = 0)		Irregularity sample (*MISSTMT* = 1)		OP-CTRL subsample (*OP-MISSTMT* = 0)		OP-IRREG subsample (*OP-MISSTMT* = 1)	
	Obs.	Percent	Obs.	Percent	Obs.	Percent	Obs.	Percent
[−1%, 0)	7316	7.32%	101	10.46%				
[0, 1%]	**17,049**	**17.06%**	**247**	**25.57%**	**17,049**	**30.15%**	**131**	**49.43%**
(1%, 2%]	18,309	18.32%	201	20.81%	18,309	32.38%	80	30.19%
(2%, 3%]	8728	8.73%	84	8.70%	8728	15.43%	23	8.68%
(3%, 4%]	4117	4.12%	41	4.24%	4117	7.28%	12	4.53%
(4%, 5%]	2182	2.18%	22	2.28%	2182	3.86%	9	3.40%
(5%, 6%]	1324	1.32%	3	0.31%	1324	2.34%	1	0.38%
(6%, 7%]	876	0.88%	7	0.72%	876	1.55%	4	1.51%
(7%, 8%]	589	0.59%	3	0.31%	589	1.04%	0	0.00%
(8%, 9%]	407	0.41%	1	0.10%	407	0.72%	1	0.38%
(9%, 10%]	348	0.35%	1	0.10%	348	0.62%	0	0.00%
(10%, 11%]	285	0.29%	1	0.10%	285	0.50%	1	0.38%
(> 11%)	2333	2.33%	10	1.04%	2333	4.13%	3	1.13%
Total	99,951	100%	966	100%	56,547	100%	265	100%

Panel C: the earnings forecast benchmark ($BMK = BMK^{FCST}$)

Bins	(1) Combined irregularity and control samples used in Panel C of Table 5				(2) Combined OP-IRREG and OP-CTRL subsamples used in Panel C of Table 7			
	Control sample (*MISSTMT* = 0)		Irregularity sample (*MISSTMT* = 1)		OP-CTRL subsample (*OP-MISSTMT* = 0)		OP-IRREG subsample (*OP-MISSTMT* = 1)	
	Obs.	Percent	Obs.	Percent	Obs.	Percent	Obs.	Percent
Miss > 10¢	3695	7.76%	65	11.05%				
Miss 10¢	359	0.75%	7	1.19%				
Miss 9¢	403	0.85%	5	0.85%				
Miss 8¢	500	1.05%	3	0.51%				
Miss 7¢	559	1.17%	5	0.85%				
Miss 6¢	679	1.43%	8	1.36%				
Miss 5¢	874	1.84%	9	1.53%				
Miss 4¢	1068	2.24%	10	1.70%				
Miss 3¢	1377	2.89%	13	2.21%				
Miss 2¢	1886	3.96%	28	4.76%				
Miss 1¢	2999	6.30%	47	7.99%				
Just Meet	**7360**	**15.46%**	**113**	**19.22%**	**7360**	**22.16%**	**91**	**35.83%**
Beat 1¢	**6521**	**13.69%**	**84**	**14.28%**	**6521**	**19.63%**	**62**	**24.41%**
Beat 2¢	4678	9.82%	61	10.37%	4678	14.08%	40	15.75%
Beat 3¢	3315	6.96%	37	6.29%	3315	9.98%	20	7.87%
Beat 4¢	2356	4.95%	18	3.06%	2356	7.09%	9	3.54%
Beat 5¢	1727	3.63%	16	2.72%	1727	5.20%	6	2.36%
Beat 6¢	1288	2.70%	10	1.70%	1288	3.88%	8	3.15%
Beat 7¢	960	2.02%	6	1.02%	960	2.89%	2	0.79%
Beat 8¢	784	1.65%	5	0.85%	784	2.36%	0	0.00%
Beat 9¢	582	1.22%	3	0.51%	582	1.75%	2	0.79%

(*continued on next page*)

Table 6 (*continued*)

Panel C: the earnings forecast benchmark ($BMK = BMK^{FCST}$)

Bins	(1) Combined irregularity and control samples used in Panel C of Table 5				(2) Combined OP-IRREG and OP-CTRL subsamples used in Panel C of Table 7			
	Control sample (*MISSTMT* = 0)		Irregularity sample (*MISSTMT* = 1)		OP-CTRL subsample (*OP-MISSTMT* = 0)		OP-IRREG subsample (*OP-MISSTMT* = 1)	
	Obs.	Percent	Obs.	Percent	Obs.	Percent	Obs.	Percent
Beat 10¢	486	1.02%	7	1.19%	486	1.46%	5	1.97%
Beat > 10¢	3162	6.64%	28	4.76%	3162	9.52%	9	3.54%
Total	47,618	100%	588	100%	33,219	100%	254	100%

Notes

See Appendix A for definitions of all variables.

We conduct several additional analyses to Table 5. First, we include all three benchmark-beating measures (BMK^{CHG}, BMK^{LVL}, and BMK^{FCST}) in Eq. (7) and repeat the analyses in Table 5. In untabulated results, we find that BMK^{CHG} is no longer significant in the presence of BMK^{LVL} and BMK^{FCST} where both BMK^{LVL} and BMK^{FCST} are significant at the 0.01 or better level. In addition, these three benchmarks together are useful for detecting accounting irregularities incremental to discretionary accruals, F-scores, and discretionary and F-scores combined. These three benchmarks together dominate discretionary accruals in a one-on-once horse race but are still dominated by F-scores. Second, we define a new benchmark-beating variable, BMK^{ANY}, and re-estimate Eq. (7), where $BMK^{ANY} = 1$ if $BMK^{CHG} = 1$, $BMK^{LVL} = 1$, or $BMK^{FCST} = 1$, and 0 otherwise. We find that BMK^{ANY} is significant in all regressions. In addition, BMK^{ANY} dominates discretionary accruals in a one-on-one horse race but is dominated by F-scores.

Third, a concern for our findings in Table 5 is that the firms in the irregularity sample are fundamentally different from firms in the control sample. The differences in firm characteristics between the irregularity sample and control sample, rather than benchmark-beating, may explain why firms in the irregularity sample commit intentional misreporting. To alleviate such a concern, we construct an industry, year-quarter, size, and ROA matched control sample for our irregularity sample. Specifically, for each firm-quarter observation in the irregularity sample, we choose firms in the same industry (according to two-digit SIC code), same year-quarter, and same size (total assets) quintile from the control sample as matching candidates. Among these matching candidates, we choose one observation that has the closest ROA with the treatment observation. Using the above procedure, we obtain a one-to-one matched irregularity and control sample for each of the three benchmarks (the earnings change benchmark, the earnings level benchmark, and earnings forecast benchmark) in Table 5. We re-estimate Model 1 and Model 5 of Table 5 for each benchmark using the matched irregularity and control sample. We find that the coefficient on each benchmark-beating measure (BMK^{CHG}, BMK^{LVL}, or BMK^{FCST}) remains significantly positive in both Model 1 and Model 5 (untabulated). That is, results in Model 1 and Model 5 in Table 5 are robust to industry, year-quarter, size, and ROA matching.

4.4. Benchmark-beating and opportunistic accounting irregularities

To address our third research question (RQ3) of whether benchmark-beating is more useful for detecting *opportunistic* accounting irregularities than accounting irregularities in general, we partition our irregularity sample into an opportunistic irregularity (OP-IRREG) subsample and a non-opportunistic irregularity subsample following Badertscher et al. (2012). Specifically, we classify an observation from the irregularity sample into the OP-IRREG subsample if one of two conditions is met: (1) first-reported (i.e., manipulated) earnings are greater than or equal to an earnings benchmark but restated (i.e., non-manipulated) earnings are less than the benchmark and (2) restated (i.e., non-manipulated) earnings are greater than first-reported (i.e., manipulated) earnings and first-reported earnings are greater than or equal to the earnings benchmark.

Observations meeting Condition (1) are classified as opportunistic because these observations were below an earnings benchmark without manipulation. Managers manipulated earnings *up* so that manipulated earnings (i.e., first-reported earnings) meet or beat their benchmark. So, the purpose of manipulation for these observations appears opportunistic: to meet or beat this benchmark. Observations meeting Condition (2) are classified as opportunistic because these observations without manipulation were substantially above an earnings benchmark. Managers manipulated earnings *down* but keeping manipulated earnings (i.e., first-reported earnings) above the benchmark. So, the purpose of manipulation for these observations appears opportunistic: to create "cookie jar" reserves for boosting future earnings. From the above procedure, we obtain the OP-IRREG subsample from the irregularity sample. We exclude the remaining, non-opportunistic irregularity subsample.

We estimate the following logistic model using the combined OP-IRREG subsample and the opportunistic control subsample (discussed below) to investigate the association between benchmark-beating and opportunistic accounting irregularities:

$$OP\text{–}MISSTMT = \delta_0 + \delta_1 BMK + \delta_2 DA + \delta_3 FSCORE + \delta_4 LEV + \delta_5 ROA + \delta_6 AVGAT + \delta_7 BigN + \varepsilon, \quad (8)$$

where *OP-MISSTMT* is a dummy variable that equals 1 for observations from the OP-IRREG subsample and 0 for observations from the opportunistic control subsample.

When the dependent variable (*OP-MISSTMT*) equals one (i.e., *OP-MISSTMT* = 1), it identifies the OP-IRREG subsample where first-reported earnings meet or *beat* (both slightly and greatly) a benchmark. When the independent variable (*BMK*) equals one (i.e., *BMK* = 1), it identifies benchmark beaters whose earnings meet or *slightly* beat the benchmark. Thus, benchmark beaters (*BMK* = 1) are a subset of the OP-IRREG subsample (*OP-MISSTMT* = 1).[17] This suggests a mechanical relation between the dependent (*OP-MISSTMT*) and independent variable (*BMK*). To address this issue, we construct the opportunistic control (OP-CTRL) subsample from the control sample in the same way as we construct the OP-IRREG subsample from the irregularity sample. That is, the OP-CTRL subsample includes only observations with reported earnings equal to or *above* (both slightly and greatly) a benchmark. Thus, benchmark beaters (*BMK* = 1) is also a subset of the OP-CTRL subsample (*OP-MISSTMT* = 0).[18] Eq. (8) examines whether benchmark-beating firms (*BMK* = 1) are more likely to engage in opportunistic earnings manipulation (*OP-MISSTMT* = 1) than non-benchmark-beating firms (*BMK* = 0), given the fact that earnings in both the OP-IRREG subsample and OP-CTRL subsample are equal to or above a benchmark.

Table 6 shows how the OP-IRREG subsample is constructed from the irregularity sample and how the OP-CTRL subsample is constructed from the control sample. We only discuss the construction procedure for the earnings forecast benchmark in Panel C since the procedure is similar for the other two benchmarks (Panel A and Panel B). First, the column (1) shows the distribution of 588 observations in the irregularity sample (*MISSTMT* = 1) and 47,618 observations in the control sample (*MISSTMT* = 0). The combined sample of 48,206 observations (= 588 + 47,618) is used in Panel C of Table 5. There are 113 (84) observations that just meet analyst earnings forecasts (beat the forecasts by 1¢) in the irregularity sample (*MISSTMT* = 1) whereas there are 7360 (6521) observations that just meet analyst earnings forecasts (beat the forecasts by 1¢) in the control sample (*MISSTMT* = 0). Among 14,078 (= 113 + 84 + 7360 + 6521) benchmark beaters ($BMK^{FCST} = 1$), 1.40% ($= \frac{113+84}{14,078}$) intentionally misreport earnings. In contrast, among 34,128 (= 47,618 + 588–14,078) non-benchmark beaters ($BMK^{FCST} = 0$), 1.15% ($= \frac{588-113-84}{34,128}$) intentionally misreport earnings. Thus, benchmark-beating firms is 21.74% ($= \frac{1.40\% - 1.15\%}{1.15\%}$) more likely to misreport earnings than non-benchmark-beating firms (see also Panel B of Table 3).

Second, we construct the OP-IRREG subsample from the irregularity sample following Badertscher et al. (2012) and the OP-CTRL subsample from the control sample similarly (i.e., requiring that reported earnings equal to or *above* the benchmark). As shown in the column (2) of Panel C, Table 6, 153 (= 91 + 62) benchmark beaters are in the OP-IRREG subsample and 13,881 (= 7360 + 6521) benchmark beaters are in the OP-CTRL subsample. Among 14,034 (= 153 + 13,881) benchmark beaters, 1.09% misreport earnings for opportunistic reasons. In contrast, among 19,439 (= 254 + 33,219–14,034) non-benchmark beaters, 0.52% misreport earnings for opportunistic reasons. Benchmark-beating firms are 109.62% ($= \frac{1.09\% - 0.52\%}{0.52\%}$) more likely to misreporting earnings for opportunistic reasons than non-benchmark-beating firms. This percentage (109.62%) is larger than its counterpart (21.74%) in the column (1). This implies that benchmark-beating is more positively associated with opportunistic accounting irregularities than accounting irregularities in general.[19]

Badertscher et al. (2012, p. 348) argue that if benchmark-beating is a reliable indicator of opportunistic earnings management, then the majority of the opportunistic earnings management subsample should fall in the "Just Meet" and "Beat 1¢" bins. Indeed, 60.24% (= 35.83% + 24.41%) of our OP-IRREG subsample falls in the "Just Meet" and "Beat 1¢" bins. In contrast, only 41.79% (= 22.16% + 19.63%) of the OP-CTRL subsample falls in these two bins. Our evidence thus suggests that benchmark-beating is a good indicator of opportunistic accounting irregularities.

We formally investigate our third research question (RQ3) of whether benchmark-beating is more useful for detecting *opportunistic* accounting irregularities than accounting irregularities in general in two steps. First, we examine whether benchmark-beating is more positively associated with opportunistic accounting irregularities than accounting irregularities in general. We estimate Eq. (8) using the combined OP-IRREG subsample and OP-CTRL subsample. Table 7 reports our findings. Panel A presents the findings for the earnings change benchmark (BMK^{CHG}). The coefficient on BMK^{CHG} in Model 1 is 0.77 (*p*-value = 0.00), which suggests that benchmark-beating firms are 115.98% (= $e^{0.77}$–1) more likely to engage in *opportunistic* earnings manipulation than non-benchmark-beating firms. The coefficient is more than twice as large as its counterpart in Table 5 (0.77 vs 0.35), indicating that meeting or slightly beating the earnings change benchmark is more likely to detect *opportunistic* accounting irregularity than accounting irregularity *in general*. The χ^2 test for the difference in the coefficients on BMK^{CHG} in Model 1 between Table 7 and Table 5 (0.77 vs 0.35) is strongly significant ($\chi^2 = 12.93$, *p*-value = 0.0003), rejecting the null that these two coefficients are equal (untabulated). Model 1 in Panels B and C, Table 7, shows that the coefficient on BMK^{LVL} and that on BMK^{FCST} are 0.27 and 0.77, respectively. Their counterparts in Table 5 are 0.57 and 0.27. The untabulated χ^2 tests suggest that the coefficient on BMK^{LVL} in Model 1 between Table 7 and Table 5 (0.27 vs. 0.57) is insignificant ($\chi^2 = 2.49$, *p*-value = 0.1142) and that the coefficient on BMK^{FCST} in Model 1 between Table 7 and Table 5 (0.77 vs. 0.27) is strongly significant ($\chi^2 = 27.48$, *p*-value = 0.0000). In summary, BMK^{CHG} and BMK^{FCST} are more positively associated with opportunistic accounting irregularities than accounting irregularities in general whereas BMK^{LVL} is as strongly associated with opportunistic accounting irregularities as accounting irregularities in general.

Second, we conduct a one-on-one horse race between benchmark-beating and discretionary accruals and between benchmark-beating and F-scores for detecting opportunistic accounting irregularities. As shown in Table 7, the Vuong test shows that benchmark-beating outperforms (ties with) discretionary accruals in two (one) races (race), and ties with (is outperformed by) F-scores in two (one) races (race).

[17] Let's take the OP-IRREG subsample (*OP-MISSTMT* = 1) column in Table 6, Panel A, for an example. The bold numbers (164 and 62.12%) represent benchmark beaters ($BMK^{CHG} = 1$), which are a subset of 264 observations in the OP-IRREG subsample.

[18] Let's take the OP-CTRL subsample (*OP-MISSTMT* = 0) column in Table 6, Panel A, for an example. The bold numbers (25,353 and 46.46%) represent benchmark beaters ($BMK^{CHG} = 1$), which are a subset of 54,564 observations in the OP-CTRL subsample.

[19] Among 588 firm-quarter observations in our irregularity sample (see Panel C of Table 6), 254 observations (43.20%) are classified as opportunistic accounting irregularities. Badertscher et al. (2012) classify 214 firm-year observations (68.59%) in their restatement sample of 312 observations as opportunistic earnings manipulation (their earnings benchmark is also analyst forecasts). Our opportunistic percentage is lower than theirs. This is most likely due to their sample being annual and ours being quarterly. If a firm manipulates its annual earnings opportunistically, it does not need to manipulate all four quarterly earnings opportunistically. The firm may only need to opportunistically manipulate quarterly earnings once or a couple of times in a year for that firm to be classified as an opportunistic manipulator on an annual basis.

Table 7
Two-way Clustering Logistic Regressions of OP-MISSTMT on Measures of Benchmark-Beating.

Panel A: the earning change benchmark (*BMK* = BMK^{CHG})

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−5.39 (0.00)	−5.40 (0.00)	−6.41 (0.00)	−6.11 (0.00)	−6.41 (0.00)
BMK^{CHG}	0.77 (0.00)	0.78 (0.00)	0.70 (0.00)		0.71 (0.00)
DA		0.64 (0.00)		0.54 (0.00)	0.63 (0.00)
FSCORE			0.96 (0.00)	0.99 (0.00)	0.95 (0.00)
LEV	0.36 (0.11)	0.36 (0.11)	0.14 (0.63)	−0.04 (0.89)	0.15 (0.61)
ROA	−0.54 (0.30)	−0.48 (0.28)	−0.95 (0.05)	−0.61 (0.19)	−0.82 (0.07)
AVGAT/1000	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
BigN	−0.69 (0.00)	−0.68 (0.00)	−0.66 (0.00)	−0.52 (0.00)	−0.66 (0.00)
No. of OP-IRREG obs.	264	264	264	264	264
No. of OP-CTRL obs.	54,564	54,564	54,564	54,564	54,564
Pseudo R^2	0.0191	0.0203	0.0370	0.0293	0.0380

1. R^2 Decomposition	
Incremental R^2 by BMK^{CHG}:	$R^2_{Model\ 5} - R^2_{Model\ 4} = 0.0380–0.0293 = 0.0087$
Incremental R^2 by *DA*:	$R^2_{Model\ 5} - R^2_{Model\ 3} = 0.0380–0.0370 = 0.0010$
Incremental R^2 by *FSCORE*:	$R^2_{Model\ 5} - R^2_{Model\ 2} = 0.0380–0.0203 = 0.0177$
2. Vuong (1989) Test	
Model 3 (BMK^{CHG}) vs. Model 4 (*DA*):	Vuong's z-statistic = 2.37 (0.02)
Model 2 (BMK^{CHG}) vs. Model 4 (*FSCORE*):	Vuong's z-statistic = −1.58 (0.11)

Panel B: the earnings level benchmark (*BMK* = BMK^{LVL})

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−4.17 (0.00)	−4.17 (0.00)	−5.00 (0.00)	−4.74 (0.00)	−4.98 (0.00)
BMK^{LVL}	0.27 (0.05)	0.27 (0.05)	0.30 (0.03)		0.30 (0.03)
DA		0.98 (0.00)		0.90 (0.02)	0.90 (0.02)
FSCORE			0.72 (0.00)	0.69 (0.00)	0.71 (0.00)
LEV	0.36 (0.22)	0.35 (0.25)	0.24 (0.44)	0.09 (0.76)	0.23 (0.46)
ROA	−0.60 (0.00)	−0.59 (0.00)	−0.59 (0.00)	−0.63 (0.00)	−0.59 (0.00)
AVGAT/1000	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
BigN	−0.93 (0.00)	−0.93 (0.00)	−0.94 (0.00)	−0.94 (0.00)	−0.93 (0.00)
No. of OP-IRREG obs.	265	265	265	265	265
No. of OP-CTRL obs.	56,547	56,547	56,547	56,547	56,547
Pseudo R^2	0.1053	0.1065	0.1147	0.1140	0.1154

1. R^2 Decomposition	
Incremental R^2 by BMK^{LVL}:	$R^2_{Model\ 5} - R^2_{Model\ 4} = 0.1154–0.1140 = 0.0014$
Incremental R^2 by *DA*:	$R^2_{Model\ 5} - R^2_{Model\ 3} = 0.1154–0.1147 = 0.0007$
Incremental R^2 by *FSCORE*:	$R^2_{Model\ 5} - R^2_{Model\ 2} = 0.1154–0.1065 = 0.0089$
2. Vuong (1989) Test	
Model 3 (BMK^{LVL}) vs. Model 4 (*DA*):	Vuong's z-statistic = 0.34 (0.47)
Model 2 (BMK^{LVL}) vs. Model 4 (*FSCORE*):	Vuong's z-statistic = −2.12 (0.03)

Panel C: the earnings forecast benchmark ($BMK = BMK^{FCST}$)

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	−4.49 (0.00)	−4.50 (0.00)	−5.67 (0.00)	−5.35 (0.00)	−5.66 (0.00)
BMK^{FCST}	0.77 (0.00)	0.77 (0.00)	0.69 (0.00)		0.69 (0.00)
DA		1.12 (0.05)		0.80 (0.10)	0.89 (0.07)
FSCORE			1.06 (0.00)	1.10 (0.00)	1.05 (0.00)
LEV	1.06 (0.00)	1.05 (0.00)	0.84 (0.01)	0.79 (0.03)	0.84 (0.01)
ROA	−1.10 (0.06)	−0.67 (0.44)	−1.73 (0.00)	−1.62 (0.00)	−1.50 (0.01)
AVGAT/1000	0.01 (0.03)	0.01 (0.03)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
BigN	−1.31 (0.00)	−1.30 (0.00)	−1.28 (0.00)	−1.27 (0.00)	−1.27 (0.00)
No. of OP-IRREG obs.	254	254	254	254	254
No. of OP-CTRL obs.	33,473	33,473	33,473	33,473	33,473
Pseudo R^2	0.0442	0.0451	0.0648	0.0554	0.0653

1. R^2 Decomposition
   - Incremental R^2 by BMK^{FCST}: $R^2_{Model\ 5} - R^2_{Model\ 4}$ = 0.0653–0.0554 = 0.0099
   - Incremental R^2 by *DA*: $R^2_{Model\ 5} - R^2_{Model\ 3}$ = 0.0653–0.0648 = 0.0005
   - Incremental R^2 by *FSCORE*: $R^2_{Model\ 5} - R^2_{Model\ 2}$ = 0.0653–0.0451 = 0.0202

2. Vuong (1989) Test
   - Model 3 (BMK^{FCST}) vs. Model 4 (*DA*): Vuong's z-statistic = 2.56 (0.01)
   - Model 2 (BMK^{FCST}) vs. Model 4 (*FSCORE*): Vuong's z-statistic = −1.52 (0.13)

Notes

See Appendix A for definitions of all variables. Eq. (8) is estimated using the combined OP-IRREG subsample and OP-CTRL subsample. $BMK = BMK^{CHG}$, BMK^{LVL}, and BMK^{FCST} in Panel A, Panel B, and Panel C, respectively.

Numbers in parentheses beside coefficient estimates are *p*-values of a two-tailed test based on the two-way clustered standard errors (Petersen, 2009). Numbers in parentheses under Vuong (1989) Test are *p*-values of a two-tailed test for Vuong's z-statistic.

These performances are better than their counterparts in Table 5. Overall, the answer to our third research question (RQ3) is that benchmark-beating is more useful for detecting *opportunistic* accounting irregularities (i.e., Table 7) than accounting irregularities in general (i.e., Table 5).

Similar to the sensitivity test to Table 5, we construct a matched irregularity and control sample for each of the three benchmarks (the earnings change benchmark, the earnings level benchmark, and earnings forecast benchmark) in Table 7. We re-estimate Model 1 and Model 5 of Table 7 for each benchmark using the matched irregularity and control sample. We find that the coefficient on each benchmark-beating measure (BMK^{CHG}, BMK^{LVL}, or BMK^{FCST}) remains significantly positive in both Model 1 and Model 5 (untabulated). That is, results in Model 1 and Model 5 in Table 7 are robust to industry, year-quarter, size, and ROA matching.

5. Conclusion

Prior literature commonly uses firms that meet or slightly beat an earnings benchmark (e.g., earnings of the same quarter last year, zero earnings, or consensus analyst earnings forecasts) as a proxy for earnings management. However, there is only limited evidence in the extant literature linking benchmark-beating to unequivocal earnings management. As Dechow et al. (2010) point out, "[t]the totality of the evidence indicates that the use of small profits as a proxy for earnings management more generally is unsubstantiated." We seek to provide evidence on a link between benchmark-beating and earnings management in this paper. In addition, we investigate whether benchmark-beating is useful for detecting accounting irregularities both incremental and relative to discretionary accruals and F-scores, and whether it is more useful for detecting opportunistic accounting irregularities than accounting irregularities in general.

We identify a sample of irregularity firms that restate their earnings due to intentional misreporting and construct a control sample where earnings are not restated. First, we compare the benchmark-beating sample with the non-benchmark-beating sample. In univariate analyses, we find that benchmark beaters are 21.74% - 66.28% more likely to intentionally misreport earnings than non-benchmark beaters. Our multivariate regression results are consistent with the univariate evidence and suggest that benchmark beaters are 31.00% - 76.83% more likely to intentionally misreport earnings than non-benchmark beaters after controlling for other determinants of misreporting.

Second, we compare benchmark-beating with discretionary accruals, arguably the most widely used proxy for earnings management, and with F-scores, arguably the most powerful detector of accounting misstatements. We find that benchmark-beating is useful for detecting accounting irregularities incremental to (i) discretionary accruals, (ii) F-scores, and (iii) discretionary accruals and F-scores combined. These findings suggest that benchmark-beating, discretionary accruals, and F-scores capture different aspects of earnings management. In a one-on-one horse race, benchmark-beating ties with (outperforms) discretionary accruals in two (one) races (race), although it is outperformed by F-scores in all three races.

Finally, we examine whether benchmark-beating is more useful for detecting *opportunistic* accounting irregularities. We find that benchmark-beating is more positively associated with opportunistic accounting irregularities than accounting irregularities in general. Moreover, benchmark-beating ties with (outperforms) discretionary accruals in one (two) races and it ties with (is outperformed by) F-scores in two (one) in detecting opportunistic accounting irregularities. These findings suggest that benchmark-beating is especially useful for detecting opportunistic accounting irregularities (a more harmful form of earnings manipulation) than accounting irregularities in general. Collectively, our findings validate the use of benchmark-beating as a proxy for earnings management.

Data availability

Data used in this study are available from the sources identified in the study.

Appendix A. Variable definitions

MISSTMT	=	a dummy variable for accounting irregularities or intentional misreporting = 1 for observations from the irregularity sample and 0 for observations from the control sample.
EARN	=	earnings in million dollars = first-reported net income (Compustat mnemonic: NIQR) from the COMPUSTAT Unrestated Quarterly database for the irregularity sample, and = bottom-line net income (NIQ) from the COMPUSTAT Fundamental Quarterly database for the control sample.
CHG	=	scaled earnings changes = $(EARN_t - EARN_{t-4})/MKTCAP_{t-1}$, where $MKTCAP$ = PRCCQ (stock price) × CSHOQ (common shares outstanding) at the beginning of quarter t.
BMK^{CHG}	=	a dummy variable for meeting or slightly beating the earnings change benchmark (i.e., earnings of the same quarter last year) = 1 if $0 \le CHG < 1\%$ and 0 otherwise.
LVL	=	scaled earnings levels = $EARN_t/MKTCAP_{t-1}$.
BMK^{LVL}	=	a dummy variable for meeting or slightly beating the earnings level benchmark (i.e., zero earnings) = 1 if $0 \le LVL < 1\%$ and 0 otherwise.
EPS_ACT	=	the I/B/E/S-reported actual earnings per share from the I/B/E/S Summary database unadjusted for stock splits and dividends.
EPS_MED	=	the median analyst earnings forecast in the month immediately before an earnings announcement from the I/B/E/S Summary database unadjusted for stock splits and dividends.
SURP	=	earnings surprises per share = $EPS\_ACT - EPS\_MED$.
BMK^{FCST}	=	a dummy variable for meeting or slightly beating the earnings forecast benchmark (i.e., consensus analyst earnings forecasts) = 1 if $SURP$ = 0¢ or 1¢ and 0 otherwise.
BMK	=	a dummy variable for meeting or slightly beating an earnings benchmark = BMK^{CHG}, BMK^{LVL}, or BMK^{FCST}.
AVGAT	=	average total assets = the average of total assets (ATQ) at the beginning and end of a quarter.

DA = discretionary accruals. For the control sample, *DA* = performance adjusted discretionary accruals = the residual from the following cross-sectional Kothari et al. (2005) model that controls for performance in the modified Jones (1991) model: $TACC = \alpha_0 + \alpha_1 Q2 + \alpha_2 Q3 + \alpha_3 Q4 + \alpha_4(\Delta REV - \Delta AR) + \alpha_5 PPE + \alpha_6 ROA + \varepsilon$, where *TACC* is total accruals calculated following Hribar and Collins (2002) as income before extraordinary items (IBCQ) minus operation cash flows (OANCFQ) plus extraordinary items and discontinued operations (XIDOCQ). $Q2 - Q4$ are dummy variables, equal to one if the fiscal quarter is the second, third, and fourth quarter, respectively, and zero otherwise. ΔREV and ΔAR are the changes in sales (SALEQ) and in accounts receivable (RECTQ), respectively, from the previous quarter to the current quarter. *PPE* is the gross property, plant, and equipment (PPEGTQ). PPEGTQ is missing for a large proportion of observations from the COMPUSTAT Fundamental Quarterly database when the fiscal quarter is not the fourth quarter. In such a case, we assume PPEGTQ in a non-fourth quarter to be equal to its fourth quarter value. *ROA* is income before extraordinary items (IBC) scaled by average total assets. All variables in the model except the intercept are scaled by average total assets (*AVGAT*). The model is estimated in the cross-section for each two-digit SIC code and year combination with at least 10 observations using the control sample. For the irregularity sample, $DA = TACC - [\hat{\alpha}_0 + \hat{\alpha}_1 Q2 + \hat{\alpha}_2 Q3 + \hat{\alpha}_3 Q4 + \hat{\alpha}_4(\Delta REV - \Delta AR) + \hat{\alpha}_5 PPE + \hat{\alpha}_6 ROA]$, where $\hat{\alpha}_0 - \hat{\alpha}_6$ are the coefficient estimates from estimating the McNichols (2002) model in the cross-section using the control sample as described above and all variables are defined as before but measured using the first-reported values from the COMPUSTAT Unrestated Quarterly database.

FSCORE = F-scores. Larger values of F-scores indicate higher probabilities of earnings misstatements. For the control sample, $FSCORE = \frac{Probability}{Unconditional\ Expectation\ of\ Misstatement}$, where *Unconditional Expectation of Misstatement* = the number of observations in the irregularity sample divided by the total number of observations in the irregularity sample and control sample; $Probability = \frac{e^{Predicted\ Value}}{(1 + e^{Predicted\ Value})}$; $Predicted\ Value = \hat{\beta}_0 + \hat{\beta}_1 RSST_{ACC} + \hat{\beta}_2 CH_{REC} + \hat{\beta}_3 CH_{INV} + \hat{\beta}_4 SOFT_{ASSET} + \hat{\beta}_5 CH_{CS} + \hat{\beta}_6 CH_{ROA} + \hat{\beta}_7 ISSUE$, where $\hat{\beta}_0 - \hat{\beta}_7$ are coefficient estimates from estimating the following model using our combined irregularity and control samples: $MISSTMT = \beta_0 + \beta_1 RSST_{ACC} + \beta_2 CH_{REC} + \beta_3 CH_{INV} + \beta_4 SOFT_{ASSET} + \beta_5 CH_{CS} + \beta_6 CH_{ROA} + \beta_7 ISSUE + \varepsilon$ (our coefficient estimates are $\hat{\beta}_0 = 6.333$, $\hat{\beta}_1 = 0.165$, $\hat{\beta}_2 = 1.246$, $\hat{\beta}_3 = 1.989$, $\hat{\beta}_4 = 1.672$, $\hat{\beta}_5 = 0.434$, $\hat{\beta}_6 = 0.465$, and $\hat{\beta}_7 = 0.777$); *MISSTMT* is defined earlier; $RSST_{ACC}$ = a broad measure of accruals based on Richardson, Sloan, Soliman, and Tuna (2005) = (ΔWC (change in non-cash working capital) + ΔNCO (change in net non-current operating assets) + ΔFIN (change in net financial assets))/*AVGAT*; after collecting terms, $RSST_{ACC}$ = [ΔTotal Assets (ATQ) – ΔTotal Liabilities (LTQ) – ΔCash and Short-term Investments (CHEQ) + ΔShort-term Investments (IVSTQ) – ΔPreferred Stock (PSTKQ)]/*AVGAT*; CH_{REC} = change in receivables = ΔReceivables (RECTQ)/*AVGATQ*; CH_{INV} = change in inventory = ΔInventories (INVTQ)/*AVGATQ*; $SOFT_{ASSET}$ = percentage of soft assets = (ATQ – Net Property, Plant, and Equipment (PPENTQ) – CHEQ)/*AVGAT*; CH_{CS} = percentage change in cash sales in quarter $t = (CSALE_t - CSALE_{t-1})/CSALE_{t-1}$, where *CSALE* = cash sales = Sales (SALEQ) – ΔReceivables (RECTQ); CH_{ROA} = change in return on assets = (Earnings$_t$ (IBQ)/$AVGAT_t$ – Earnings$_{t-1}$/$AVGAT_{t-1}$; and *ISSUE* = a dummy variable for actual issuance = 1 if the firm issued equity (SSTKQ > 0) or debt (DLTISQ > 0) in quarter *t* and 0 otherwise. For the irregularity sample, *FSCORE* is calculated using the same procedure as above except all accounting variables for calculating *Predicted Value* are measured using the first-reported values.

LEV = leverage = (Long-term Debt (DLTTQ) + Debt in Current Liabilities (DLCQ))/*AVGAT*.

ROA = return on assets = the bottom-line net income (NIQ)/*AVGAT*.

BigN = a dummy variable = 1 if a firm's auditor is Arthur Andersen, Coopers & Lybrand, Ernst & Young, Deloitte & Touche, KPMG, or PriceWaterhouseCoopers and 0 otherwise.

$EARN^{RS}$ = restated earnings for the irregularity sample = the bottom-line net income (NIQ) from the Fundamental Quarterly database since COMPUSTAT Fundamental Quarterly database provide a firm's financial data with restated data *replacing* the originally-reported values.

EPS_ACT^{RS} = adjusted or as if restated I/B/E/S-reported actual EPS for the irregularity sample = *EPS_ACT* + *EPS_RS* (restatement amount per share), where *EPS_RS* = EPSPXQ – EPSPXQR when the I/B/E/S primary/diluted index (PDI) is equal to primary (P), EPSPXQ = *basic* earnings per share before extraordinary items and discontinued operations from the Fundamental Quarterly database, and EPSPXQR = first-reported *basic* earnings per share before extraordinary items and discontinued operations from the Unrestated Quarterly database; *EPS_RS* = EPSFXQ – EPSFXQR when the I/B/E/S primary/diluted index (PDI) is equal to diluted (D), EPSFXQ = *diluted* earnings per share before extraordinary items and discontinued operations, and EPSFXQR = first-reported *diluted* earnings per share before extraordinary items and discontinued operations. We adjust *EPS_ACT* (I/B/E/S-reported actual EPS) by (EPSPXQ – EPSPXQR) or (EPSFXQ – EPSFXQR), not by restated amount of net income (NIQ – NIQR), because what analysts forecast is not net income but an earnings measure similar to earnings before extraordinary items and discontinued operations (Collins, Li, & Xie, 2009; Doyle, Lundholm, & Soliman, 2003).

OP-MISSTMT = a dummy variable for opportunistic accounting irregularities = 1 for observations from the opportunistic irregularity (OP-IRREG) subsample and 0 for observations from the opportunistic control (OP-CTRL) subsample. OP-IRREG subsample is constructed as follows. We classify an observation from the irregularity sample into the OP-IRREG subsample if (1) first-reported earnings (*EARN* or *EPS_ACT*) are greater than or equal to an earnings benchmark (i.e., earnings of the same quarter last year, zero earnings, and consensus analyst earnings forecasts) but restated earnings ($EARN^{RS}$ or EPS_ACT^{RS}) are less than the benchmark or (2) restated earnings ($EARN^{RS}$ or EPS_ACT^{RS}) are greater than first-reported earnings (*EARN* or *EPS_ACT*) and first-reported earnings (*EARN* or *EPS_ACT*) are greater than or equal to the earnings benchmark. We construct the OP-CTRL subsample similarly by requiring reported earnings to equal or above an earnings benchmark.

References

American Institute of Certified Public Accountants (AICPA) (2002). *Consideration of Fraud in a Financial Statement Audit. Statement on Auditing Standards No. 99.* New York, NY: AICPA.

Ayers, B., Jiang, J. X., & Yeung, P. E. (2006). Discretionary accruals and earnings management: An analysis of pseudo targets. *The Accounting Review, 81*(3), 617–652.

Badertscher, B. A., Collins, D. W., & Lys, T. Z. (2012). Discretionary accounting choices and the predictive ability of accruals with respect to future cash flows. *Journal of Accounting and Economics, 53*(1–2), 330–352.

Beaver, W., McNichols, M., & Nelson, K. (2003). Management of the loss reserve accrual and the distribution of earnings in the property-casualty insurance industry. *Journal of Accounting and Economics, 35*(3), 347–376.

Beatty, A., Ke, B., & Petroni, K. (2002). Earnings management to avoid earnings declines across publicly and privately held banks. *The Accounting Review, 77*(3), 547–570.

Boone, J., Khurana, I., & Raman, K. (2012). Audit market concentration and auditor tolerance for earnings management. *Contemporary Accounting Research, 29*(4), 1171–1203.

Brown, L. D., & Caylor, M. L. (2005). A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences. *The Accounting Research, 80*(2), 423–440.

Burgstahler, D., & Chuk, E. (2015). Do scaling and selection explain earnings discontinuities? *Journal of Accounting and Economics, 60*(1), 168–186.

Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics, 24*(1), 99–126.

Cahan, S., Zhang, W., & Veenman, D. (2011). Did the Waste Management audit failures signal lower firm-wide audit quality at Arthur Andersen? *Contemporary Accounting Research, 28*(3), 859–891.

Caramanis, C., & Lennox, C. (2008). Audit effort and earnings management. *Journal of Accounting and Economics, 45*(1), 116–138.

Carcello, J., & Li, C. (2013). Costs and benefits of requiring an engagement partner signature: Recent experience in the United Kingdom. *The Accounting Review, 88*(5), 1511–1546.

Cheng, Q., & Warfield, T. (2005). Equity incentives and earnings management. *The Accounting Review, 80*(2), 441–476.

Collins, D., Li, O., & Xie, H. (2009). What drives the increased informativeness of earnings announcements over time? *Review of Accounting Studies, 14*(1), 1–30.

Davis, L., Soo, B., & Trompeter, G. (2009). Auditor tenure and the ability to meet or beat earnings forecasts. *Contemporary Accounting Research, 26*(2), 517–548.

Dechow, P. M., Ge, W., Larson, C., & Sloan, R. (2011). Predicting material accounting misstatements. *Contemporary of Accounting Research, 28*(1), 17–82.

Dechow, P. M., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics, 50*(2–3), 344–401.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *Journal of Business, 72*(1), 1–34.

Donelson, D., McInnis, J., & Mergenthaler, R. (2013). Discontinuities and earnings management: Evidence from restatements related to securities litigation. *Contemporary Accounting Research, 30*(1), 242–268.

Doyle, J., Lundholm, R., & Soliman, M. (2003). The predictive value of expenses excluded from pro forma earnings. *Review of Accounting Studies, 8*(2–3), 145–174.

Durtschi, C., & Easton, P. (2005). Earnings management? The shapes of the frequency distributions of earnings metrics are not evidence ipso facto. *Journal of Accounting Research, 43*(4), 557–592.

Durtschi, C., & Easton, P. (2009). Earnings management? Erroneous inferences based on earnings frequency distribution. *Journal of Accounting Research, 47*(5), 1249–1281.

Fama, E., & French, K. (1997). Industry costs of equity. *Journal of Financial Economics, 43*(2), 153–193.

Francis, J., Schipper, K., & Vincent, L. (2002). Expended disclosures and the increased usefulness of earnings announcements. *The Accounting Review, 77*(3), 515–546.

Frankel, R., Johnson, M., & Nelson, K. (2002). The relation between auditors' fees for nonaudit services and earnings management. *The Accounting Review, 77*(Supplement), 71–105.

Gunny, S. (2010). The relation between earnings management using real activities manipulation and future performance: Evidence from meeting earnings benchmarks. *Contemporary Accounting Research, 27*(3), 855–888.

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics, 20*(2), 125–152.

Hennes, K. M., Leone, A. J., & Miller, B. P. (2008). The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review, 83*(6), 1487–1519.

Hillegeist, S. A., Keating, E. A., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies, 9*(1), 5–34.

Hollie, D., Livnat, J., & Segal, B. (2012). Earnings revisions in SEC filings from prior preliminary announcements. *Journal of Accounting, Auditing and Finance, 27*(1), 3–31.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: John Wiley & Sons.

Hribar, P., & Collins, D. (2002). Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research, 40*(1), 105–134.

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

Jones, K. L., Krishnan, G. V., & Melendrez, K. D. (2008). Do models of discretionary accruals detect actual cases of fraudulent and restated earnings? An empirical analysis. *Contemporary Accounting Research, 25*(2), 499–531.

Kanagaretnam, K., Lim, C., & Lobo, G. (2010). Auditor reputation and earnings management: International evidence from the banking industry. *Journal of Banking and Finance, 34*(10), 2318–2327.

Keung, E., Lin, Z., & Shin, M. (2010). Does the stock market see a zero or small positive earnings surprises as a red flag? *Journal of Accounting Research, 48*(1), 105–135.

Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics, 31*(1–3), 105–231.

Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics, 39*(1), 163–197.

Lee, T. A., Ingram, R. W., & Howard, T. P. (1999). The difference between earnings and operating cash flow as an indicator of financial reporting fraud. *Contemporary Accounting Research, 16*(4), 749–786.

Livnat, J., & Mendenhall, R. (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research, 44*(1), 177–205.

McInnis, J. M., & Collins, D. (2011). The effect of cash flow forecasts on accrual quality and benchmark beating. *Journal of Accounting and Economics, 51*(3), 219–239.

McNichols, M. F. (2002). Discussion of The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review, 77*(Supplement), 61–69.

Minutti-Meza, M. (2013). Does auditor industry specialization improve audit quality? *Journal of Accounting Research, 51*(4), 779–817.

Payne, J., & Thomas, W. (2003). The implications of using stock-split adjusted I/B/E/S data in empirical research. *The Accounting Review, 78*(4), 1049–1067.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22*(1), 435–480.

Phillips, J., Pincus, M., & Rego, S. (2003). Earnings management: New evidence based on deferred tax expense. *The Accounting Review, 78*(2), 491–521.

Reichelt, K., & Wang, D. (2010). National and office-specific measures of auditor industry expertise and effects on audit quality. *Journal of Accounting Research, 48*(3), 647–686.

Richardson, S., Sloan, R., Soliman, M., & Tuna, I. (2005). Accrual reliability, earnings persistence, and stock prices. *Journal of Accounting and Economics, 39*(3), 437–485.

Srinidhi, B., Gul, F., & Tsui, J. (2011). Female directors and earnings quality. *Contemporary Accounting Research, 28*(5), 1610–1644.

Subramanyam, K. R. (1996). The pricing of discretionary accruals. *Journal of Accounting and Economics, 22*(1–3), 249–281.

Vafeas, N. (2005). Audit committees, boards, and the quality of reported earnings. *Contemporary Accounting Research, 22*(4), 1093–1122.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307–333.

Watts, R., & Zimmerman, J. (1986). *Positive accounting theory.* New Jersey: Prentice-Hall.