Received: 21 December 2018 | Revised: 5 August 2020 | Accepted: 7 August 2020

DOI: 10.1111/jbfa.12494

JBFA

ARTICLE

Rounding up performance measures in German firms: Earnings cosmetics or earnings management on a larger scale?

Sebastian Lebert[1] | **Ulf Mohrmann**[2,4] | **Ulrike Stefani**[3]

[1] Independent, Munich, Germany

[2] Department of Accounting, Auditing, and Law, Norwegian School of Economics, Norway

[3] Department of Economics, University of Konstanz, Germany

[4] Universität Konstanz

Correspondence
Ulf Mohrmann, Norwegian School of Economics, Department of Accounting, Auditing, and Law, Helleveien 30, N- 5045 Bergen, Norway.
Email: ulf.mohrmann@nhh.no

Abstract

We use Benford's Law to provide evidence that German firms round up both their net income and earnings per share. We use the introduction of the euro to show that round earnings numbers are likely the result of earnings management. The incentive to round up comes from stakeholders' left-digit bias when processing the information in financial statements. Since round numbers are natural benchmarks, stakeholders perceive the performance metrics directly below such thresholds as abnormally lower. However, rounding up is objectionable only if it involves large-scale earnings management, but not in cases of negligible 'earnings cosmetics'. Because the difference between the pre-managed and reported earnings is unobservable, we investigate whether the prevalence of rounding up coincides with specific levels of several earnings characteristics and proxies for audit quality. If the rounding up is cosmetic, then it should occur independently of these characteristics. In contrast, if firms use earnings management on a larger scale, then it might not be possible to simultaneously round up and achieve other objectives of earnings management. Our evidence is in line with substantial earnings management.

KEYWORDS
audit quality, Benford's Law, cognitive thresholds, earnings characteristics, earnings management, left-digit bias, rounding up

JEL CLASSIFICATION
C46, M41, M42

1 | INTRODUCTION

Users of financial information tend to interpret a performance measure that is slightly below a critical value as abnormally lower than a performance measure that just beats this target. The most prominent benchmarks are analyst forecasts, prior year's earnings, and zero earnings (Burgstahler & Dichev, 1997; Degeorge, Patel, & Zeckhauser, 1999). However, less distinctive benchmarks, such as multiples of 10, also act as performance thresholds, because stakeholders commonly use them as cognitive reference points (Rosch, 1975). This behavior is similar to the €1.99 pricing phenomenon in retailing. If the realized performance falls below a given threshold, then managers might use their accounting discretion to shift the reported number just at or even above the critical value. This special kind of benchmark beating is usually called rounding up.

Beginning with Carslaw (1988), a large body of literature has found empirical evidence that firms round up their reported performance measures (e.g., Carslaw, 1988; Kinnunen & Koskela, 2003; Niskanen & Keloharju, 2000; Thomas, 1989). There are also results on the *conditions* that amplify rounding up, including the latitude of the accounting system and firms' exposure to the capital market (Kinnunen & Koskela, 2003; Niskanen & Keloharju, 2000). However, whether investors should be concerned about rounding up remains an open question. The lack of evidence regarding the assessment of rounding up is surprising, because Thomas (1989, p. 787) had already called for research on 'whether rounding up is a harmless practice'.

If rounding up occurs only if the pre-managed performance measure is slightly below a threshold, then the manipulation is most likely marginal and therefore not necessarily problematic from the addressees' perspective. Niskanen and Keloharju (2000) introduced the term *earnings cosmetics* for rounding up because of its presumed small-scale nature. This kind of manipulation could even prevent irrational investment decisions: without rounding up, the likelihood exists that investors will wrongfully discount a performance number that is directly below a threshold because of the left-digit bias that results from cognitive constraints (Bizer & Schindler, 2005; Lacetera, Pope, & Sydnor, 2012; Thomas & Morwitz, 2005). However, firms could also report rounded performance measures in cases where the difference between the pre-managed number and the next threshold is substantial. Rounding up on a large scale could harm addressees, since the reported earnings number is misleading.

Benford's Law has become an established tool for analyzing large data sets of accounting information to detect whether rounding up is present in a sample. The approach goes back to the work of Newcomb (1881) and Benford (1938) and describes the frequency distribution of the numerals of certain sets of numbers. For analyses of performance measures, the second and third digits (from the left) are essential. Rounding up results in a higher than expected frequency of the numeral zero as the second or third digit, while the numeral nine occurs less often than predicted. For illustration, consider a pre-managed net income with a nine as the second digit, for example, €3,980,000. If a substantial share of the firms in the sample report a net income of €4,000,000 to meet the next cognitive reference point, fewer nines and more zeros than expected will occur as the second digit.[1]

[1] Because managers' incentives reverse if a firm reports a loss, we expect a lower occurrence of zeros and a higher occurrence of nines than predicted by the Benford distribution for loss firms (Thomas, 1989).

The appealing feature of Benford's Law is its potential to detect a large set of accounting manipulations without specifying management's motivation or the accounting methods used to achieve a performance threshold (Skousen, Guan, & Wetzel, 2004). One drawback is that the tests cannot directly attribute the rounded numbers to intentional earnings management. We thus use the introduction of the euro as the reporting currency in Germany as an exogenous shock to firms' incentives to manage earnings. Since firms had incentives to round their performance measures reported in DMark (in euros) before (after) the change in the reporting currency, a change in rounding behavior around the euro's introduction is evidence in line with the earnings management explanation.

According to Benford's Law, we can expect around 12% of the firms in a sample to have a zero as the second digit of their performance measure. This percentage is higher if firms intentionally round up. However, it is impossible to differentiate firms that round up from the 12% of firms predicted to do so, because the pre-managed numbers are unobservable. For the same reason, the magnitude of the manipulation is indeterminable. Thus, whether rounding up is the result of earnings cosmetics or an indication of substantial earnings management remains unclear.

To investigate the extent of rounding up, we therefore analyze the cross-sectional variation in the rounding-up behavior between subsamples of German firms that differ in their earnings and auditor characteristics. If the earnings management is only cosmetic, rounding up should not be connected to the earnings and auditor characteristics, since all the firms in the sample have some incentives for rounding up. If the extent of rounding up is small, there should be no association with empirically observed earnings characteristics. The auditor's characteristics should not restrict management's options to slightly round up either. In contrast, using larger amounts of earnings management to round up the performance measures reported could interfere with other objectives related to earnings management. In this case, firms have to decide between realizing the preferred level of a specific earnings characteristic and reporting a rounded performance measure. For example, using a large amount of earnings management to round up could be incompatible with moderate (or even negative) levels of discretionary accruals, having a high-quality auditor, and achieving a smooth net income.

Our analysis indicates that the net income and earnings per share (EPS) of German firms reporting a profit reveal the characteristics that are associated with rounding up. Revenue, operating income, and cash flow from operations, in contrast, do not seem to be connected to rounding up. We show that the German firms in our sample round up either net income or EPS, but not both metrics simultaneously. However, for firms with a round net income and a round number of shares, a round net income mechanically leads to a round EPS. There are more firms with a round net income than with a round EPS. For the firms that report round numbers for both net income and EPS, the proportion with round EPS numbers is larger than the proportion with round net income numbers. Therefore, we restrict our main analyses to net income.

Using the introduction of the euro as an exogenous shock, we present evidence that is in line with the argument that round net income numbers are the result of earnings management. We show that firms rounded the net income reported in euros only in the periods after 2001 when reporting in euros became mandatory. In contrast, if we convert the net income originally reported in the 1990s in DMarks into euros, the data do not show evidence of rounding. Conversely, firms rounded the net income reported in DMarks in the 1990s, but not the net income converted into DMarks after 2001. This structural break in the change in the reporting currency points to deliberate actions of management as the reason for the deviation of the net income numbers from the Benford distribution.

Moreover, the results of our cross-sectional tests indicate that rounding up is concentrated in certain subsamples with specific earnings and auditor characteristics. We find deviations from the Benford distribution in subsamples of firms with high levels of discretionary accruals, low levels of earnings smoothing, less persistent net income, less timely loss recognition, nonzero extraordinary items, a non-Big 4 auditor, and an audit firm that is not an industry specialist. Complementary subsamples do not show deviations from the Benford distribution. Because rounding up occurs at specific levels of earnings characteristics, the argument that firms use only cosmetic earnings manipulations to achieve the targeted reference point is not evident.

For firms with negative net income, we find no link between earnings and auditor characteristics and rounding up. This result could be attributable to the opposing incentives that result from 'big bath' accounting (i.e., loss firms try to create cookie jar reserves for the future by maximizing reported losses).

Our analyses contribute to the literature in several ways. First, we provide evidence that one group of German firms rounds up net income, whereas a different group rounds up EPS. The German research has used Benford's Law as an analytical audit procedure (Quick & Wolz, 2003), but not to detect rounding up. Second, we use the introduction of the euro to show that rounding up is most likely the result of earnings management. Third, our results suggest that the term *earnings cosmetics* (Kinnunen & Koskela, 2003) that is associated with rounding up could be misleading. Thus, in the case in which the numeral zero is the net income's second digit, addressees should do additional analyses to avoid erroneously investing in these firms. Avoiding the risk of being fooled by rounded income numbers could warrant the additional costs of these analyses.

We proceed as follows: In Section 2, we review the literature and develop the hypotheses. In Section 3, we describe our research design and present the earnings and auditor characteristics for our empirical analyses. In Sections 4 and 5, we present our main results and additional analyses, respectively. Section 6 concludes the paper.

2 | BACKGROUND

2.1 | Benford's Law in an accounting context

The rounding-up literature commonly uses Benford's Law as a benchmark for the expected frequency distribution of specific numerals in numbers (Das & Zhang, 2003, and Ullmann & Watrin, 2017, are notable exceptions from using Benford's Law). Newcomb (1881) and Benford (1938) independently established this distribution by analyzing several sets of numbers drawn from populations, areas of rivers, atomic weights, and several other categories. Both authors conclude that the occurrence of numerals follows a logarithmic distribution. For the second digit (from the left), this distribution is given by

$$P(D_2 = d_2) = \sum_{d_1=1}^{9} \log_{10}\left(1 + \frac{1}{d_1 d_2}\right) \tag{1}$$

where d_1 (d_2) is the numeral that occurs as the first (second) digit of a number. The expected occurrence of numerals decreases from zero to nine, that is, more numbers should exist with a zero as the second digit than numbers with a nine.

Unfortunately, there is no clear-cut definition of the conditions that need to be fulfilled for a data set to follow Benford's Law (Hill, 1998). However, Hill (1995, p. 360) states that 'if probability distributions are selected at random and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant-digit frequencies of the combined sample will converge to the logarithmic distribution'. Although it is not possible to prove that a data set conforms to these conditions, 'in many real-life sampling procedures, they appear to be reasonable assumptions' (Hill, 1995, p. 361). Among different applications, Hill (1995) explicitly names accounting data as one area where the assumptions should hold and reviews anecdotal evidence for the conformity of different sets of accounting data with Benford's Law. Others discuss the appropriateness of Benford's Law on theoretical grounds (e.g., Durtschi, Hillison, & Pacini, 2004; Nigrini & Mittermaier, 1997) and conclude that most accounting data follow Benford's Law. Some of these studies also test individual accounts of specific companies for fraud (e.g., Nigrini & Mittermaier, 1997 detect the production of fictitious invoices). Additional evidence for the applicability of Benford's Law to accounting data is obtained from Amiram, Bozanic, and Rouen (2015). They show that the numbers collected from restated financial statements better conform to Benford's Law than the (misstated) numbers originally disclosed. In contrast to these applications, we follow Carslaw (1988), Thomas (1989), and others and apply Benford's Law to the pooled numbers of different performance measures that we collected from a large sample of firms. Pooling makes the assumption that the numbers are random samples drawn from different distributions even more likely. We argue that firms use earnings management to round up their performance measures and that this earnings management leads to a deviation from the Benford distribution.

For rounding up to occur, we have to assume that managers have (and will make use of) the opportunity to manage the reported performance measures. Healy and Wahlen (1999, p. 368) state that 'earnings management occurs when managers use judgment in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company or to influence contractual outcomes that depend on reported accounting numbers'. We do not predict *how* firms round up their performance measures, because each firm probably uses different actions: 'In fact, in many cases, each individual observation is likely to have been transformed by more than one earnings management action' (Burgstahler & Chuk, 2017, p. 739). However, since rounding up is reasonable only when management knows the unmanaged outcome—that is, after it has recorded the transactions that occurred during the fiscal year—the channels that involve *real* earnings management (i.e., timing and the use of artificial transactions) are no longer available. Instead, management will focus on different ways of *accounting* earnings management to round up (Xu, 2016).

To investigate benchmark beating, the literature usually uses (scaled) earnings.[2] Taking earnings as a target is in line with the survey conducted by Graham, Harvey, and Rajgopal (2005), in which a majority (51%) of the responding chief financial officers named earnings as the performance measure that is most important for external addressees. Only a few managers referred to alternatives, for instance, revenue, cash flows, and pro forma earnings, such as operating income (12% each). From these performance measures, firms can manage net income, operating income, and revenue with their accrual choices. Therefore, we follow the literature on rounding up and base our main analyses on net income and EPS. We also test for rounding up of revenues and operating income. Although cash flow from operations can no longer be managed when rounding is supposed to happen, we still test for rounding in the cash flow as a placebo. Given our understanding of the rounding process, we do not expect to find rounding in the cash flow from operations.

2.2 | Empirical evidence for the rounding up of performance measures

Carslaw (1988) presents evidence of rounding up in the net income of a sample of firms from New Zealand. The author finds that the second digit is rarely a nine but that the zero is more frequent. This result also holds for samples from the United States (Guan, He, & McEldowney, 2008a; Jordan & Clark, 2011; Thomas, 1989), the United Kingdom (Van Caneghem, 2002), Finland (Niskanen & Keloharju, 2000), Japan (Skousen et al., 2004), and Taiwan (Guan, Lin, & Fang, 2008b). In a cross-country study, Kinnunen and Koskela (2003) find the highest likelihood of rounding up in firms from Spain, Hong Kong and Singapore, and the lowest likelihood in firms from Norway, the United Kingdom and Sweden.

Evidence also exists of incentives for rounding up. Firms with a higher exposure to the capital market (Niskanen & Keloharju, 2000) and firms that apply bonus schemes (Kinnunen & Koskela, 2003) are more likely to deviate from Benford's Law. The institutional environment plays a role as well: Rounding up is more pronounced if Generally Accepted Accounting Principles (GAAP) allow greater discretion (Kinnunen & Koskela, 2003), but lessened since the passage of the Sarbanes–Oxley Act (Jordan & Clark, 2011). Moreover, an audit *per se* decreases the likelihood of rounding up (Guan, He, & Yang, 2006), and an auditor with a high degree of industry specialization appears to further restrict rounding up (Van Caneghem, 2004). There are also industry differences in the prominence of rounding up (Guan et al., 2008a).

[2] Most Benford analyses use unscaled earnings (following the example of Carslaw, 1988), while the literature on the zero earnings benchmark uses either earnings scaled by the market value of the firm (Burgstahler & Dichev, 1997; Degeorge, Patel, & Zeckhauser, 1999) or the EPS (Xu, 2016).

The German evidence is scarce. Quick and Wolz (2003) investigate single accounts from German firms and find that the data follow Benford's Law, although there are deviations when only balance sheet data are used. However, the authors focus on detecting deviations from Benford's Law in general, and not specifically on providing evidence of rounding up.

2.3 | Hypotheses on rounding up, earnings characteristics, and audit quality

There is a wide range of evidence that firms try to meet or beat earnings benchmarks (Burgstahler & Chuk, 2017). The most prominent earnings benchmarks are analyst forecasts, the prior year's earnings, and zero earnings (Dechow, Ge, & Schrand, 2010). Although the percentage of firms that report earnings slightly above such important benchmarks is higher than expected, it is still relatively low. For example, Cheng and Warfield (2005) find that around 25% of all firms meet analysts' consensus forecast, and Barth, Landsman, and Lang (2008) report that 13% of firms meet the zero-earnings benchmark. Put differently, for a majority of firms, these benchmarks are not relevant or are unachievable. However, they could still have incentives to meet less prominent benchmarks. Burgstahler and Dichev (1997) propose two general conditions that, in combination, create incentives for beating the benchmark. These conditions also hold for multiples of 10 as benchmarks.

First, firms are better off if they report higher rather than lower earnings, because higher earnings indicate higher firm value (Cornell & Shapiro, 1987; Graham et al., 2005), improve the terms of explicit and implicit contracts with stakeholders (Bowen, DuCharme, & Shores, 1995; Graham et al., 2005), and increase managers' bonus payments (Healy, 1985; Holthausen, Larcker, & Sloan, 1995; Indjejikian & Nanda, 2002; Murphy, 2001).[3] Moreover, analysts integrate their expectations that firms will beat the benchmark into their forecasts (Burgstahler & Eames, 2003). Therefore, 55% of annual EPS forecasts on I/B/E/S have the numerals zero or five in the penny location (Herrmann & Thomas, 2005). While investors correct this analyst bias for long-term forecasts, they fail to do so for short-term forecasts (Eames & Kim, 2012).

Second, investors and creditors usually use heuristics in their decision making, particularly if their deliberation costs are sufficiently high (Simon, 1955). Conlisk (1996) presents arguments supporting the use of heuristics in decision making in a general context, while Hirshleifer (2001) reviews the evidence for the use of heuristics in asset pricing and, thus, specifically in investors' decision making. More precisely, Hirshleifer (2001, p. 1545) refers to the 'cognitive efficiency of mentally discretizing continuous variables', which means that decision makers intuitively memorize the first digit(s) of a number but do not round the number correctly. The heuristic of focusing on the left digit(s) could go back to the fact that rounding up is more complex than rounding down, which involves simply dropping off the rightmost digits of a number (Bizer & Schindler, 2005; Brenner & Brenner, 1982). The left-digit bias (Bizer & Schindler, 2005; Lacetera et al., 2012; Thomas & Morwitz, 2005) leads to kinks in the investors' utility functions around multiples of 10; that is, investors perceive a value that is directly below a round number (e.g., €3,999,999) as disproportionately lower than a value directly above a round number (e.g., €4,000,001). In line with this argument, Rosch (1975) shows experimentally that multiples of 10 act as cognitive reference points.

In an investor-specific context, Bhattacharya, Holden, and Jacobsen (2012) find evidence of the use of rounded prices as reference points in trading. They find higher buy–sell ratios for liquidity demanders at all price points one penny below integers, half-dollars, quarters, dimes, and nickels, but excess selling by liquidity demanders at all price points one penny above these reference points. The authors find the greatest imbalance between buys and sells around integers.

In addition to the conditions referred to by Burgstahler and Dichev (1997), there are further incentives for managers to round up firm performance measures: Since bonus plans often have a floor, managers have incentives to overstep this threshold (Indjejikian & Nanda, 2002; Murphy, 2001). The bonus can further increase if the key

[3] An exception is when firms have incentives for a big bath (Healy, 1985).

performance indicators exceed additional thresholds (Câmara, 2001; Holthausen et al., 1995).[4] If earnings are used in covenants, management has a strong incentive to round up earnings to avoid technical default (Dichev & Skinner, 2002; Guan et al., 2008b). Although the role of balance sheet covenants has declined over time, the role of earnings-based covenants in debt contracting remains stable (Demerjian, 2011).

Taken together, we expect that German firms will round up some of their performance measures. Formally, we investigate the following hypothesis (stated in alternative form):

H1: German firms use earnings management to round up their performance measures.

The literature finds evidence that management achieves rounding up by its accrual choices. However, whether the earnings management used to achieve the benchmark is problematic for investors remains an open question. Niskanen and Keloharju (2000) and Kinnunen and Koskela (2003) argue that only small-scale manipulations are required and introduce the term *cosmetic earnings management* for rounding up. The level of earnings management necessary to reach the cognitive thresholds of investors and creditors could be so small that it remains below any materiality threshold. However, firms still benefit from rounding up if addressees use heuristics in their decision making. Rounding up could even prevent suboptimal decisions, because addressees no longer downgrade those firms that report earnings directly below the threshold. However, based on investors' left-digit bias, management's incentive to report a performance measure above a reference point could be so strong that firms use earnings management on a larger scale to overstep the threshold. Although rounding up is most likely not an option for a pre-managed net income of, for example, €3,400,000, the strategy is less clear for a pre-managed net income of €3,750,000. If firms use more than earnings cosmetics to reach the benchmark, rounding will likely reduce the decision usefulness of financial statements.

A direct evaluation of the magnitude of earnings management is not possible, because the pre-managed performance measures are not observable. Therefore, we relate the prevalence of rounding up to different earnings characteristics. Our underlying assumption is the independence between the incentives for rounding up and our earnings characteristics, because addressees' cognitive biases should not be related to any earnings characteristics. If rounding up occurs only when the pre-managed performance measure is directly below the threshold, we do not expect a difference in the digit distribution for the subsamples built on high versus low levels of earnings characteristics. For example, firms with more or less timely loss recognition and firms with more or less smooth earnings could slightly round up their net income to meet the next threshold without changing the earnings characteristics. However, if managers use nontrivial amounts of earnings management, rounding up should be associated with our earnings characteristics. The direction of the association depends on whether the specific characteristic complements or interferes with rounding up. For example, less timely loss recognition could be complementary to rounding up. In contrast, earnings smoothing could be opposed to rounding up if the target for smoothing does not coincide with a zero as the performance measure's second digit. Similarly, if the firm uses substantial amounts of earnings management to round up its performance measures, we expect an auditor of higher quality to limit this behavior. If, however, earnings management is only cosmetic, it probably falls below the auditor's materiality threshold. These examples illustrate that we do not expect a causal link between rounding up and earnings quality. Even in the case where the earnings characteristic and rounding up are complementary, we do not claim that the earnings management used in the rounding-up process is the driving force behind the formation of the earnings characteristic.

Formally, we investigate the following hypothesis (stated in null form):

H2: There is no relation between rounding up and the earnings and auditor characteristics that we investigate.

[4] Holthausen, Larcker, and Sloan (1995) do not present evidence of the use of rounded numbers as thresholds, because they linearly transform all thresholds to ensure confidentiality. We adopt this assumption from Kinnunen and Koskela (2003) and Thomas (1989).

3 | RESEARCH DESIGN

3.1 | Benford test

Rounding affects the frequency distribution of the numerals in the reported performance measures. The zero and nine as second digits are important for testing our hypotheses. If firms round up, the numeral zero (nine) should be over-represented (under-represented) in firms with a positive value for the performance measure. We expect the reverse pattern for firms with a negative value. Benford's Law determines the expected frequencies. For our primary test, we rely on the *Z*-statistic (Carslaw, 1988; Thomas, 1989) that compares—separately for each numeral—the relative frequency observed with the frequency predicted by Benford's Law:

$$Z = \frac{|p - p_0| - \frac{1}{2n}}{\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{2}$$

where p is the observed proportion of numerals in the reported performance measure's second digit, p_0 is the expected proportion according to Benford's Law, and n is the sample size. The term $1/2n$ is a continuity correction that is used only if the correction term is smaller than the absolute value term (Thomas, 1989).

To test hypothesis H1, we investigate the empirical distribution of the second digit of different performance measures. Specifically, we use net income, EPS, revenue, operating income, and cash flow from operations (as a placebo test). We further distinguish between positive and negative values of the performance measure, because the expected pattern of the rounding-up manipulation differs between these cases (Thomas, 1989).

The Benford test cannot identify the reasons for the deviations from the Benford distribution. To strengthen our earnings management explanation, we compare the performance measures of German firms in euros and DMarks around the introduction of the euro. If rounding up is the result of management action, only the performance measures in the reporting currency should be rounded. To test this assumption, we convert the performance measures for 2001 and later years into DMarks. The converted DMark amounts should not indicate rounding. Similarly, we compare the distributions of the performance measures for the years before 1999, reported in DMarks, with those converted into euros.[5] A problem of this test is the DMark–euro exchange rate, where DM1 = €0.511292. If we convert numbers in DMark with an even first digit into euros, the converted number will keep the same second digit as the original number. For example, a net income of DM20,900 converts into €10,686. To avoid this mechanical effect, we run our analysis only on observations whose respective performance measure starts with an odd numeral (e.g., a net income of DM30,600 converts into €15,645).[6]

To test hypothesis H2, we form subsamples that we construct conditionally on frequently used earnings and auditor characteristics and use Benford's Law to evaluate whether the performance measures in these subsamples show patterns consistent with rounding up.

Because our test procedure essentially consists of 10 separate tests (one for each numeral), Cleary and Thibodeau (2005) warn against the danger of Type I errors. Therefore, we perform a chi-squared goodness-of-fit test as a second test (Carslaw, 1988; Thomas, 1989):

$$\chi^2 = \sum_{d=0}^{9} \frac{\left(p^d - p_0^d\right)^2}{p_0^d} \tag{3}$$

[5] We exclude the years 1999 and 2000 because in these years German firms could choose between reporting in DMarks or in euros. Unfortunately, Worldscope does not include information on the reporting currency.

[6] In untabulated analyses, we repeat all analyses with only observations with odd first digits. All the inferences are the same.

A disadvantage of this test is the possibility that irregularities other than rounding up are causal for rejecting the null. Moreover, Shikano and Mack (2011) argue that the test can erroneously indicate deviations from the Benford distribution. They show that inflated test statistics are likely if the maximum value of the density distribution is at least above 0.01. In our sample, this value is only 0.0011 and, thus, the chi-squared test should be uncompromised.

3.2 | Earnings characteristics

To construct our earnings characteristics, we build on the overview in Exhibit 1 of Dechow et al. (2010). From the proxies discussed there, we choose discretionary accruals, smoothing, persistence and conservatism in the form of timely loss recognition.

The literature uses discretionary accruals (*DA*) to estimate the portion of accruals that the firm uses to manage its earnings (Dechow et al., 2010). We apply the modified Jones model to estimate discretionary accruals (Dechow, Sloan, & Sweeney, 1995):

$$TACC_{it}/TA_{it-1} = \beta_0 1/TA_{it-1} + \beta_1 \left(\Delta REV_{it} - \Delta REC_{it}\right)/TA_{it-1} + \beta_2 PPE_{it}/TA_{it-1} + \varepsilon_{it} \tag{4}$$

where *TACC* is total accruals, defined as net income minus cash flow from operations; *TA* is total assets; *REV* is revenue; *REC* is accounts receivables; and *PPE* is gross property, plant and equipment. We estimate equation 4 for every industry–year combination with at least 10 observations, where industries are classified following Barth, Beaver, and Landsman (1998). The discretionary accruals are defined as the residuals of equation 4.

For our analysis, we form terciles based on discretionary accruals.[7] If rounding up is only cosmetic, we do not expect differences between the firms in the terciles of discretionary accruals with regard to rounding up. However, if profit firms use substantial amounts of discretionary accruals to round up their metrics, the deviation from the Benford distribution should be strongest in the third tercile.

We use two metrics of Leuz, Nanda, and Wysocki (2003) that have been frequently applied in accounting research (e.g., Perotti & Wagenhofer, 2014) to measure the smoothness of net income: *SMOOTH1*, the ratio of the firm-level standard deviation in net income to the firm-level standard deviation in cash flow from operations, and *SMOOTH2*, the firm-level Spearman correlation between the change in accruals (scaled by lagged total assets) and the change in cash flow from operations (scaled by lagged total assets). In both measures, the cash flow from operations is the reference point for the firm's unmanaged performance; larger values of *SMOOTH1* and *SMOOTH2*, respectively, indicate earnings that are less smooth. We drop all firms for which fewer than 10 observations are available to compute the standard deviations or correlations.[8] Based on each metric, we form two subsamples based on a median split. If firms use substantial earnings management to round up their performance measures, they might not be able to smooth these numbers simultaneously, because the target earnings for income smoothing might be any number (e.g., €2.4 million if the profits were €2.2 million in t–2 and €2.3 million in t–1).

Managers have incentives to report persistent earnings numbers because investors perceive this pattern as less risky (Graham et al., 2005). We estimate persistence with the following regression model:

$$NI_{it}/TA_{it-1} = \beta_0 + \beta_1 NI_{it-1}/TA_{it-2} + \varepsilon_{it} \tag{5}$$

[7] We form our subsamples based on all observations of the respective earnings and auditor characteristics of a firm instead of separately for firms with positive and negative net income, because some of the earnings characteristics are only available at the firm level and not at the firm–year level (e.g., smoothing and persistence metrics). However, in untabulated analyses, we show that all the inferences are the same if we use the alternative formation of the subsamples.

[8] All the inferences are the same if we reduce the number of required observations per firm to five. Moreover, we compare firms with fewer than 10 observations with the remaining firms. The firms with fewer than 10 observations are, on average, smaller (with a logarithm of total assets of 17.384 versus 18.441, significant at the 1% level). However, there are no significant differences in profitability (e.g., return on assets), risk (e.g., leverage), or, most importantly, earnings characteristics.

where *NI* is net income. We estimate equation 5 at the firm level for all firms with at least 10 observations.[9] Our estimate of persistence is β_1. We use a median split to form two subsamples that represent firms with either high or low persistence. Similar to earnings smoothing, in the case of substantial earnings management in the rounding-up process, both goals might not be possible, resulting in a concentration of rounding up in the low-persistence group.

Conservatism in the form of timely loss recognition, that is, the asymmetric recognition of profits and losses, increases the decision usefulness of financial statements (Watts, 2003). We approximate the level of conservatism at the firm–year level with the C-score of Khan and Watts (2009), which measures the level of conservatism as a function of firm size, the market-to-book ratio, and the leverage ratio:

$$\begin{aligned} EPS_{it} = {} & \beta_1 + \beta_2 D_{it} + R_{it}\left(\mu_1 + \mu_2 SIZE_{it} + \mu_3 MB_{it} + \mu_4 LEV_{it}\right) \\ & + D_i R_i \left(\lambda_1 + \lambda_2 SIZE_{it} + \lambda_3 MB_{it} + \lambda_4 LEV_{it}\right) \\ & + \left(\delta_1 SIZE_{it} + \delta_2 MB_{it} + \delta_3 LEV_{it} + \delta_4 D_{it} SIZE_{it} + \delta_5 D_{it} MB_{it} + \delta_6 D_{it} LEV_{it}\right) + \varepsilon_{it} \end{aligned} \tag{6}$$

where *EPS* is earnings per share, *D* is a dummy variable that equals one if the firm reports a negative return, *R* is the annual return (calculated from month −9 to month +3, relative to the fiscal year end), *SIZE* is the natural logarithm of the market value of equity, *MB* is the market-to-book ratio, and *LEV* is the leverage ratio (defined as total liabilities divided by the market value of equity). We estimate Equation 6 with annual cross-sectional regressions. The C-score is defined as:

$$CSCORE_{it} = \hat{\lambda}_1 + \hat{\lambda}_2 SIZE_{it} + \hat{\lambda}_3 MB_{it} + \hat{\lambda}_4 LEV_{it} \tag{7}$$

where $\hat{\lambda}_i$ is the coefficient estimate from equation 6. A higher C-score indicates earnings that are more conservative, that is, the asymmetry in the recognition of profits and losses is greater. We form two subsamples based on a median split that represent firms with either higher or lower levels of conditional conservatism. Managers might recognize revenues earlier or defer the recognition of losses to the future to round up their performance measures. If this manipulation is not just cosmetic, we expect firms that round up to be concentrated in the subsample with low C-score values.

3.3 | Proxies for auditor quality

As proxies for audit quality, we choose the size and industry specialization of the audit firm from the measures discussed by DeFond and Zhang (2014) (see their Table 2 for an overview). We measure the audit firm's size using the Big 4/non-Big 4 dichotomy (*BIG4*). There is empirical evidence that Big 4 auditors restrict the usage of discretionary accruals and improve several other measures of earnings quality (Becker, DeFond, Jiambalvo, & Subramanyam, 1998; Francis, Maydew, & Sparks, 1999). Because this variable is already bivariate, we form two subsamples: one comprises firm–years audited by a Big 4 auditor and the other comprises firm–years audited by a non-Big 4 auditor.

The empirical audit research often uses the auditor's industry specialization (*EXPERT*) as a measure of audit quality (DeFond & Zhang, 2014). The best way to measure industry specialization is the market share based on audit fees (Audousset-Coulier, Jeny, & Jiang, 2016). However, the disclosure of audit fees was not mandatory during the full sample period. Accordingly, we measure the industry specialization of an audit firm by its market share measured by total assets audited in a given industry–year combination as the second-best alternative (Audousset-Coulier et al., 2016). The industries are again defined following Barth et al. (1998). We base our subsamples for the Benford analyses on a median split. If earnings management is nontrivial, high-quality auditors should restrict rounding up. Alternatively, if firms use only earnings cosmetics, rounding up should arise with both types of auditors.

[9] All inferences are unchanged if we reduce the number of required observations per firm to five.

4 | EMPIRICAL ANALYSIS

4.1 | Sample description

The initial sample consists of the group accounts of all German listed firms that are covered by the Worldscope database from 2001 to 2012 (11,004 firm–years).[10] We drop all firm–years with no information on net income (2,573 observations). The remaining sample of 8,431 observations splits into 5,556 firm–years with a positive net income and 2,875 firm–years with a negative net income. The data requirements of the earnings characteristics reduce the sample size further. To maintain the power of the tests as high as possible, we drop firm–years with missing data only for the test for which the specific variable is missing.[11] For the discretionary accruals and smoothing samples, we further drop all financial institutions, because their accrual measures differ from those of the remaining firms. Panel A of Table 1 summarizes the sample selection and shows the sample sizes for the different earnings quality metrics.

The mandatory introduction of the International Financial Reporting Standards (IFRS) in 2005 represents a potential structural problem. We address this issue in two ways (results for both analyses are untabulated). First, we separately compare the fit of net income to Benford's Law for the IFRS and German GAAP (*Handelsgesetzbuch*) subsamples and find no differences between these subsets. Second, we replicate all of our tests with only IFRS firm–years and the results are no different from those reported below.[12] In Panel B of Table 1, we report the descriptive statistics for the performance measures. Net income ranges from a loss of €196 million at the first percentile to a profit of €2.2 billion at the 99th percentile.

4.2 | Rounding up by German firms

Table 2 presents the results for the Benford test on the subsamples of firm–years with positive and negative values in different performance measures separately, but without taking the earnings characteristics into account. This test checks whether these performance measures are rounded up in the German group accounts and whether the likelihood of rounding up differs between profit and loss firms.

In Panel A of Table 2, we report our evidence for net income. Firm–years with a positive net income have higher numbers of zeros as the second digit than predicted by Benford's Law. As expected under the rounding-up hypothesis, we also find a lower than predicted number of nines in this subsample. Most of the other numerals appear as often as predicted, except the numeral six, which appears less often than expected. The deviation is also economically significant: the difference between the expected and observed percentages of zeros as the second digit is 0.973. Thus, for approximately 7.5% (0.973%/12.941%) of all profit firms that report a round number, the zero as the second digit results from rounding up. The magnitude of rounding up we observe is in line with earlier research (Kinnunen & Koskela, 2003; Niskanen & Keloharju, 2000; Thomas, 1989). The chi-squared test indicates that the second digits of the positive net income as reported by the German firms in our sample do not follow Benford's Law.

For loss firms, only the numeral one occurs significantly less often than predicted by Benford's Law, so rounding up seems less likely. A possible explanation for this finding is the competing incentive that results from big bath accounting. In the case of a loss, managers can maximize the reported loss to create the potential to report higher earnings in the future (Healy, 1985; Kirschenheiter & Melumad, 2002). If the big bath incentive is stronger than the incentive to avoid overstepping cognitive thresholds, we should not observe an excess of nines, because firms maximize the loss instead of trying to avoid rounded negative numbers. For example, under the rounding hypothesis, we expect a firm

[10] For the test on the introduction of the euro, we also use the years from 1990 to 1998.

[11] There are only 1,823 observations with a positive net income and all available earnings and auditor characteristics (of the 5,556 observations with a positive net income). However, we replicate all our analyses with this restricted sample (not tabulated) and find qualitatively similar results.

[12] The sample of German-GAAP firm–years is too small to conduct a similar analysis.

TABLE 1 Descriptive statistics

Panel A: Sample selection	Observations	Positive *NI*	Negative *NI*
Initial sample	11,004		
Firm–years with zero or missing net income (*NI*)	2,573		
Firm–years for the estimation of the earnings characteristics	8,431	5,556	2,875
Firm–years with non-missing *DA*	4,921	3,403	1,518
Firm–years with non-missing *SMOOTH1*	4,425	3,216	1,209
Firm–years with non-missing *SMOOTH2*	4,776	3,444	1,332
Firm–years with non-missing *PERSISTENCE*	4,956	3,568	1,388
Firm–years with non-missing *CSCORE*	4,529	3,389	1,140
Firm–years with non-missing *EXTRAORD*	7,774	5,231	2,543
Firm–years with non-missing auditor information	7,969	5,309	2,660

Panel B: Descriptive statistics	*N*	Mean	SD	P1	P25	P50	P75	P99
NI	8,452	6.66E+7	6.80E+8	−1.96E+8	−6.53E+5	8.60E+5	8.78E+6	2.20E+9
EPS	7,854	3.06	110.34	−40.78	−0.20	0.23	1.18	58.08
OPINCOME	5,621	1.70E+8	9.08E+8	−2.23E+8	−7.58E+5	1.85E+6	2.24E+7	5.11E+9
REVENUES	8,050	2.12E+9	1.00E+10	0.01E+0	9.85E+6	5.42E+7	2.86E+8	5.64E+10
CFO	6,771	2.59E+8	1.51E+9	−1.10E+8	−2.38E+5	4.15E+6	2.65E+7	6.63E+9

Notes: Panel A presents the sample selection and sample sizes for the subsample analyses. Panel B provides the descriptive statistics for our sample, where *NI* is net income, *DA* is discretionary accruals estimated with the modified Jones model, *SMOOTH1* is the ratio of the standard deviation of net income and the standard deviation of cash flow from operations, *SMOOTH2* is the Spearman correlation between the change in total accruals and the change in cash flow from operations, *PERSISTENCE* is the coefficient for the regression of net income scaled by lagged total assets on lagged net income scaled by total assets of t-2, *CSCORE* is the C-score of Khan and Watts (2009), *EXTRAORD* is a dummy variable that equals one if the firm reports non-zero extraordinary items, *EPS* is earnings per share, *OPINCOME* is operating income, *REVENUES* is revenues, and *CFO* is cash flow from operations.

with a net income of −€605,000 to manage the net income to −€599,000, while we expect it to report even higher losses under the big bath hypothesis. Moreover, if investors perceive the occurrence of a loss as a stronger signal than the magnitude of the loss (Hayn, 1995), managers might not have an incentive to round a negative net income.

In Panel B of Table 2, we repeat the analysis for EPS. Thomas (1989) argues that, for EPS, the third digit (from the left) is the most relevant for rounding and finds more than the expected number of zeros and fives in the sample, which indicates an additional reference point for EPS. Thus, firms do not round to the next full euro, but to the next multiple of five cents. We restrict our sample to firms with EPS between €1.00 and €9.99 to ensure that the third digit measures the same unit. As predicted by the rounding hypothesis, we find significantly more than the expected number of zeros (and ones) for firms with positive EPS. However, we do not find a significant lack of the numeral nine. Thus, our evidence for rounding up is weak. Similar to the net income analysis, we find no evidence of rounding in the observations with negative EPS.

In Panel C of Table 2, we use operating income as our performance measure. We find no evidence of rounding for firms with either a positive or a negative operating income. We also find no deviation from the Benford distribution in the revenues (Panel D of Table 2). The surveyed managers of Graham et al. (2005) do not think that operating income or revenues are the most relevant performance measure for investors. Accordingly, it is not surprising that managers

TABLE 2 Benford analysis for the full sample

Panel A: Benford analysis for net income					
		Positive net income		Negative net income	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	11.968	12.941	0.973**	12.765	0.797
1	11.389	11.843	0.454	9.774	−1.615***
2	10.882	11.249	0.367	10.783	−0.100
3	10.433	10.601	0.168	10.157	−0.276
4	10.031	9.557	−0.474	10.017	−0.013
5	9.668	10.205	0.537	10.052	0.384
6	9.337	8.657	−0.680*	9.530	0.193
7	9.035	8.999	−0.036	8.974	−0.061
8	8.757	8.387	−0.370	9.357	0.600
9	8.500	7.559	−0.940**	8.591	0.092
Chi^2		18.549		10.124	
p-value		0.029		0.341	
Number of obs.		5,556		2,875	
Panel B: Benford analysis for EPS					
		Positive EPS		Negative EPS	
3rd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	10.178	11.886	1.707***	9.855	−0.323
1	10.138	11.193	1.055**	9.648	−0.490
2	10.097	10.418	0.321	10.186	0.089
3	10.057	9.541	−0.516	10.973	0.916
4	10.018	10.357	0.339	10.269	0.251
5	9.979	9.602	−0.376	9.979	0.001
6	9.940	8.502	−1.439***	9.524	−0.416
7	9.902	9.582	−0.320	9.772	−0.130
8	9.864	9.195	−0.669	10.145	0.281
9	9.827	9.725	−0.102	9.648	−0.179
Chi^2		35.491		3.738	
p-value		<0.001		0.928	
Number of obs.		4,905		2,415	
Panel C: Benford analysis for operating income					
		Positive operating income		Negative operating income	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	11.968	12.037	0.069	12.929	0.961
1	11.389	11.660	0.271	10.396	−0.993
2	10.882	10.989	0.107	10.501	−0.381

(Continues)

TABLE 2 (Continued)

Panel C: Benford analysis for operating income					
		Positive operating income		Negative operating income	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
3	10.433	10.854	0.421	9.868	−0.565
4	10.031	9.296	−0.735	11.293	1.262*
5	9.668	8.974	−0.694	10.026	0.359
6	9.337	9.404	0.066	9.235	−0.103
7	9.035	9.054	0.019	7.863	−1.172
8	8.757	9.027	0.270	8.602	−0.155
9	8.500	8.705	0.205	9.288	0.788
Chi^2		5.300		11.537	
p-value		0.807		0.241	
Number of obs.		3,722		1,895	

Panel D: Benford analysis for revenues			
		Revenues	
2nd digit	Benford (%)	Observed (%)	Difference
0	11.968	12.225	0.258
1	11.389	11.426	0.037
2	10.882	11.134	0.252
3	10.433	10.880	0.447
4	10.031	9.471	−0.560*
5	9.668	9.598	−0.070
6	9.337	9.433	0.095
7	9.035	8.874	−0.161
8	8.757	8.645	−0.112
9	8.500	8.315	−0.184
Chi^2		5.646	
p-value		0.775	
Number of obs.		7,877	

Panel E: Benford analysis for cash flow from operations					
		Positive operating cash flow		Negative operating cash flow	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	11.968	12.050	0.082	12.152	0.184
1	11.389	10.782	−0.608	12.099	0.709
2	10.882	11.027	0.145	10.225	−0.657
3	10.433	10.352	−0.081	9.957	−0.476
4	10.031	9.472	−0.559	9.957	−0.074

(Continues)

TABLE 2 (Continued)

Panel E: Benford analysis for cash flow from operations					
		Positive operating cash flow		Negative operating cash flow	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
5	9.668	10.291	0.623	9.636	–0.032
6	9.337	9.902	0.564	8.887	–0.451
7	9.035	9.002	–0.034	9.154	0.119
8	8.757	8.940	0.183	9.690	0.933
9	8.500	8.183	–0.316	8.244	–0.256
Chi^2		7.655		4.472	
p-value		0.569		0.878	
Number of obs.		4,888		1,868	

Notes: This table presents the expected distribution according to Benford's Law (Benford) and actual occurrences in the data set (observed) for different performance measures. The *, **, and *** denote significance at the 10%, 5%, and 1% levels (based on the *Z*-statistic), respectively. The chi-squared test reported at the bottom of the panels tests the overall fit with the Benford distribution.

do not allocate resources to managing these measures. In Panel E, we investigate the cash flow from operations, which should not be rounded, according to our understanding of the rounding process. Indeed, we find no significant deviation from the Benford distribution.

To sum up, our results indicate rounding in firms with both a positive net income and positive EPS. However, for most firms, it is not possible to round up both metrics simultaneously. One reason is that firms have only limited discretion over the number of shares. To verify this argument, we rerun the analysis for EPS (net income) separately for firms that have potentially rounded their net income (EPS) and those that have not. Table 3 presents the results.

In Panel A of Table 3, we repeat the net income analysis for the subsamples based on whether EPS were potentially rounded or not. We find evidence of rounding the net income in the subsample without a round EPS, but not in the subsample with potentially rounded EPS. That is, most firms round either EPS or net income, but not both measures simultaneously. Unfortunately, we cannot test whether firms prefer rounding net income over rounding EPS (or vice versa). We also cannot determine whether rounding one of these measures is just easier because the pre-rounded value is closer to the next round number. Additionally, a limited number of firms have round numbers in both their EPS and net income. Panel B presents weak evidence of rounding up EPS in both subsamples of firms that potentially have or have not rounded their net income. However, this is a mechanical effect due to the number of shares outstanding: firms with a round net income and a round number of shares will also report a round EPS. This group comprises a significant part of those firms with a round EPS number, but not the group with a round net income. Therefore, we restrict our further analyses to net income.

Next, we turn to the question of whether the deviation from Benford's Law results from earnings management. The fact that only the numerals zero and nine deviate from the expected values is initial evidence that earnings management is the most likely explanation for the round numbers. As a more formal test, we compare the net income reported in euros with the net income converted into DMarks for 2002 to 2012 and the net income reported in DMarks with the net income converted into euros for 1990 to 1998. Table 4 presents the results.

In Panel A of Table 4, we investigate the observations from 2002 to 2012, when the euro was the reporting currency. The euro analysis is identical to the main analysis in Table 2 and therefore indicates rounding. However, we find no evidence of rounding in the values converted into DMarks. In Panel B, we repeat the analysis for the 1990s, when the DMark was the reporting currency. In this analysis, we use only observations with a positive net income that have

TABLE 3 Simultaneous rounding of net income and EPS

Panel A: Benford analysis for net income, conditional on the third digit of the EPS					
		Zero 3rd digit in EPS		Nonzero 3rd digit in EPS	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	11.968	13.474	1.506	12.874	0.907*
1	11.389	11.201	−0.188	11.923	0.534
2	10.882	10.714	−0.168	11.316	0.434
3	10.433	10.065	−0.368	10.668	0.235
4	10.031	9.253	−0.778	9.595	−0.436
5	9.668	10.714	1.047	10.142	0.474
6	9.337	9.253	−0.084	8.583	−0.754*
7	9.035	8.604	−0.431	9.049	0.013
8	8.757	9.416	0.659	8.259	−0.498
9	8.500	7.305	−1.195	7.591	−0.909**
Chi^2		3.822		17.037	
p-value		0.923		0.048	
Number of obs.		616		4,940	

Panel B: Benford analysis for EPS, conditional on the second digit of the net income					
		Zero 2nd digit in net income		Nonzero 2nd digit in net income	
3rd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	10.178	12.750	2.571**	11.754	1.575***
1	10.138	13.978	3.841***	10.766	0.629
2	10.097	9.524	−0.573	10.555	0.458
3	10.057	8.449	−1.609	9.709	−0.349
4	10.018	9.217	−0.801	10.531	0.513
5	9.979	9.217	−0.762	9.661	−0.317
6	9.940	7.988	−1.952	8.580	−1.360***
7	9.902	9.524	−0.378	9.591	−0.311
8	9.864	9.524	−0.340	9.144	−0.720
9	9.827	9.831	0.004	9.709	−0.118
Chi^2		19.052		25.600	
p-value		0.025		0.002	
Number of obs.		651		4,254	

Notes: This table presents the expected distribution according to Benford's Law (Benford) and the actual occurrences in the data set (observed) for net income (Panel A) and EPS (Panel B). The *, **, and *** denote significance at the 10%, 5%, and 1% levels (based on the *Z*-statistic), respectively. The chi-squared test reported at the bottom of the panels tests the overall fit with the Benford distribution.

TABLE 4 Benford analysis for net income in euros and in DMark

Panel A: Benford analysis for net income in the years after the euro's introduction					
		Net income in euros		Net income converted into DMark	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	11.968	12.941	0.973**	12.332	0.364
1	11.389	11.843	0.454	11.267	−0.122
2	10.882	11.249	0.367	11.010	0.128
3	10.433	10.601	0.168	11.050	0.617
4	10.031	9.557	−0.474	9.471	−0.560
5	9.668	10.205	0.537	9.215	−0.453
6	9.337	8.657	−0.680*	9.629	0.292
7	9.035	8.999	−0.036	9.155	0.120
8	8.757	8.387	−0.370	8.287	−0.470
9	8.500	7.559	−0.940**	8.583	0.084
Chi^2		18.549		7.072	
p-value		0.029		0.630	
Number of obs.		5,556		5,556	

Panel B: Benford analysis for net income in the years before the euro's introduction					
		Net income in DMark		Net income converted into euros	
2nd digit	Benford (%)	Observed (%)	Difference	Observed (%)	Difference
0	11.968	13.688	1.720*	13.089	1.121
1	11.389	11.421	0.032	4.974	−6.415***
2	10.882	10.811	−0.071	4.799	−6.083***
3	10.433	10.898	0.465	5.323	−5.110***
4	10.031	8.806	−1.225	5.497	−4.533***
5	9.668	11.595	1.928**	9.948	0.280
6	9.337	9.677	0.340	14.311	4.973***
7	9.035	7.759	−1.276	13.089	4.054***
8	8.757	7.062	−1.695**	15.358	6.601***
9	8.500	8.282	−0.217	13.613	5.113***
Chi^2		15.241		277.300	
p-value		0.085		<0.001	
Number of obs.		1,147		1,147	

Notes: This table presents the expected distribution according to Benford's Law (Benford) and the actual occurrences in the data set (observed) for the net income. In Panel A, we use observations after 2001, when reporting in euros became mandatory. In Panel B, we use observations for 1998 and earlier, when reporting in DMarks was still mandatory. In this panel, we only use observations where the net income in DMarks has an odd first digit, to ensure that the converted euro values do not mechanically have the same second digit as the DMark value because of the exchange rate. The *, **, and *** denote significance at the 10%, 5%, and 1% levels (based on the *Z*-statistic), respectively. The chi-squared test reported at the bottom of the panels tests the overall fit with the Benford distribution.

TABLE 5 Analysis for firms with a positive net income, conditional on earnings and auditor characteristics

		Findings	
Earnings characteristics	**Subsample**	**Zeros**	**Nines**
Discretionary accruals	1st tercile of *DA*	No deviation	Fewer nines
	2nd tercile of *DA*	No deviation	No deviation
	3rd tercile of *DA*	More zeros	No deviation
SMOOTH1	Less smooth *NI*	No deviation	Fewer nines
	Smoother *NI*	No deviation	No deviation
SMOOTH2	Less smooth *NI*	More zeros	Fewer nines
	Smoother *NI*	No deviation	No deviation
Persistence	Less persistent *NI*	No deviation	Fewer nines
	More persistent *NI*	No deviation	No deviation
Conservatism	Less conservative *NI*	More zeros	Fewer nines
	More conservative *NI*	No deviation	Fewer nines
Audit firm size	Non-Big 4	More zeros	Fewer nines
	Big 4	No deviation	No deviation
Auditor's industry specialization	No industry expert	More zeros	Fewer nines
	Industry expert	No deviation	No deviation

Note: This table summarizes the Benford analyses conditional on the earnings and auditor characteristics. It denotes for each subsample whether the deviation from Benford's Law is significant at the 5% level or not. The full results are reported in Table OA2 in the Online Appendix.

an odd number as the first digit. We find some evidence of rounding: there are more zeros than expected; although the numeral nine occurs as often as predicted, we find significantly fewer eights than expected. For the values converted into euros, we note that the distribution strongly deviates from Benford's Law for nearly all numerals, which is not in line with rounding. This finding reflects the effect of the exchange rate and the use of only odd-numbered DMark values. Importantly, this special sample selection does not necessarily lead to such an unusual digit distribution. If we use only the net income numbers with odd first digits in the firm–years that originally reported in euros, we still find evidence of rounding up, but the numerals one to eight follow Benford's Law (we report the results in the Online Appendix). Taken together, we find evidence of rounding up the net income only for the reporting currency (i.e., in situations where firms have incentives to round), but not for the alternative currency. Moreover, when the incentives change due to the change in the reporting currency, firms adjust their targets for rounding from the net income in DMark to the net income in euros. This targeting strongly points to earnings management as an explanation for the deviations from the Benford distribution (Burgstahler & Chuk, 2017).

4.3 | Analyses based on earnings characteristics and proxies for audit quality

Having established that firms use earnings management to round up their reported net income, we next investigate whether the earnings management is only earnings cosmetics or of greater magnitude. We report a summary of the results for our indirect analyses based on earnings and auditor characteristics for the firm–years with a positive net income in Table 5. The full results are reported in the Online Appendix.

For the first tercile of discretionary accruals (i.e., income-decreasing accruals), we find fewer than expected nines. However, there are also fewer than expected zeros (although the difference is nonsignificant). Given this pattern, whether the evidence is attributable to rounding up remains unclear. For the second tercile, we also do not find evidence of rounding up. For the third tercile (i.e., income-increasing accruals), we find significantly more zeros and ones than the Benford distribution predicts. However, there are not significantly fewer nines than expected, and, thus, there is only weak evidence of rounding up.

For the smoothing metrics *SMOOTH1* and *SMOOTH2*, we find significantly fewer than expected nines in the groups with less smooth earnings. For *SMOOTH2*, we also find significantly more zeros than expected. For both smoothing measures, we find no evidence of rounding up in the firms with smoother earnings. Thus, large-scale earnings management is more likely.

For our persistence proxy, we find fewer nines than predicted in the firms with less persistent net income, whereas the difference for the zeros is nonsignificant. There is no evidence of rounding up for those firms with more persistent net income. There is thus no clear evidence of differences in the rounding up in the different subsamples based on the persistence metric.

We find the typical pattern of rounding up for those firms with lower levels of conservatism. The numeral zero (nine) occurs marginally significantly more (less) often than expected. For those firms with higher levels of conservatism, we find no significant deviations for the numeral zero from the Benford distribution. We interpret this result as weak evidence of the differences between the subsamples, which makes large-scale earnings management likely.

We find the typical pattern of rounding up for those firms with a non-Big 4 auditor, in which significantly fewer than expected nines and more than expected zeros occur. By contrast, for firms with a Big 4 auditor, we find no significant deviation from the Benford distribution. Tests based on the auditors' industry specialization lead to similar conclusions. Thus, the tests based on auditor characteristics also indicate that large-scale earnings management is the most likely explanation for rounding up.

To further support our interpretation, we investigate the number of firms that are close to the next round number. In the first tercile of discretionary accruals, 2.7% of the observations would have to increase their reported net income by 1% to reach the next round number. In the second and third terciles, the values are 3.0% and 2.9%, respectively. That is, although the number of zeros is higher than expected only in the third tercile, there is approximately the same percentage of firms directly below the threshold in all terciles.[13] Thus, the excess zeros reported in the third tercile result from rounding up from a starting point that is far from the threshold. We find no difference between the subsamples in the number of firms directly below a round number for the other earnings characteristics either.

The results from the subsample analyses are also economically significant. For example, there is a 13% chance that a firm with a zero as the second digit and high levels of positive discretionary accruals has rounded up its net income. There is a similar likelihood for firms with less smooth earnings or a non-Big 4 auditor. If a firm reports a zero as the second digit, we encourage addressees to use additional analyses (e.g., the empirical tools of Amiram et al., 2015, and Henselmann, Ditter, & Scherr, 2015) to avoid being fooled by a rounded number.

If we repeat our analyses for firm–years with a negative net income, we find no evidence that is consistent with rounding up in any of the subgroups built on our earnings and auditor characteristics (we report the results in the Online Appendix).

5 | THIRD-DIGIT ANALYSIS

We stressed the fact that a direct test of the association between our earnings characteristics and rounding up is not possible. To rule out the possibility that our earnings characteristics are poor indicators of the significance of rounding up, we repeat our tests with the third digit. Suppose a net income of €6,480,000. It is probably not an option for management to round this number up to €7,000,000; however, management could still round up to €6,500,000. Since this kind of rounding is less likely to require substantial amounts of earnings management, we expect rounding in the third

[13] The results are qualitatively unchanged if we take the firms that would have to increase their net income by 5%.

digit to occur independently of the values obtained for our earnings characteristics. By contrast, if the concentration of firms rounding the second digit in specific subsamples is for reasons other than earnings management, we expect a similar pattern as in the main analysis.

For the full sample, we find strong support for rounding up. The numeral zero occurs more often than expected, whereas firms report every other numeral less often than expected (the difference is significant for the numerals two, four, seven, and, most importantly, nine). The chi-squared test is also highly significant. Thus, firms prefer the numeral zero not only for the second but also for the third digit (we report the results in the Online Appendix). For the analyses conditional on earnings and auditor characteristics, we find an excess of the numeral zero in every subsample, irrespective of earnings or auditor characteristics or their level. The chi-squared tests also reject the Benford distribution in every subsample (except for the first tercile of discretionary accruals). The result that our accounting quality metrics are connected to rounding up the second digit of net income but not the third supports our interpretation.

6 | CONCLUSION

We investigate German firms' rounding up of reported performance measures that target addressees' cognitive reference points. The most prominent reference points are multiples of 10, where rounding up results in a lack of nines as the second digit and an excess of zeros. We use Benford's Law to confirm rounding up in German group accounts among firms with both a positive net income and positive EPS.

Further analyses show that rounding up net income and EPS are separate issues, since most firms only round up one of these performance measures. Our tests indicate that earnings management is the most likely explanation for the observed deviations from Benford's Law. Additionally, we are interested in whether the earnings management used to beat given benchmarks is only cosmetic or greater. Our analyses show that rounding up net income is more pronounced for firms with positive discretionary accruals, less smooth earnings, less conservative earnings, non-Big 4 auditors, and non-specialist auditors. The pattern of more than the expected number of zeros and fewer than the expected number of nines holds in one specific subgroup throughout these different proxies, but not for the respective complementary subgroup(s). Specifically, only firms with earnings characteristics at certain levels round up net income, whereas the other firms cannot or choose not to. Because we expect all firms to have some incentives to round up, it seems unlikely that they forgo small-scale earnings management and thus miss a critical benchmark. The most likely explanation for our findings is, therefore, that large-scale earnings management conflicts with specific levels of earnings characteristics. In that case, rounding up can misdirect investors' and creditors' decision making. The term *cosmetic earnings management*, a frequent label for rounding up (e.g., Kinnunen & Koskela, 2003; Niskanen & Keloharju, 2000), can thus be misleading. Addressees and auditors should be aware of this risk and should apply additional analyses to separate firms with rounded earnings from those with unmanaged earnings.

Our design does not allow for the determination of the exact magnitude of rounding up. Thus, our analyses present only associations from indirect tests. The relations between earnings characteristics, auditor characteristics, and rounding up can actually not derive from substantial earnings management. For example, firms with low levels of discretionary accruals, high levels of smoothing, conservative accounting, Big 4 auditors, and industry specialist auditors could *choose* not to round up their earnings, even though they would only need cosmetic earnings management to meet the earnings target. In contrast, firms that round up their net income *happen* to be firms with high discretionary accruals and low degrees of smoothing, even though nonmaterial amounts of discretionary accruals are used for rounding up. This explanation holds if, in contrast to our assumption, the incentives for rounding up coincide with preferences for specific levels of earnings characteristics. However, our proxies represent different dimensions of earnings quality. Moreover, the sign of the associations between rounding up and earnings maximization is the opposite of that between rounding up and smoothing. Thus, the argument of pure coincidence is arguably unlikely.

ACKNOWLEDGEMENTS

We thank Peter F. Pope (the Editor), an anonymous reviewer, Susumu Shikano, Jan Riepe, and seminar participants at the University of Tübingen and the 2015 Annual Congress of the European Accounting Association for valuable comments and insights.

Open access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data used in this study are available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest to declare.

ORCID

Ulf Mohrmann https://orcid.org/0000-0002-0388-7431

REFERENCES

Amiram, D., Bozanic, Z., & Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, *20*(4), 1540–1593.

Audousset-Coulier, S., Jeny, A., & Jiang, L. (2016). The validity of auditor industry specialization measures. *Auditing: A Journal of Practice & Theory*, *35*(1), 139–161.

Barth, M. E., Beaver, W. H., & Landsman, W. R. (1998). Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics*, *25*(1), 1–34.

Barth, M. E., Landsman, W. R., & Lang, M. H. (2008). International accounting standards and accounting quality. *Journal of Accounting Research*, *46*(3), 467–498.

Becker, C. L., DeFond, M. L., Jiambalvo, J., & Subramanyam, K. R. (1998). The effect of audit quality on earnings management. *Contemporary Accounting Research*, *15*(1), 1–24.

Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, *78*(4), 551–572.

Bhattacharya, U., Holden, C. W., & Jacobsen, S. (2012). Penny wise, dollar foolish: Buy-sell imbalances on and around round numbers. *Management Science*, *58*(2), 413–431.

Bizer, G. Y., & Schindler, R. M. (2005). Direct evidence of ending-digit drop-off in price information processing. *Psychology & Marketing*, *22*(10), 771–783.

Bowen, R. M., DuCharme, L., & Shores, D. (1995). Stakeholders' implicit claims and accounting method choice. *Journal of Accounting and Economics*, *20*(3), 255–295.

Brenner, G. A., & Brenner, R. (1982). Memory and markets, or why are you paying $2.99 for a widget? *Journal of Business*, *55*(1), 147–158.

Burgstahler, D., & Chuk, E. (2017). What have we learned about earnings management? Integrating discontinuity evidence. *Contemporary Accounting Research*, *34*(2), 726–749.

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

Burgstahler, D., & Eames, M. J. (2003). Earnings management to avoid losses and earnings decreases: Are analysts fooled? *Contemporary Accounting Research*, *20*(2), 253–294.

Câmara, A. (2001). The pricing of relative performance based incentives for executive compensation. *Journal of Business Finance & Accounting*, *28*(9-10), 1149–1188.

Carslaw, C. A. P. N. (1988). Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review*, *63*(2), 321–327.

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

Cleary, R., & Thibodeau, J. C. (2005). Applying digital analysis using benford's law to detect fraud: The dangers of type i errors. *Auditing: A Journal of Practice and Theory*, *24*(1), 77–81.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, *34*(2), 669–700.

Cornell, B., & Shapiro, A. C. (1987). Corporate stakeholders and corporate finance. *Financial Management*, *16*(1), 5–14.

Das, S., & Zhang, H. (2003). Rounding-up in reported eps, behavioral thresholds, and earnings management. *Journal of Accounting and Economics*, *35*(1), 31–50.

Dechow, P., Ge, W. L., & 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.

Dechow, P., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *The Accounting Review*, *70*(2), 193–225.

DeFond, M., & Zhang, J. (2014). A review of archival auditing research. *Journal of Accounting and Economics*, *58*(2-3), 275–326.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *Journal of Business*, *72*(1), 1–33.

Demerjian, P. R. (2011). Accounting standards and debt covenants: Has the "balance sheet approach" led to a decline in the use of balance sheet covenants? *Journal of Accounting and Economics*, *52*(2-3), 178–202.

Dichev, I., & Skinner, D. J. (2002). Large-sample evidence on the debt covenant hypothes. *Journal of Accounting Research*, *40*(4), 1091–1123.

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, *5*, 17–34.

Eames, M. J., & Kim, Y. (2012). Analyst vs market forecasts of earnings management to avoid small losses. *Journal of Business Finance & Accounting*, *39*(5), 649–674.

Francis, J. R., Maydew, E. L., & Sparks, H. C. (1999). The role of big 6 auditors in the credible reporting of accruals. *Auditing: A Journal of Practice and Theory*, *18*(2), 17–34.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, *40*(1-3), 3–73.

Guan, L., He, D., & Yang, D. (2006). Auditing, integral approach to quarterly earnings, and cosmetic earnings management. *Managerial Auditing Journal*, *21*(6), 569–581.

Guan, L., He, S. D., & McEldowney, J. (2008a). Window dressing in reported earnings. *Commercial Lending Review*, *23*(3), 28–33.

Guan, L., Lin, F., & Fang, W. (2008b). Goal-oriented earnings management: Evidence from taiwanese firms. *Emerging Markets Finance and Trade*, *44*(4), 19–32.

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, *20*(2), 125–153.

Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics*, *7*(1-3), 85–107.

Healy, P. M., & Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, *13*(4), 365–383.

Henselmann, K., Ditter, D., & Scherr, E. (2015). Irregularities in accounting numbers and earnings management - a novel approach based on sec xbrl filings. *Journal of Emerging Technologies in Accounting*, *12*(1), 117–151.

Herrmann, D., & Thomas, W. B. (2005). Rounding of analyst forecasts. *The Accounting Review*, *80*(3), 805–823.

Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science*, *10*(4), 354–363.

Hill, T. P. (1998). The first digit phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data. *American Scientist*, *86*(4), 358–363.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance*, *56*(4), 1533–1597.

Holthausen, R. W., Larcker, D. F., & Sloan, R. G. (1995). Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics*, *19*(1), 29–74.

Indjejikian, R. J., & Nanda, D. (2002). Executive target bonuses and what they imply about performance standards. *The Accounting Review*, *77*(4), 793–819.

Jordan, C., & Clark, S. (2011). Detecting cosmetic eamings management using benford's law. *CPA Journal*, *81*(2), 32–37.

Khan, M., & Watts, R. L. (2009). Estimation and empirical properties of a firm-year measure of accounting conservatism. *Journal of Accounting and Economics*, *48*(2-3), 132–150.

Kinnunen, J., & Koskela, M. (2003). Who is miss world in cosmetic earnings management? A cross-national comparison of small upward rounding of net income numbers among eighteen countries. *Journal of International Accounting Research*, *2*(1), 39–68.

Kirschenheiter, M., & Melumad, N. D. (2002). Can "big bath" and earnings smoothing co-exist as equilibrium financial reporting strategies? *Journal of Accounting Research*, *40*(3), 761–796.

Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, *102*(5), 2206–2236.

Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: An international comparison. *Journal of Financial Economics*, *69*(3), 505–527.

Murphy, K. J. (2001). Performance standards in incentive contracts. *Journal of Accounting and Economics*, *30*(3), 245–278.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, *4*(1), 39–40.

Nigrini, M. J., & Mittermaier, L. J. (1997). The use of benford's law as an aid in analytical procedures. *Auditing: A Journal of Practice and Theory*, *16*(2), 52–67.

Niskanen, J., & Keloharju, M. (2000). Earnings cosmetics in a tax-driven accounting environment: Evidence from finnish public firms. *European Accounting Review*, *9*(3), 443–452.

Perotti, P., & Wagenhofer, A. (2014). Earnings quality measures and excess returns. *Journal of Business Finance & Accounting*, *41*(5-6), 545–571.

Quick, R., & Wolz, M. (2003). Benford's law in deutschen rechnungslegungsdaten. *Betriebswirtschaftliche Forschung und Praxis*, *55*(2), 208–224.
Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*(4), 532–547.
Shikano, S., & Mack, V. (2011). When does the second-digit benford's law-test signal an election fraud? Facts or misleading test results. *Jahrbücher für Nationalökonomie und Statistik*, *231*(5-6), 719–732.
Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.
Skousen, C. J., Guan, L., & Wetzel, T. S. (2004). Anomalies and unusual patterns in reported earnings: Japanese managers round earnings. *Journal of International Financial Management & Accounting*, *15*(3), 212–234.
Thomas, J. K. (1989). Unusual patterns in reported earnings. *The Accounting Review*, *64*(4), 773–787.
Thomas, M., & Morwitz, V. (2005). Penny wise and pound foolish: The left-digit effect in price cognition. *Journal of Consumer Research*, *32*(1), 54–64.
Ullmann, R., & Watrin, C. (2017). Detecting target-driven earnings management based on the distribution of digits. *Journal of Business Finance & Accounting*, *44*(1-2), 63–93.
Van Caneghem, T. (2002). Earnings management induced by cognitive reference points. *British Accounting Review*, *34*(2), 167–178.
Van Caneghem, T. (2004). The impact of audit quality on earnings rounding-up behaviour: Some uk evidence. *European Accounting Review*, *13*(4), 771–786.
Watts, R. L. (2003). Conservatism in accounting part i: Explanations and implications. *Accounting Horizons*, *17*(3), 207–221.
Xu, Y. (2016). Accruals management to avoid losses. *Journal of Business Finance & Accounting*, *43*(9), 1095–1120.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lebert S, Mohrmann U, Stefani U. Rounding up performance measures in German firms: Earnings cosmetics or earnings management on a larger scale? *J Bus Fin Acc.* 2021;48:564–586.
https://doi.org/10.1111/jbfa.12494