Self-Reported Earnings Management Practices: Experimental Evidence from Private Firms

Jannis Bischof
University of Mannheim

Yuhan Liu
University of Mannheim

Davud Rostam-Afschar
University of Mannheim

Running Head: Self-Reported Earnings Management Practices: Experimental Evidence from Private Firms

Acknowledgments: We appreciate helpful comments and suggestions from Sara Bormann, Nicole Cade, Luzi Hail, Katharina Hombach, Reeyarn Li (discussant), David Oesch, Sönke Sievers, Lauren Vollon (discussant), and workshop participants at University of Mannheim, Goethe University Frankfurt, 2024 TRR Annual Conference, EAA Annual Congress 2025, ESA World Meeting 2025, and AAA Annual Meeting 2025.

Affiliation and contact details: Jannis Bischof, University of Mannheim, University of Mannheim, 68131 Mannheim, E-mail: jbischof@uni-mannheim.de; Yuhan Liu, University of Mannheim, University of Mannheim, 68131 Mannheim, E-mail: yuhan.liu@uni-mannheim.de; Davud Rostam-Afschar, University of Mannheim, University of Mannheim, 68131 Mannheim, E-mail: rostam-afschar@uni-mannheim.de

Competing interests: The authors have no relevant interests to disclose.

Compliance with ethical standards: This experiment was reviewed by the Ethics Committee of the University of Mannheim (EC-3/2023) and registered in the AEA RCT Registry (AEARCTR-0009794).

Data Availability: For the empirical analyses, we used the German Business Panel. The German Business Panel is a recurrent survey of firm decision makers in Germany. It is run by some of the coauthors of this submission (Jannis Bischof, Yuhan Liu, and Davud Rostam-Afschar) and other researchers at the University of Mannheim. The German Business Panel survey data are subject to GDPR. The regulations prevent us from making the data accessible online. However, we can provide onsite or controlled remote data access to the original data for an independent third-party replicator after signing a Data Use Agreement. We will provide all program codes that generated all results for online access without any restrictions.

Funding: Financial support from the German Research Foundation (DFG) for the project 'Accounting for Transparency' (Grant No.: SFB/TRR 403041268) is gratefully acknowledged.

JEL Classifications: C83, D91, G10, M41, N24

Keywords: Earnings Management; List Experiment; Survey Methods; Social Desirability Bias

Self-Reported Earnings Management Practices: Experimental Evidence from Private Firms

ABSTRACT

We examine earnings management (EM) practices in private firms. Unlike public firms, where EM is shaped by capital market pressures, private firms primarily use earnings for private communication with stakeholders. While prior research documents incentives for EM in private firms, much less is known about whether managers view EM as socially undesirable and whether this varies by motives. In a large-scale online survey experiment, we randomly assign managers to a direct question, an indirect question, or a list experiment to elicit managers' report of EM. About 20 percent report accrual-based EM, with similar rates under direct questioning and the list experiment, suggesting that EM is viewed as a routine business practice. However, managers are less likely to disclose opportunistic motives under direct questioning, such as securing better credit terms or influencing business partners. When responses are masked in the list experiment, disclosure of these motives increases. No such difference emerges for managers who report a tax savings motive. The results collectively suggests that tax-motivated EM is perceived as more socially acceptable than EM aimed at influencing lenders or other business partners. Overall, admitting EM is not particularly sensitive among private firms, but undesirable motives can lead to under-reporting of such practice.

I. INTRODUCTION

In contrast to public firms, where financial reporting is shaped by capital market pressure, earnings of private firms primarily serve contracting, tax reporting, and information exchange with stakeholders such as lenders and business partners (Ball and Shivakumar 2005; Burgstahler, Hail, and Leuz 2006; Minnis and Shroff 2017; Lisowsky and Minnis 2020). Private firms are predominant in economies. In the European Union, the overwhelming majority of firms

(99.87%) are privately held, accounting for 42.8% of aggregated corporate assets and 61.8% of total employment (Beuselinck, Elfers, Gassen, and Pierk 2023).[1] Despite their economic importance, research on earnings in private firms remains limited relative to the extensive literature on public firms, largely due to weaker disclosure requirements and restricted access to financial statement data (Sutherland 2025). Even where financial statement data are available, as in many European countries where private firms are subject to reporting requirements, fundamental differences in reporting incentives and stakeholder scrutiny imply that evidence from public firms does not automatically generalize to private firms (e.g. Graham, Harvey, and Rajgopal 2005; Dichev, Graham, Harvey, and Rajgopal 2013; Lisowsky and Minnis 2020). Existing studies primarily examine reporting quality and differences in accounting choices between public and private firms (Hope, Thomas, and Vyas 2013; Lisowsky and Minnis 2020). While this literature documents that private firms manage earnings in response to contractual and tax incentives, evidence on earnings management (EM) in private firms is largely inferred from observed reporting outcomes and institutional settings (e.g. Ball and Shivakumar 2005; Burgstahler et al. 2006). As a result, it remains unclear how managers perceive EM and its underlying motives. Understanding managers' perceptions is important because outcome-based evidence alone makes it difficult to interpret observed behavior without insight into managers' underlying views. We address this gap by providing experimental evidence on private firm managers' perceptions of EM and its motives.

Accrual-based EM is inherently difficult for outsiders to detect because the underlying reporting process is shaped by the interaction of a firm's economic fundamentals, accounting standards, and subtle accounting choices by managers. Therefore, the identification of EM behavior using archival accounting data faces well-known limitations, including measurement error in accrual-based proxies and challenges in separating managerial discretion from economic fundamentals (P. Dechow, Ge, and Schrand 2010; Gerakos 2012; Ball 2013; Leuz and Wysocki 2016). Importantly, these identification challenges are debated even for public firms with rich disclosure environments and are likely to be even more questionable for private firms, where

1. In the United States, 33.18 million firms (99.9%) have fewer than 500 employees and collectively employ 61.7 million workers, accounting for 46.4% of private-sector employment (U.S. Small Business Administration, Office of Advocacy 2023).

information asymmetry is more likely to be resolved by an "insider access" model and public financial statements play a more limited role in stakeholder decision-making (Ball and Shivakumar 2005; Burgstahler et al. 2006; Beuselinck et al. 2023). Survey-based studies help address these limitations by directly engaging with decision-makers, thus avoiding strong assumptions about managers' use of reporting discretion. However, survey responses can be biased if managers view their engagement in EM practices as socially undesirable. Prior literature acknowledges this concern and attempts to mitigate social desirability bias through indirect elicitation techniques, such as asking about practices in other firms (the "other-people" approach, see Dichev et al. (2013)) or using list experiments which is particularly effective for eliciting honest responses on sensitive topics (Frye, Gehlbach, Marquardt, and Reuter 2017; Tourangeau and Yan 2007; Cade, Gunn, and Vandenberg 2024). In contrast to indirect questions that shift the focus to other firms, list experiment has the advantage that it retains questions about respondents' own behavior while masking individual responses through aggregation, thereby reducing incentives to misreport sensitive practices.

We combine the "other-people" approach and the list experiment method that has been used in prior literature by conducting a large-scale online survey experiment among firms in Germany. In our experiment, firm participants are randomly assigned to one of three groups: a List Experiment (LIST) group, a DIRECT group, or an INDIRECT group. Following Dichev et al. (2013), we focus on accrual-based EM, which we define as the practice of taking advantage of accounting choices and discretion for a firm's own benefit to misrepresent earnings. In the first part of our analysis, we estimate the prevalence of earnings management by asking firm participants directly or indirectly whether they have engaged in such practices. To address potential response bias, we adopt a list experiment design similar to that of Cade et al. (2024). In a list experiment, participants are randomly assigned to either a treatment or a control group. Both groups receive a list of non-sensitive statements and are asked to report the number of items that apply to them, without identifying which ones. The only difference is that the treatment group receives an additional item describing the sensitive behavior. In our experiment, this statement reads: "We have taken advantage of available accounting choices and managers' discretion for own benefit to misrepresent earnings." Assuming random assignment yields comparable views on the non-sensitive items, the difference in the average number of statements selected between the treatment and control groups reflects the estimated prevalence of EM. We compare this estimate to those derived from the DIRECT group and the INDIRECT group. Participants in the DIRECT group receive a straightforward question about EM within their own firm: "In past years, has your company taken advantage of accounting choices and discretion for own benefit to misrepresent earnings?" In contrast, participants in the INDIRECT group are asked about EM practices in the peer companies within the same industry using two approaches. First, similar to the DIRECT group, the indirect question is formulated in a "Yes/No" format: "In past years, have companies in your industry taken advantage of accounting choices and discretion for own benefit to misrepresent earnings?" Second, similar to Dichev et al. (2013), we include a second indirect question: "What percentage of companies in your industry (in %) haven taken advantage of accounting choices and discretion for their own benefit to misrepresent earnings?"

In the second part of our analysis, we investigate managers' stated motivations for engaging in EM. All participants, independent of the assigned experimental group, are asked the same indirect question: "Which of the following reasons are most relevant for such practices in your industry?" This design allows us to collect responses also from firms that do not engage in EM themselves. Based on prior literature, we compile a list of motivations, including: *i) Better lending opportunities. ii) Higher compensation for employees (e.g., higher bonuses). iii) Give customers/suppliers more certainty about the stable development of business. iv) Give the owners or investors more security about the stable development of business. v) To save taxes.* Next, we classify participants in the DIRECT and LIST groups into two categories based on their likelihood of engaging in EM. For the DIRECT group, this is based on their direct self-report in the survey. For the List group, because individual behavior is not directly observable, we estimate each respondent's probability of managing earnings using the one-step maximum likelihood estimator proposed by Imai, Park, and Greene (2015), based on firm characteristics. We then examine whether managers of EM firms in the LIST group report different motivations compared to their counterparts in the DIRECT group. Taking the responses from the non-EM firms as the baseline for industry-wide perceptions of EM motivations, this allows us to test whether greater privacy offered by the list experiment results in more candid disclosure of socially sensitive motivation. By assessing whether managers in the DIRECT group strategically select which motivations to disclose, we offer insights into whether managers consider certain motivations to be socially undesirable and consequently under-report them in a survey experiment when being asked directly.

Previous research offers mixed implications regarding how managers might respond in such settings. On the one hand, discretionary accounting decisions are legal and often used to meet earnings targets, which is an important signal of credibility in debt markets (Francis, LaFond, Olsson, and Schipper 2005; Graham, Li, and Qiu 2008). In addition, Burgstahler et al. (2006) argues that private firms face less pressure to manage earnings for tax minimization purposes, given the relatively lower importance of external financing reporting compared to public firms. On the other hand, managers may hesitate to disclose EM due to reputational concerns. For example, Hewitt, Hodge, and Pratt (2020) find that trust is impaired when EM suggests that managers are prioritizing personal interests over those of stakeholders. Similarly, Tahir, Ibrahim, and Nurullah (2019) and Cohen and Zarowin (2008) show that incentive structures, such as performance-based compensation, may motivate managers to conceal EM for personal gain. Building on these insights, we hypothesize that managers' willingness to truthfully report EM motivations in a survey setting depends on the social desirability of those motivations. Specifically, we expect that socially undesirable motives (e.g., personal gain) are more likely to be under-reported than widely accepted ones (e.g., tax savings).

We obtain the following results. First, regarding the prevalence of earnings management, we find that approximately 20 percent of firms in our sample report having engaged in EM in recent years. The INDIRECT group reports a much higher prevalence estimate of approximately 50 percent, which significantly exceeds the estimates from both the Direct and LIST groups. Interestingly, there is no significant difference between the prevalence estimates from the DIRECT and LIST groups. This suggests that, on average, admitting to EM does not trigger social desirability bias, and participants in the DIRECT group appear to respond truthfully. Despite having private firms in the sample, this result is consistent with Cade et al. (2024), who find a similar pattern among public firms in the U.S.

We further examine how firm managers report their motivations for EM and compare responses from the Direct and LIST groups. We focus specifically on managers in firms identified as engaging in EM, using responses from non-EM firms as the baseline. Surprisingly, we find that EM firms in the Direct and the LIST group report different sets of perceived motivations, but non-EM firms do not. After controlling for non-EM responses, the difference is particularly evident for motivations such as improving credit terms, enhancing perceptions by suppliers or customers, increasing employee compensation, and improving perceptions by owners or investors. These differences are also observed in an additional analysis where we constrain our sample to firms with disclosure obligation.

We interpret these differences as evidence that managers' reported motivations serve as a proxy for their true underlying motivations. In the DIRECT group, where EM behavior is disclosed directly, managers appear to strategically avoid selecting motivations they perceive as socially undesirable. By aligning their responses with those of non-EM firms, they may aim to avoid reputational risks associated with certain motives. In contrast, managers in the LIST group benefit from the privacy provided by the list experiment design. Because their EM behavior is not directly revealed, they face less pressure to conform to socially desirable norms and thus appear more willing to report potentially sensitive motivations. Interestingly, for the most commonly cited motivation, tax savings, we observe no difference in reporting across groups. This suggests that managing earnings for tax purposes is widely accepted and not perceived as sensitive or undesirable by firm managers. Overall, our results indicate that while acknowledging EM behavior itself does not appear to be sensitive, managers do engage in selective reporting when it comes to their motivations. The list experiment method mitigates this bias by enhancing privacy, thereby encouraging more candid responses.

Our paper makes several contributions to the accounting literature. First, we extend prior research on EM in private firms by providing direct experimental evidence on managers' reporting of EM practice and its motives. Prior work emphasizes differences in reporting quality, disclosure regulation, and contracting environments between public and private firms, but largely relies on inferred incentives and observed reporting outcomes (Ball and Shivakumar 2005; Hope et al. 2013; Leuz and Wysocki 2016; Minnis and Shroff 2017; Lisowsky and Minnis 2020; Sutherland 2025). Using data from Germany where accounting requirements depend on legal form rather than listing status, we provide evidence on how private firm managers perceive and justify EM in environments characterized by limited public scrutiny.

Second, we extend Cade et al. (2024) by examining private firm managers' perception of EM practice and motives. While Cade et al. (2024) document that EM itself is not uniformly perceived as socially undesirable among US public firms, we show similar results among Germany private firms, and show additionally that this assessment differs across motives. Comparing direct questioning with a list experiment, we find that managers are less willing to disclose opportunistic motives, such as influencing lenders or business partners, than tax-related motives, even when they report engaging in EM. This extension underscores the importance of distinguishing between the EM practice and the motives underlying it. This evidence also informs the interpretation of archival findings by highlighting that observed reporting patterns may reflect not only economic incentives but also differences in the perceived legitimacy of motives, which are largely unobservable in archival data.

Finally, we make a methodological contribution by introducing the list experiment to the study of EM in private firms. Building on Cade et al. (2024) and following the two-step approach of Imai et al. (2015), we recover firm-level probabilities of engaging in EM from list experiment responses and relate these probabilities to managers' reported motivations across survey designs. Comparing responses from the list experiment with direct questioning, we show that the additional privacy protection afforded by the list experiment increases managers' willingness to report EM behavior.

II. RELATED LITERATURE AND CONCEPTUAL UNDERPINNINGS

The literature identifies a wide range of common incentives EM. For instance, Franz, HassabElnaby, and Lobo (2014) find that firms facing tighter debt covenant constraints are significantly more likely to engage in EM. In a theoretical framework, Dutta and Fan (2014) examine how managerial compensation contracts influence the likelihood of earnings manipulation. Empirically, Kim, Kyung, and Ng (2022) document a positive association between the dispersion in pay-performance sensitivity and reduced earnings quality. In product market settings, Kim and Luo (2022) find that the Sarbanes-Oxley Act led firms with low customer concentration to reduce EM more than firms with high customer concentration. Other studies highlight the role of tax incentives in shaping EM behavior (Cazier, Rego, Tian, and Wilson 2015; Blaylock, Gaertner, and Shevlin 2015).

A common feature of these studies is their focus on identifying specific contexts in which incentives to manage earnings are particularly strong, linking these settings to observable EM behavior. Most of the studies rely on archival methods, which benefit from large-scale, time-series data that are publicly available and relatively easy to process. However, a key limitation lies in the measurement of EM. Specifically, EM proxies based on financial statement data struggle to capture managerial intent, making it difficult to separate discretionary from non-discretionary components of earnings (Ball 2013; Gerakos 2012). When EM is measured with substantial error, the correlation between EM activities and underlying motives is attenuated, limiting the ability to draw causal inferences.

To address this limitation, survey-based approaches offer a complementary perspective. Unlike archival studies, surveys can directly elicit managers' views and do not rely on assumptions about intent. However, self-reported data on sensitive topics like EM are also susceptible to measurement issues, including non-response, recall bias, and social desirability bias, which describes the tendency to under-report socially undesirable behaviors and over-report desirable ones (Graham et al. 2005). To mitigate these challenges, prior research has adopted various techniques. For example, Graham et al. (2005) used hypothetical scenarios in their survey of financial executives, finding that managers preferred real activities manipulation over accrual-based methods. Dichev et al. (2013) expanded on this approach by surveying 169 CFOs from public firms and 206 from private U.S. firms, employing an "other-people" approach to reduce social pressure and provide more candid insights into EM motivations. More recently, Cade et al. (2024) introduced a novel survey design incorporating list experiments to estimate the prevalence of five forms of EM. This method embeds sensitive items in randomized lists and encourages more truthful responses, because participants' answer to the sensitive item is masked in their aggregated responses. The authors report higher prevalence estimates for certain EM behaviors, 29.9 percent for real EM and 12.4 percent for accounting fraud, though the list experiment does not yield a higher estimate for accrual-based EM.

Building on these findings, our study investigates how the disclosure of EM motivations varies with actual engagement in EM. In contrast to prior research that focuses on prevalence estimates, we explore whether social desirability bias affects respondents' willingness to report different motives for EM. We further assess whether some motivations are perceived as more socially acceptable than others. Understanding the relationship between self-reported motivations and EM behavior is critical for both researchers and regulators. It has implications for the design of empirical tests for EM and for the effectiveness of regulatory interventions.

Prior literature suggests that not all EM motivations are perceived equally. Some may be viewed as legitimate or even aligned with shareholder interests—such as meeting earnings benchmarks to signal performance or gain access to capital markets (Francis et al. 2005; Graham et al. 2008), or managing earnings to influence valuation (Kasznik and McNichols 2002; Myers, Myers, and Skinner 2007). Johnson, Fleischman, Valentine, and Walker (2011) argue that managers may rationalize such decisions when they appear to serve the firm's long-term interests, even if they raise ethical concerns. Likewise, Dyreng, Hillegeist, and Penalva (2022) show that EM related to debt covenants can benefit shareholders, challenging the view that all EM reflects agency conflicts.

Nevertheless, reputational and ethical considerations may constrain managers' willingness to disclose EM behavior, particularly in the wake of major scandals and regulatory responses like the Sarbanes-Oxley Act (Cohen and Zarowin 2008). Jensen (2005) frames EM as a form of deception that can erode trust. Consistent with this view, Hewitt et al. (2020) find that shareholder trust deteriorates when EM appears driven by managerial self-interest. Compensation-based incentives can further amplify this concern, motivating managers to obscure EM to protect personal gains (Tahir et al. 2019; Cohen and Zarowin 2008). Ethical standards also shape perceptions: Fischer and Rosenzweig (1995) show that acceptability of EM varies with education and experience, while Commerford, Hermanson, Houston, and Peters (2019) find that auditors are less willing to retain clients engaged in real EM, highlighting reputational risks.

While most of this evidence comes from public firms, understanding EM in private firms is equally important. Private companies represent a significant share of economic activity in terms of employment, investment, and innovation (Gassen and Muhn 2018). These firms differ in several key respects. Public firms typically have dispersed ownership and communicate through standardized financial reports. In contrast, private firms often have concentrated ownership structures and rely more on informal communication channels (Burgstahler et al. 2006). They also face higher proprietary costs, which can deter voluntary disclosure (Bernard 2016). These structural differences influence both the frequency and the motivations for EM. For example, Burgstahler et al. (2006) find that private firms face less pressure to minimize taxable income, suggesting a different incentive landscape from that of public firms.

Building on these insights, we hypothesize that managers' willingness to disclose EM motivations truthfully depends on the social desirability of those motivations. Specifically, we expect socially undesirable motives (e.g., personal gain) to be under-reported, whereas more widely accepted motives (e.g., tax savings) are more likely to be disclosed.

III. EXPERIMENTAL DESIGN AND PROCEDURES

Accrual-Based Earnings Management

In this study, we explicitly exclude GAAP violations and focus on legal accrual-based earnings management (EM) for several reasons. First, illegal accounting practices such as GAAP violations are relatively rare, resulting in limited sample sizes and reduced statistical power. Findings based on such cases are therefore less generalizable to the broader, more subtle forms of EM that are more commonly encountered in practice (see, for example, P. M. Dechow, Sloan, and Sweeney (1996) and Beneish (1999)). In contrast, accrual-based EM is more prevalent and nuanced, making it a more suitable focus for examining variation in managers' reporting behavior. Second, accrual-based EM has a direct and immediate impact on reported earnings, which allows for more precise inferences about financial reporting quality. Third, standardized financial reporting requirements under recognized accounting frameworks enhance the comparability of earnings figures across firms and industries. In contrast, real EM that is achieved through actions such as altering operational decisions is more difficult to measure consistently, as it often depends on firm-specific contexts. Finally, we note that prior literature has found a trade-off between real EM and accrual-based EM (Zang 2012; Cohen and Zarowin 2008). In this study, we focus on accrual-based EM, while real EM remains a promising avenue for future research.

Following Dichev et al. (2013), we define accrual-based EM as the use of available accounting choices and managerial discretion to misrepresent earnings for own benefit[2]. The specific underlying motivations are covered in a follow-up question (Details in Section 3.2.5).

Experimental Design

Our experiment includes three experimental groups (LIST group, DIRECT group and INDIRECT group), each designed to elicit information about earnings management (EM) in a different way. After a short introduction and a first comprehension question on EM, firm participants are randomly assigned to one of these three groups. Next, all participants are asked about their perceived underlying motives for EM. The survey concludes with questions about firm and respondent characteristics. Details on the experimental design, including the exact survey questions in German and their English translations, are provided in the Online Appendix. Descriptive statistics on the sample are presented and discussed in Section 4.1.

Earnings Management Practices

At the beginning of the survey, all firm participants are asked a general question to assess their understanding of earnings management (EM): "Which of the following can be used to manage earnings in compliance with GAAP?" The responses are summarized in Figure 1, which shows the proportion of participants who classify various practices as permissible under Generally Accepted Accounting Principles (GAAP). The results indicate that 72.5 percent of respondents correctly identify the choice of depreciation method as a legal EM practice, the highest recognition rate among the listed practices. Similarly, 68.3 percent correctly classify impairments as GAAP-compliant, while 50.2 percent recognize the acceleration or deferral of revenues and expenses as a legal form of EM. Only 12.8 percent identify cost allocation as

2. We intentionally leave "own benefit" open to interpretation – encompassing motivations that serve either the firm's strategic interests or the manager's personal incentives.

legitimate, and 8.10 percent incorrectly consider backdating invoices to be GAAP-compliant[3]. Overall, these findings suggest that a substantial share of respondents can correctly distinguish legal EM techniques.

List Experiment Group

Similar to Cade et al. (2024), we design a double list experiment. This method is particularly effective for addressing sensitive topics and has been widely applied in social research (e.g., Frye et al. (2017) and Traunmüller, Kijewski, and Freitag (2019)). Panel A of Table 1 presents an example of a standard list experiment. Participants are randomly assigned to either a treatment or a control group. Both groups receive a list of items and are asked to report the total number of items that apply to them, without specifying which ones. The lists given to both groups contain the same control statements (1-4), but the treatment group receives an additional statement about the sensitive behavior. In our setting, that is the accrual-based EM. Since individuals only provide an aggregated answer regarding the total number applied for the list of items, her true response to the sensitive item is masked (Question: *How many of the following measures has your company taken in recent years?* Answer: *We have taken __ of measures.*). If the two groups have similar views on the non-sensitive items (which is often ensured by the random assignment of participants), the prevalence of EM in the treatment group can be estimated by comparing the average number of items reported between the two groups. In other words, the prevalence can only be estimated through group comparison, not on the individual level. Participant's privacy is thus preserved.

To gain more statistical power, we implement an extended version of a single list experiment, namely a double list experiment (DLE) (Glynn 2013). Panel B of Table 1 illustrates the double list experiment that we use in our study. Firm participants are randomly assigned to either List Experiment Group 1 or 2, and participants from both groups receive two different lists. While firm participants assigned to the List Experiment Group 1 serve as the treatment observations for List A and as control observations for List B, the opposite applies for participants in List Experiment Group 2. While both lists contain different control items, depending on

3. For the main analyses below, we restrict the sample to firm participants who do not classify "backdating invoices" as GAAP-compliant. The inference does not change when the whole sample is used.

the treatment status, they include the same sensitive statement regarding EM: "We have taken advantage of available accounting choices and managers' discretion for own benefit to misrepresent earnings." The final single estimate of the prevalence of the sensitive behavior is derived by pooling the responses from both lists. The order of the items in both lists is randomized to avoid order effects.

A successful list experiment needs to fulfill several criteria: (i) absence of ceiling and floor effects, (ii) absence of design effect, and (iii) random assignment of treatment status (Kuklinski, Cobb, and Gilens 1997; Blair and Imai 2012). First, the potential existence of ceiling/floor effects could distort the results from the list experiment. In our setting, the ceiling effect is of major concern. If a respondent approves all items in the treatment list and reports truthfully, this implicitly signals that the respondent has conducted EM as well. In this case, the list experiment loses its advantage of sensitivity protection, and the respondent may just indicate one item less to answer in a socially desirable way. If this phenomenon is systematic, this would lead to an underestimation of EM. To check whether the ceiling effect exists, we summarize the responses to the list experiment in Appendix B, Table C.1. For both list A and B, the proportion of respondents who approve all items in the control group is minimal (3.41 percent for list A and 4.67 percent for list B). For control groups, the mean reported items for list A and B are 1.41 and 1.54 respectively. The relatively small mean values rule out concerns about a potential ceiling effect.

Second, the design effect should be avoided. Specifically, the addition of the sensitive item to the baseline list should not alter the respondents' answers to the baseline list. Imagine that a respondent has selected 3 items out of 4 on the baseline list. If the sensitive item is added additionally to the list, based on whether the respondent has selected the sensitive one or not, her answer should either remain 3 or increase to 4. If she reports 2 after the inclusion of the sensitive item, this would be a sign of design effect. We run the test proposed by Blair and Imai (2012) and the insignificant results indicate that there is no obvious design effects (untabulated).

Third, the assignment of treatment status should be random. Successful randomization requires that participants in each condition are sufficiently similar so that on average they respond similarly to the same number of control statements in each list. To ensure this, we randomly assign participants to the experimental conditions in the survey software Qualtrics during the field phase. Ex-post, we assess whether randomization was effective by comparing participants' self-reported manager and firm characteristics across experimental groups. In support of successful randomization, the chi-square tests to test the distribution of categorical variables reported in Table 4.1 fail to reject the null hypothesis of no difference across List Experiment Groups 1 and 2 (untabulated).

When these assumptions are fulfilled, the prevalence of earnings management in the List Experiment Group can be derived from the following regression:

$$Y_i = \lambda + \tau T_i + \beta List_i + \varepsilon_i. \tag{1}$$

Y_i is a discrete variable that ranges from 0 to 5 for treatment lists and from 0 to 4 for control lists. T_i is a binary variable equal to one if the response comes from a treatment list, and zero otherwise. *List* is an indicator variable equal to one for list A, and zero for List B. This variable controls for the different innocuous items in the two treatment lists. Since each firm participant views two lists, we cluster standard errors at the individual level. λ indicates the average number of items answered affirmatively in the control group. The main parameter of interest is τ, which corresponds to the difference between the average number of items reported in the treatment list and the average number of items reported in the control list. This difference indicates the prevalence of the sensitive statement, which is accrual-based earnings management in our case. β controls for the underlying difference between the two lists and ε is an error term. For a more comprehensive review on the derivation of the list experiment prevalence estimates, please refer to Appendix B.

Direct Group

To collect data on a specific characteristic among individuals in a population, the most straightforward approach is to ask survey participants directly. Firm participants allocated to the DIRECT group receive a direct question that asks for EM activities directly within their own firm: "*In past years, has your company taken advantage of accounting choices and discretion for own benefit to misrepresent earnings?*" The prevalence of EM in the DIRECT group could be derived from the simple regression:

$$direct_i = \beta_0 + \varepsilon_i, \tag{2}$$

where β_0 is the sample mean of *direct*, which equals to one if the participant answers "Yes".

Participants in this group also receive one of the control lists from the double list experiments to ensure that both the DIRECT group and the LIST groups receive the same number of questions. However, due to potential respondent reluctance or intentional misreporting of sensitive information, direct surveys often face challenges related to reliability and accuracy (Elffers, Weigel, and Hessing 1987). To test whether this concern is relevant in the context of EM, we compare the estimates derived from the LIST and DIRECT groups. The results and relevant discussion are in Section 4.

Indirect Group

An alternative approach is the indirect questioning technique. For example, survey respondents may provide information about people they know, either in addition to or instead of disclosing personal details (Breza, Chandrasekhar, McCormick, and Pan 2020). A typical question is "How many X's do you know?", where X refers to persons in the respondents network. In a context closely related to ours, Dichev et al. (2013) employed this method in a US survey of CFOs to estimate the prevalence of EM[4].

One primary advantage of this method lies in its preservation of privacy and enhanced participation, as respondents are not required to disclose their own status. However, the effectiveness of this approach depends on several critical assumptions. First, it assumes that respondents can accurately recall the individuals within their personal networks. Second, respondents must know whether each person in their network belongs to the hidden population. Third, it assumes that all individuals have an equal probability of knowing someone within the hidden population (Laga, Bao, and Niu 2021). Although previous studies have sought to relax

4. This method has been applied across various contexts, such as estimating the population of female sex workers (Jing, Lu, Cui, Yu, and Wang 2018), key populations for HIV (Teo et al. 2019), and more recent studies on COVID-19 prevalence (Srivastava et al. 2023; Garcia-Agundez et al. 2021).

these assumptions, the use of poorly designed questionnaires can still lead to response bias, recall errors, transmission errors, and barrier effects when collecting aggregated relational data (Laga et al. 2021)[5].

For the INDIRECT group, we inquire about EM practices in the participating firm's peer companies within the same industry using two different approaches. First, similar to the DIRECT group, the indirect question is formulated in a "Yes/No" format: "*In past years, have companies in your industry taken advantage of accounting choices and discretion for own benefit to misrepresent earnings?*" Second, similar to Dichev et al. (2013), we include a second indirect question: "*What percentage of companies in your industry (in %) haven taken advantage of accounting choices and discretion for their own benefit to misrepresent earnings?*" Similar to the prevalence derived for the DIRECT group, the mean response is calculated to enable a comparison across experimental groups:

$$indirect_i = \beta_1 + \varepsilon_i, \tag{3}$$

where β_1 is the sample mean of *indirect*, which is the estimated industry prevalence reported by firm participants.

Earnings Management Motivations

A key distinction between our study and Cade et al. (2024) is that beside prevalence, we also examine how managers perceive the motivations underlying EM — whether they serve the firm's interests, personal gain, or other purposes. These motivations shape both the frequency and nature of EM, thereby influencing overall financial reporting quality. Drawing on prior literature, we compile a list of relevant incentives, including improved credit terms, higher employee compensation, enhanced perception by suppliers or customers, favorable views by owners or investors, and tax savings. Considering this range of factors provides a comprehensive perspective on what drives EM across different organizational contexts.

5. Previous research has identified instances where the technique may not perform optimally; for instance, Rigol, Hussam, and Roth (2017) demonstrate strategic misreporting in a peer elicitation method, exacerbated when respondents believe their reports influence grant distribution. Moreover, Phillips (1981) indicate the method yields unreliable results if informants are insufficiently informed about others' practices.

To investigate the relevant motivations, participants from all experimental groups receive a question not about their own firm's behavior but about common motivations for EM in their industry. Specifically, respondents are asked: "*There are several potential reasons why a company might take advantage of accounting choices and managerial discretion for its own benefit to misrepresent earnings. Which of the following reasons are the most relevant for such practices in your industry?*"

Our experimental design essentially follows a two-step structure. In the first step, we estimate the prevalence of EM, either based on a list experiment, a direct question, or an indirect question; in the second step, we explore EM motivations in an indirect way. Linking direct and indirect measures improves the interpretability of the results by incorporating different perspectives (Schedler and Sarsfield 2007). [6] The indirect framing on motivation helps mitigate sensitivity concerns by not requiring firms to disclose their own EM behavior. Moreover, it allows us to collect responses from both EM and non-EM firms, using the latter as a benchmark for industry-wide perceptions. By combining information on EM prevalence and perceived motivations, our design enables us to examine whether social desirability bias affects disclosure of motivations on top of the acknowledgment of EM itself. This distinction is important, as firms that engage in EM may view or report motivations differently based on their experience with the practice.

In a nutshell, we aim to draw implications on whether managers selectively under-report motivations that are considered as socially undesirable when being asked directly by comparing responses from the LIST group with that from the DIRECT group. In addition, we condition their stated perceived motivations on their EM practices. For participants in the DIRECT group, the stated EM practices are directly observable. Assuming linear relationships for illustration purposes, the propensity to select certain motivations could be modeled as

$$Motiv_i = \alpha + \gamma_1 EMD_i + \beta^T X_i + \varepsilon_i, \tag{4}$$

6. Theoretically, it is possible to design different list experiments tailored to investigate different motivations behind EM. However, from a practical perspective, these motivations are rarely unique, as firms often have multiple, overlapping reasons for engaging in such practices. Conducting multiple list experiments for each potential motive would be a burden for the participants, potentially leading to response fatigue and lower data quality.

where EMD_i is a dummy variable which equals to one if the respondent i in the DIRECT group states engaging in EM affirmatively. X_i is a vector of respondents' characteristics. γ_1 captures the differences in the perceived EM motivations between firms that state engaging EM and firms that do not.

For participants in the LIST group, by design, it is unclear whether the firm has managed earnings or not. Formally, the propensity to select certain motives can be modeled as

$$Motiv_i = \alpha + \gamma Z^*_{i,J+1} + \zeta Y^*_i + \beta^T X_i + \varepsilon_i, \tag{5}$$

where $Z^*_{i,J+1}$ denotes respondent i's latent truthful response to the sensitive item, Y^*_i is the latent number of control items answered affirmatively given the response to the sensitive item, and X_i is a vector of respondents' characteristics. The main parameter of interest is the coefficient for the latent response to EM, γ. However, we neither observe directly the answer to the sensitive item $Z^*_{i,J+1}$ nor the number of control items answered affirmatively given the response to the sensitive item Y^*_i. Thus, we use firm and respondent characteristics to predict these variables. This way, we can study the correlation between EM firms and the self-reporting of a certain motive. We adopt the methodology proposed by Imai et al. (2015) to estimate the outcome regression model. In Appendix B, we formally derive the prevalence estimator from the list experiment design and demonstrate how the outcome regression works.

IV. RESULTS

Sample Descriptives

Data are collected from the German Business Panel (GBP) between December 2022 and February 2024. The GBP is a large-scale online firm survey with the aim of enhancing the understanding of entrepreneurial behavior, change in the business landscape, and the impact of policy decisions on firms. For a comprehensive overview of the data and GBP, we refer to Bischof, Doerrenberg, Rostam-Afschar, Simons, and Voget (2025). For the main analyses, we restrict the sample to firm participants who are CEOs or owners and who do not classify "backdating invoices" as GAAP-compliant (N = 4,086). This restriction serves two purposes: (1) to exclude respondents who appear to misunderstand the legal boundaries of EM, and (2) to focus on individuals likely to possess adequate knowledge of EM practices within their firms. Importantly, our results remain robust when using the full sample, or when applying alternative classification-based exclusion criteria based on Figure 1.

Table 2 provides a summary of firm characteristics by number of employees, revenue, legal form, and disclosure obligation. Columns 1 and 2 report the distribution from the Statistical Business Register of the German Federal Statistical Office (for the 2022 reporting period) along with the firm characteristics of our sample. Our sample has a median number of employees of 4 and a median revenue of 550,000 €. It is evident that our sample firms are skewed to the right, as the mean number of employees and revenue are much higher than the median values, suggesting that we have much more small firms than large ones. Nevertheless, we observe that firms in our sample are larger than the population of the German universe. While 70.3 percent of the firms in our sample have fewer than 10 employees, according to the Statistical Register, 86.8 percent of the total population has fewer than 10 employees. Similarly, the majority of firms in the Statistical Register have revenues below 2 million (91.9 percent), while in the sample this percentage is much lower (55.9 percent), which suggests that the sample includes firms with higher revenues than those in the general population of firms. Furthermore, sole proprietorships are the most common legal form according to the Statistical Register (59.2 percent), but are less common in our sample (21.4 percent). Moreover, according to the Statistical Register, 23.7 percent of firms in the total population are corporations. In our sample, more than half of the firms are corporations. Overall, our sample appears to overrepresent mid-sized and larger firms, higher revenue firms, and corporations compared to the broader firm population in Germany. Columns 3 to 5 report firm characteristics by experimental groups. Column 6 reports the p-values from statistical tests assessing equality across three experimental groups. Since all $p \geq 0.1$, there appear to be no systematic differences across groups. This suggests that the randomized assignment of participants to the experimental groups worked as expected.

Table 3 reports respondent characteristics (education, position, and gender) in column 1. Columns 2 to 5 show the number of employees and revenue for each category. Columns 6 to 9 show the respondent characteristics by experimental groups. Overall, the statistics indicate that we have an experienced and knowledgeable sample of respondents. In column 1, we observe that all participants in our sample are owners or CEOs of the firm due to the restriction.[7] 63.3 percent of the respondents have a college degree, and it is a male-dominated sample (85.2 percent). Comparing column 2 to 3 and column 4 to 5, we again observe the skewness of the data - the mean values for number of employees or revenues are significantly higher than the median values. The *p*-values in the last column indicate that there are no differences across experimental groups.[8]

Prevalence of Earnings Management

First, we ask firm participants about their accrual-based EM practices and derive the prevalence across experimental groups. The results are presented in Table 4, from columns 1 to 4. Column 1 shows the OLS estimate from the double list experiment, indicating that 21.4 percent of our sample firms engage in EM.[9] The estimate from the DIRECT group is at 22.6 percent. The difference is insignificant, which suggests that, on average, firms assigned to the DIRECT group do not face disproportional social desirability bias and systematically under-report their EM practices. One might question whether our inference is affected by the sensitivity of EM practices, which could be insufficiently addressed by our list experiment design. If this were true, in the extreme case, no one would report the sensitive item, leading to same averages between the treatment and control groups, which further implies a prevalence of zero. This is not observed in our data. Overall, this result is consistent with Cade et al. (2024), in which the authors also find no significant under-reporting behavior for firms being asked directly about their accrual-based EM practices.

Interestingly, the perceived prevalence of EM among other firms (column 3 and 4) in the same industry is significantly higher than the estimates from the List Experiment or the DIRECT group. We formulate two different questions to get a better understanding of how firms

7. In our raw sample, 86.5 percent of the firm participants indicate that they are the owners or the CEOs.

8. The questions about firm and respondents characteristics are shown after the experiment. The insignificance of differences across groups also eliminates the concern that there is systematic priming effect incurred from the experimental design.

9. The List Experiment group has almost twice as many observations as the DIRECT group due to the double list design, which involves two separate non-sensitive item lists. Each participant is treated with one list while being controlled by the other.

perceive EM among their industry peers. First, similar to Dichev et al. (2013), we formulate the question as: "*In past years, what percentage of companies in your industry (in %) have taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?*" The average response is 45.7 percent (column 3), which is considerably higher than the 24.6 percent in Dichev et al. (2013) or 20.5 percent in Cade et al. (2024).[10] Column 5 to 7 report the estimates from prior published survey studies. We note that there are differences in sample composition and research methods across studies, which makes a structural comparison between the derived prevalence estimates interesting. While Cade et al. (2024) only consider listed U.S. firms and Dichev et al. (2013) focus on both public and private firms but all large firms, we have smaller firms in our sample. Compared to these studies, Graham et al. (2005) applies another method as an alternative to the indirect questioning, namely the hypothetical scenarios. Still, our question can be compared as it asks: "In past years, have companies in your industry have taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?" Column 4 shows that 64.3 percent of firms in our sample perceive that at least one of their industry peers have managed earnings in past years. We consider this as an upper bound estimate for EM, and a precise estimate is highly dependent on the number of industry peers of each of the survey participant or even overestimation due to double counting. Therefore, in the following analyses, if not stated explicitly, we only compare the indirect(share) group with the estimates from LIST or DIRECT groups.

EM Prevalence: Subgroup Analysis

We next examine EM prevalence in subgroups. Panel (a) shows the EM prevalence based on firm size by using the median revenue (55,000 euros) to separate firms into "smaller" and "larger than median revenue" groups. Among firms with less than median revenue, the prevalence of EM is 0.18 for both the Direct and LIST groups, with no significant difference between the two groups. Firms in the INDIRECT group perceive, on average, that 42 percent of their peers manage their earnings. For firms with revenues above the median, the prevalence of EM is higher, especially for the INDIRECT group (48 percent). The Direct and LIST groups report

10. Please refer to the Appendix A.1 for an overview of the prevalence of EM from prior survey literature.

prevalence rates of 0.26 and 0.24, respectively, which are lower but not statistically significant from each other. Collectively, these results suggest that the EM prevalence derived from the INDIRECT group depends on the size of the respondent's own firm, but when asked about EM practices at their own firms, there is no significant difference between being asked directly or in a list experiment, which is designed to preserve the privacy of the respondent.

Panel (b) examines the disclosure requirement as a subgroup variable. Among firms with the disclosure obligation to publish annual financial statements in the German Federal Gazette, the prevalence of EM is 0.17 and 0.23 for the LIST group and the DIRECT group, respectively. The point estimate of the DIRECT group is higher, but the estimates of the Direct and LIST groups are not significantly different. The indirect method shows a significantly higher prevalence of 0.50. Companies without disclosure requirements have slightly higher prevalence rates in the LIST group (0.24) than companies with disclosure obligation (0.21).

In Panel (c), we examine the local business tax rate at the firm's location. Firms are classified as *LBTR_high* if they are subject to a local business tax rate above the median of the sample. Within this subgroup, 25 percent of firms in the DIRECT group and 22 percent in the LIST group are identified as engaging in EM practices. Consistently, the prevalence of EM does not differ significantly between the two groups. Panel (d) focuses on firms' leverage levels, defined as the ratio of current liabilities to total assets. Leverage data are sourced from the Orbis database and matched to survey responses using BvDID. [11] While prevalence estimates are higher in the DIRECT group than in the LIST group, the wide confidence intervals in the LIST group prevent us from drawing firm conclusions.

In general, these results imply that there is no clear social desirability bias within subgroups. In other words, social desirability bias is independent of static demographic characteristics. In a multivariate test, we examine whether any executive or firm characteristics are useful predictors of firms' reports of EM. In short, we do not find consistent associations between any characteristic and the prevalence of EM. See Appendix C.2 for detailed analysis. The conclusion is consistent with Graham et al. (2005) and Cade et al. (2024), which also find that demographic variables generally do not affect their measures of misreporting behavior. We do

11. As not all firms publish financial statements, Orbis coverage is incomplete. This results in small sample with large confidence intervals.

find that the estimates from the INDIRECT group are highly depend on firm characteristics such as size or disclosure obligation. We interpret this as a result of the different benchmarks and pool of peer industries that firms may have in mind when responding.

Earnings Management Motivations

Descriptives

After asking about EM practices, we ask firm participants from all experimental groups about the potential motivations behind these practices: "*There are several potential reasons for why a firm might take advantage of accounting choices and managers' discretion for its own benefit to misrepresent earnings. Which of the following reasons are most relevant for such practices in your industry?*" The answer options are: *Better lending opportunities. Higher compensation for employees (e.g., higher bonuses). Give customers/suppliers more certainty about the stable development of business. Give the owners or investors more security about the stable development of business. To save taxes.* Please refer to the Online Appendix for original survey question in German.

Table 5 summarizes the EM motives selected by firm respondents across the full sample and by subgroups: experimental groups (List experiment or DIRECT group), firm size, disclosure obligation, local business tax rate, and leverage.[12]

The most frequently cited motivation across all respondents is tax savings, selected by 69.5 percent of firms. This motivation was reported by 69.9 percent of DIRECT group participants and 66.6 percent of the LIST group, with the difference being statistically significant at the 10 percent level. Smaller firms and firms with no disclosure obligation consider this motivation as more relevant. The prevalence of this motivation also appears consistent across leverage and local tax rate categories. The second most cited motivation, securing better credit terms, was selected by 37.5 percent of respondents. Large firms, firms with disclosure obligation, highly leveraged firms, and those facing higher local business tax rates are more likely to cite this motivation, suggesting greater reliance on external financing or stronger incentives to maintain

12. We exclude responses from the INDIRECT group, because we want to examine whether firms that were asked about their own EM behavior subsequently report different motivational patterns.

favorable lending conditions. "Give customers/suppliers more certainty about the stable development of business" was chosen by 35.4 percent of respondents. Interestingly, 37 percent of firms in the LIST group selected this motive, compared to just 30.4 percent in the DIRECT group. Similar to credit-related motives, this option was selected more frequently by large and highly leveraged firms, likely reflecting their need to preserve a positive external image due to their more prominent market roles. "Higher compensation for employees (e.g., higher bonuses)" was cited by 24.9 percent of participants, with 26.3 percent of the LIST group identifying this motivation versus 21.7 percent in the DIRECT group. This motive is also more frequently cited by large firms, possibly reflecting the greater use of performance-based compensation structures in such firms. Finally, "Give the owners or investors more security about the stable development of business" was the least frequently cited motive (20.2 percent), which is unsurprising given that the sample is largely composed of small, privately held firms. No significant differences were observed between large and small firms on this dimension, but between firms with and without disclosure obligation.

Overall, tax savings emerge as the dominant perceived motivation for EM across all subgroups. Financially driven motives such as improving credit conditions or customer perception are more commonly reported by larger firms or firms with disclosure obligation, underscoring their exposure to external stakeholder expectations. Notably, the higher selection rates for certain motives within the LIST group suggest that the list experiment may elicit more candid responses for sensitive motivations. In the following section, we further investigate whether participants in the List and DIRECT groups report different sets of EM motives, conditional on their own EM practices. This distinction is important because for firms that admit to engaging in EM, their selected motivations may more closely reflect the true underlying drivers of such behavior.

Relevance of Perceived Motivations by Experimental Groups

Our question on EM motivations is intentionally framed in an indirect way. Rather than directly asking firms about their own motivations, we ask about common motivations observed among peer firms in the same industry. This indirect approach allows firms across all experimental groups to respond in a comfortable manner. That is, even firms that do not engage in EM could also provide their perceptions of EM motivations in this question. However, it is important to note that firms that do (EM firms) and do not manage earnings (non-EM firms) may have different perception for motivations, as those which manage earnings in recent years have more experience with this accounting practice. To assess whether perceived motivations differ between the List and DIRECT groups among EM firms, we use the responses from non-EM firms as the baseline. Responses from the INDIRECT group are excluded from the analysis because it is impossible to predict whether a firm assigned to the INDIRECT group has managed their earnings based solely on the estimates of its industry peers.

For the DIRECT group, our approach is straightforward because we observe their responses directly in the EM question. For the LIST group, however, it is more complex, as we only observe the aggregated responses of participants in the List experiment. To examine the association between firms' EM practices and their choice of perceived motivations within the LIST group, we use the maximum likelihood estimator developed by Imai et al. (2015). This one-step estimator fully integrates the available information within a likelihood framework and therefore is fully efficient. Empirically, we calculate each respondent's predicted probability of having managed earnings in past years based on their characteristics. As we observe a difference between large and small firms in Table 5, we primarily consider firm size in the prediction model[13]. Our main outcome of interest is the participants' choice of perceived motivations.

Figure 3 presents the relevance of perceived earnings management (EM) motives across experimental groups (List vs. Direct), conditional on disclosed (DIRECT group) or predicted (LIST group) EM practices. The estimated differences in motive selection between EM and non-EM firms are reported for both experimental groups (shown in the third column of the first plot in each panel). We then test whether these estimated differences vary significantly between the Direct and LIST groups. The null hypothesis is that the controlled difference in motive selection between EM and non-EM firms does not differ across experimental groups. That is, under the assumption of random assignment, EM firms are expected to report the same set of underlying motivations regardless of whether they are assigned to the LIST or the DIRECT

13. We also consider other set of characteristics as predictors for EM practices: legal form, disclosure obligation, gender. The inferences remain the same.

group.

Panel A examines the relevance of tax savings as a potential EM motivation. The outcome variable is a binary indicator equal to one if a participant selected "tax savings" as a relevant motive. We observe that perceptions among non-EM firms are similar across experimental groups: 70 percent of non-EM firms in the DIRECT group and 67 percent in the LIST group consider tax savings a relevant motive. In contrast, the responses of EM firms diverge across groups. While 81 percent of EM firms in the DIRECT group report tax savings as a relevant motivation, only 67 percent of EM firms in the LIST group do so. However, when accounting for the baseline responses of non-EM firms (third column), we find no statistically significant difference (p-value larger than 10 percent) in the marginal relevance of tax savings between the groups.

The inferences change for other motivations. Panel B investigates "better credit terms" as a motivation for EM, the second most frequently cited motive. We find that 56 percent of EM firms in the LIST group select this motive, compared to only 35 percent of EM firms in the DIRECT group. Interestingly, 35 percent of non-EM firms also select this motive in the DIRECT group, implying that EM and non-EM firms appear indistinguishable in this regard. Once we control for the baseline (non-EM) responses, we find that EM firms in the LIST group are 28 percentage points more likely to select this motive than their non-EM counterparts—a gap that does not appear in the DIRECT group.

The pattern observed in Panel B extends to other motivations. EM firms in the DIRECT group tend to mimic the behavior of non-EM firms (with the exception of the motive "to give higher compensation to employees," as shown in Panel D), while EM firms in the LIST group provide more heterogeneous responses that likely reflect their true motivations. The incremental increase in motive selection prevalence in the LIST group ranges from 8 to 32 percentage points.

Overall, while both experimental groups derive indistinguishable prevalence estimates, our results show that the DIRECT group is more susceptible to social desirability bias when reporting motivations. We argue that participants in the DIRECT group treat the motivation question as a proxy for revealing their own behavior, prompting them to mask their true motivations to appear more socially acceptable. In contrast, participants in the LIST group are not explicitly labeled as EM or non-EM firms, reducing the pressure to self-censor. Therefore, the list experiment appears to mitigate this concern by allowing respondents to express motivations more freely. Although no true benchmark exists for the “correct” level of motive relevance, the higher estimates from the LIST group suggest these responses may better reflect respondents’ intrinsic views.

Our findings support the notion that managers perceive some motivations as more desirable than others. Specifically, consistent with the argument of Burgstahler et al. (2006), we find that managers are comfortable with reporting tax savings as a relevant EM motivation. At the same time, we also find weak evidence that managers tend to conceal EM for higher compensation, which is consistent with the insights from Tahir et al. (2019). We also document under-reporting of motivations such as better credit terms, improving perceptions among suppliers or customers, and enhancing perceptions among owners or investors. These results suggest that survey researchers studying EM motivations should consider potential social desirability bias and choose appropriate empirical methods that account for these effects.

Along similar logic, as an extension, we explore potential under-reporting behavior of firms with disclosure obligation. The results are summarized in Figure 4. The vertical lines in the graph correspond to the differences in indicated relevance of each motive between the EM and non-EM firms. Largely consistent with discussed implications in Figure 3, EM-firms assigned to the DIRECT group tend to under-report certain motivations such as improving perceptions among suppliers or customers, increasing compensation for employees, or enhancing perceptions among owners or investors.

Earnings Management Directions

In the final set of analyses, we ask about the direction in which firms manage earnings. We formulated the question in a less-sensitive way, without specifying that accounting options are used for own benefits to misrepresent earnings: “In which direction did the exercised accounting options and managers’ discretion in the past years/ in the current year influence your company’s earnings?” Beside “Increase in earnings” or “Decrease in earnings”, firm participants also have the option to select "Neutral" or "No accounting discretion exercised".

Similar to previous analyses, we examine the correlation between reported EM directions (in past years and in the current year) by experimental group and reported (for the DIRECT group) or predicted (for the LIST group) EM practices in the experiment. For both the LIST and DIRECT groups, only EM firms are considered, as the direction of EM is potentially relevant only for this group of firms. While we are able to identify EM firms directly for the DIRECT group, for the LIST group we predict whether a firm managed earnings based on its aggregated responses in the List experiment and its characteristics. The result is a dummy variable indicating which direction (up/down) was chosen by the EM firms. Figures (a)/(b) show the direction of EM in past years/current year. Figure (a) shows that compared to EM firms in the DIRECT group, EM firms in the LIST group are more likely to report their upward EM practices. At the same time, they are less likely to report their downward EM practices. This heterogeneity across experimental groups suggests that upward EM accounting practices are more sensitive than downward EM practices, leading to under-reporting in the DIRECT group. Even though the question is framed the same way for both groups, firms that receive the direct question on EM tend to hide their upward EM practices. The same implication can be observed in panel (b). While there is no difference between groups for downward EM, EM firms from the LIST group are 19 percentage points more likely to report their upward EM in the current year. Similarly, this also implies that accounting practices for upward EM are more sensitive compared to downward EM, and the List experiment works better with constructs that is potentially subject to the social desirability bias, for example, upward EM practices.

V. CONCLUSION

In this paper, we investigate how managers self-report their underlying EM motivations conditional on their EM practices. To answer this question, we conduct a large-scale survey experiment of firms in Germany by randomly assigning participants to one of the three experimental groups: DIRECT group, INDIRECT group, and a List Experiment group, which is designed to protect participants' privacy. We first estimate the prevalence of EM among firms operating in Germany to be around 20 percent. After asking about participants' EM practices, all experimental groups receive the same question on their perceived underlying EM motivations. We examine the correlation between the selection of perceived EM motivations across experimental groups and firm participants' own EM practices. We find that, even though all experimental groups receive the same question on motivations, which directs at general EM motivations among firms in the same industry, we still observe that firm participants in the Direct and the LIST group report different set of perceived motivations for EM. Our results imply that while acknowledging EM is not sensitive, reporting undesirable motives is prone to social desirability bias and can lead to under-reporting.

We interpret the results as follows. Firms in the LIST group have less pressure to report all relevant motives because their EM practices are not directly revealed in the list experiment. This bias is observable across multiple motivations, such as better credit terms or better perception by business partners. For the most commonly accepted motivation among firms in our sample, i.e., tax savings, no such bias exists. Our results suggest that the added privacy protection of a list experiment can increase executives' willingness to report truthfully about their underlying EM motivations, especially the ones that are considered as socially undesirable. This finding is valuable for researchers attempting to detect EM, as the choice of empirical identification strategies needs to incorporate the perceived social desirability feature of certain motivations.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used ChatGPT in order to check spelling or grammar. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

Ball, R. 2013. Accounting informs investors and earnings management is rife: Two questionable beliefs. *Accounting Horizons* 27 (4): 847–853.

Ball, R., and L. Shivakumar. 2005. Earnings quality in UK private firms: comparative loss recognition timeliness. *Journal of Accounting and Economics* 39 (1): 83–128.

Beneish, M. D. 1999. The detection of earnings manipulation. *Financial analysts journal* 55 (5): 24–36.

Bernard, D. 2016. Is the risk of product market predation a cost of disclosure? *Journal of Accounting and Economics* 62 (2-3): 305–325.

Beuselinck, C., F. Elfers, J. Gassen, and J. Pierk. 2023. Private Firm Accounting in the European Reporting Environment: Data and Research Perspectives. *Accounting and Business Research* 53 (4): 395–432.

Bischof, J., P. Doerrenberg, D. Rostam-Afschar, D. Simons, and J. Voget. 2025. The German Business Panel: Firm-Level Data for Accounting and Taxation Research. *European Accounting Review* 34 (4): 1499–1527.

Blair, G., W. Chou, and K. Imai. 2019. List experiments with measurement error. *Political Analysis* 27 (4): 455–480.

Blair, G., and K. Imai. 2012. Statistical analysis of list experiments. *Political analysis* 20 (1): 47–77.

Blaylock, B., F. Gaertner, and T. Shevlin. 2015. The association between book-tax conformity and earnings management. *Review of accounting studies* 20:141–172.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan. 2020. Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review* 110 (8): 2454–2484.

Burgstahler, D. C., L. Hail, and C. Leuz. 2006. The Importance of Reporting Incentives: Earnings Management in European Private and Public Firms. *The Accounting Review* 81 (5): 983–1016. Accessed April 3, 2023. http://www.jstor.org/stable/4093095.

Cade, N. L., J. L. Gunn, and A. J. Vandenberg. 2024. Measuring the Prevalence of Earnings Manipulations: A Novel Approach. *Journal of Accounting Research.*

Cazier, R., S. Rego, X. Tian, and R. Wilson. 2015. The impact of increased disclosure requirements and the standardization of accounting practices on earnings management through the reserve for income taxes. *Review of Accounting Studies* 20:436–469.

Cohen, D. A., and P. Zarowin. 2008. Economic consequences of real and accrual-based earnings management activities. *The Accounting Review* 83 (1): 758–787.

Commerford, B. P., D. R. Hermanson, R. W. Houston, and M. F. Peters. 2019. Auditor sensitivity to real earnings management: The importance of ambiguity and earnings context. *Contemporary accounting research* 36 (2): 1055–1076.

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

Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary accounting research* 13 (1): 1–36.

Dichev, I. D., J. R. Graham, C. R. Harvey, and S. Rajgopal. 2013. Earnings quality: Evidence from the field. *Journal of Accounting and Economics* 56 (2-3): 1–33.

Dutta, S., and Q. Fan. 2014. Equilibrium earnings management and managerial compensation in a multiperiod agency setting. *Review of Accounting Studies* 19:1047–1077.

Dyreng, S. D., S. A. Hillegeist, and F. Penalva. 2022. Earnings management to avoid debt covenant violations and future performance. *European Accounting Review* 31 (2): 311–343.

Elffers, H., R. H. Weigel, and D. J. Hessing. 1987. The consequences of different strategies for measuring tax evasion behavior. *Journal of Economic Psychology* 8 (3): 311–337.

Fischer, M., and K. Rosenzweig. 1995. Attitudes of students and accounting practitioners concerning the ethical acceptability of earnings management. *Journal of business ethics* 14 (6): 433–444.

Francis, J., R. LaFond, P. Olsson, and K. Schipper. 2005. The market pricing of accruals quality. *Journal of Accounting and Economics* 39 (2): 295–327.

Franz, D. R., H. R. HassabElnaby, and G. J. Lobo. 2014. Impact of proximity to debt covenant violation on earnings management. *Review of Accounting Studies* 19:473–505.

Frye, T., S. Gehlbach, K. L. Marquardt, and O. J. Reuter. 2017. Is Putin's popularity real? *Post-Soviet Affairs* 33 (1): 1–15.

Garcia-Agundez, A., O. Ojo, H. A. Hernández-Roig, C. Baquero, D. Frey, C. Georgiou, M. Goessens, R. E. Lillo, R. Menezes, N. Nicolaou, et al. 2021. Estimating the COVID-19 prevalence in Spain with indirect reporting via open surveys. *Frontiers in Public Health* 9:658544.

Gassen, J., and M. Muhn. 2018. Financial transparency of private firms: Evidence from a randomized field experiment. *Journal of Accounting Research.*

Gerakos, J. 2012. Discussion of detecting earnings management: A new approach. *Journal of Accounting Research* 50 (2): 335–347.

Glynn, A. N. 2013. What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly* 77 (S1): 159–172.

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

Graham, J. R., S. Li, and J. Qiu. 2008. Corporate misreporting and bank loan contracting. *Journal of financial economics* 89 (1): 44–61.

Hewitt, M., F. D. Hodge, and J. H. Pratt. 2020. Do shareholders assess managers' use of accruals to manage earnings as a negative signal of trustworthiness even when its outcome serves shareholders' interests? *Contemporary Accounting Research* 37 (4): 2058–2086.

Hope, O.-K., W. B. Thomas, and D. Vyas. 2013. Financial Reporting Quality of U.S. Private and Public Firms. *The Accounting Review* 88 (5): 1715–1742.

Imai, K. 2011. Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* 106 (494): 407–416.

Imai, K., B. Park, and K. F. Greene. 2015. Using the predicted responses from list experiments as explanatory variables in regression models. *Political Analysis* 23 (2): 180–196.

Jensen, M. C. 2005. Agency costs of overvalued equity. *Financial management* 34 (1): 5–19.

Jing, L., Q. Lu, Y. Cui, H. Yu, and T. Wang. 2018. Combining the randomized response technique and the network scale-up method to estimate the female sex worker population size: an exploratory study. *Public health* 160:81–86.

Johnson, E. N., G. Fleischman, S. Valentine, and K. B. Walker. 2011. Managers' ethical evaluations of earnings management and its consequences. *Contemporary Accounting Research, forthcoming.*

Kasznik, R., and M. F. McNichols. 2002. Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices. *Journal of Accounting Research* 40 (3): 727–759.

Kim, R., and W. Luo. 2022. Customer concentration and earnings management: Evidence from the Sarbanes–Oxley Act. *European Accounting Review* 31 (4): 905–936.

Kim, T., H. Kyung, and J. Ng. 2022. Top management team incentive dispersion and earnings quality. *Contemporary Accounting Research* 39 (3): 1949–1985.

Kuklinski, J. H., M. D. Cobb, and M. Gilens. 1997. Racial attitudes and the" New South". *The Journal of Politics* 59 (2): 323–349.

Laga, I., L. Bao, and X. Niu. 2021. Thirty years of the network scale-up method. *Journal of the American Statistical Association* 116 (535): 1548–1559.

Leuz, C., and P. D. Wysocki. 2016. The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of accounting research* 54 (2): 525–622.

Lisowsky, P., and M. Minnis. 2020. The Silent Majority: Private U.S. Firms and Financial Reporting Choices. *Journal of Accounting Research* 58 (2): 557–588.

Minnis, M., and N. Shroff. 2017. Why Regulate Private Firm Disclosure and Auditing? *Accounting and Business Research* 47 (5): 473–502.

Myers, J. N., L. A. Myers, and D. J. Skinner. 2007. Earnings momentum and earnings management. *Journal of accounting, auditing & Finance* 22 (2): 249–284.

Phillips, L. W. 1981. Assessing measurement error in key informant reports: A methodological note on organizational analysis in marketing. *Journal of marketing research* 18 (4): 395–415.

Rigol, N., R. Hussam, and B. Roth. 2017. Targeting high ability entrepreneurs using community information: Mechanism design in the field. *Job Market Paper,* 1–82.

Schedler, A., and R. Sarsfield. 2007. Democrats with adjectives: Linking direct and indirect measures of democratic support. *European journal of political research* 46 (5): 637–659.

Srivastava, A., J. M. Ramirez, S. Diaz, J. Aguilar, A. Ortega, A. Fernandez-Anta, and R. Lillo-Rodriguez. 2023. Estimating Temporal Trends using Indirect Surveys. *arXiv preprint arXiv:2307.06643.*

Sutherland, B. 2025. "Private Firms and the Economic Role of Accounting." Working Paper.

Tahir, M., S. Ibrahim, and M. Nurullah. 2019. Getting compensation right-The choice of performance measures in CEO bonus contracts and earnings management. *The british accounting review* 51 (2): 148–169.

Teo, A. K. J., K. Prem, M. I. Chen, A. Roellin, M. L. Wong, H. H. La, and A. R. Cook. 2019. Estimating the size of key populations for HIV in Singapore using the network scale-up method. *Sexually transmitted infections* 95 (8): 602–607.

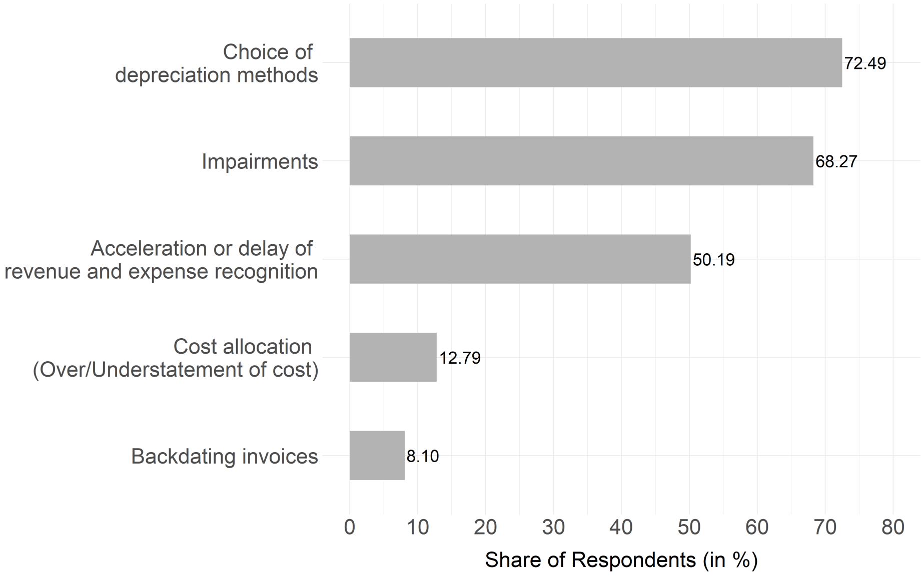
Tourangeau, R., and T. Yan. 2007. Sensitive questions in surveys. *Psychological bulletin* 133 (5): 859.

Traunmüller, R., S. Kijewski, and M. Freitag. 2019. The silent victims of sexual violence during war: Evidence from a list experiment in Sri Lanka. *Journal of Conflict Resolution* 63 (9): 2015–2042.

U.S. Small Business Administration, Office of Advocacy. 2023. Frequently Asked Questions About Small Business, 2023. *Available online: Link.*

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

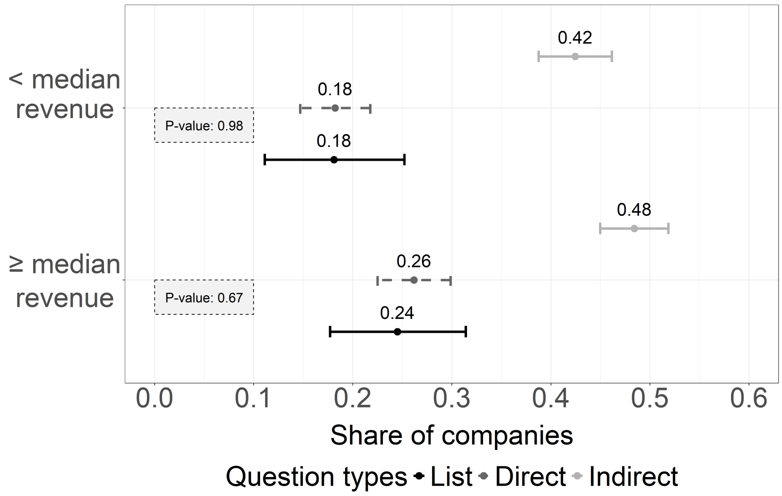
Figure 1: Participants' Classification of Earnings Management Practices



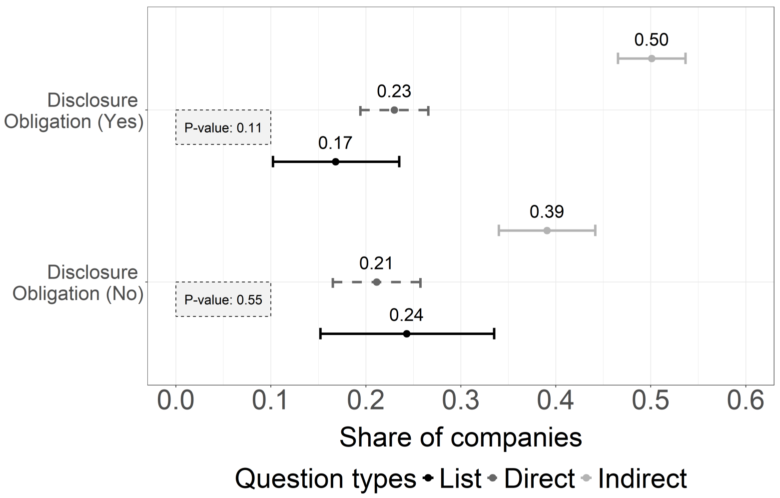
Note: This figure shows the distribution of respondents' answers to the question: "Which of the following accounting measures can be used for earnings management in compliance with the GAAP?". We include one illegal earnings management practice "Backdating invoices" to classify firm participants who are more familiar with earnings management practices.

Figure 2: Prevalence of EM: Subgroup Analysis

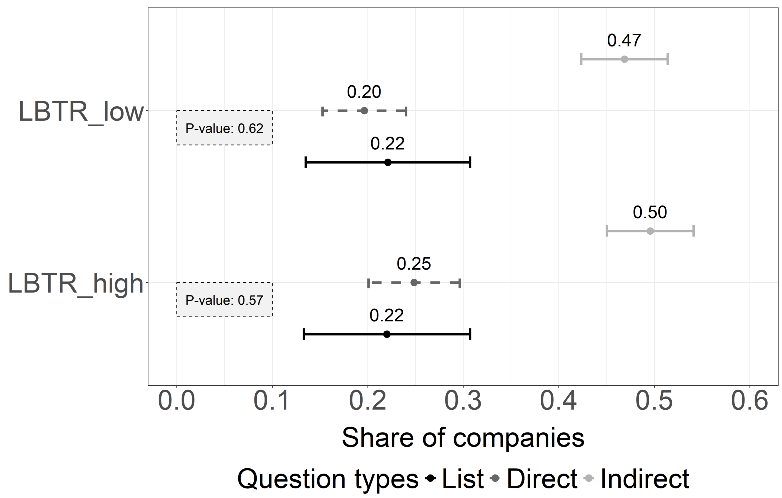
Panel A: Firm size



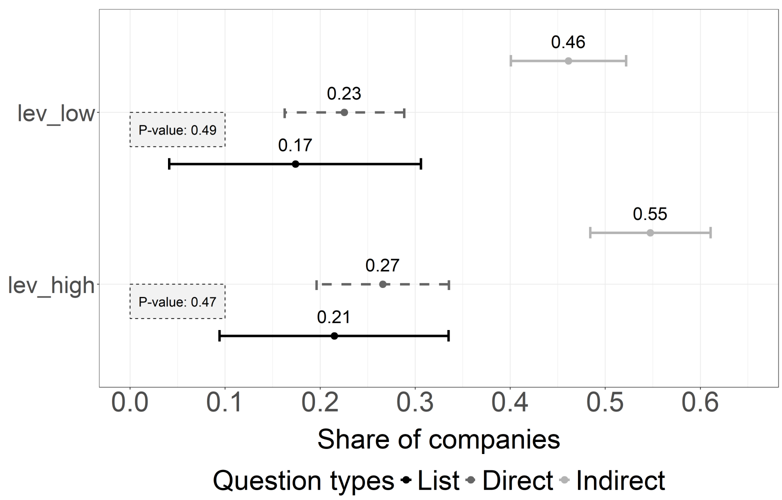
Panel B: Disclosure obligation



Panel C: Local business tax rate



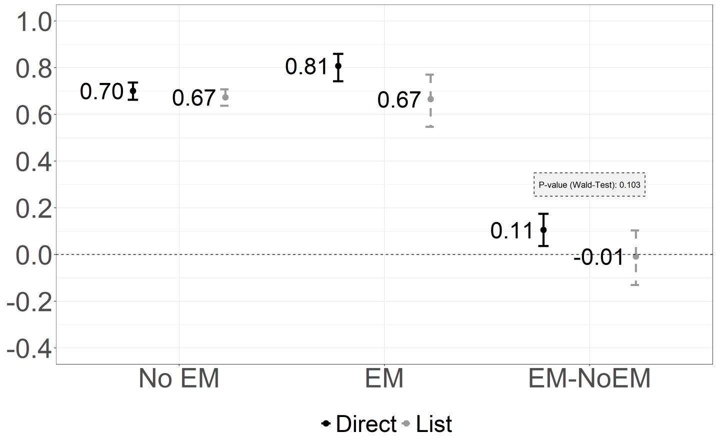
Panel D: Leverage (Current liabilities/ Total assets)



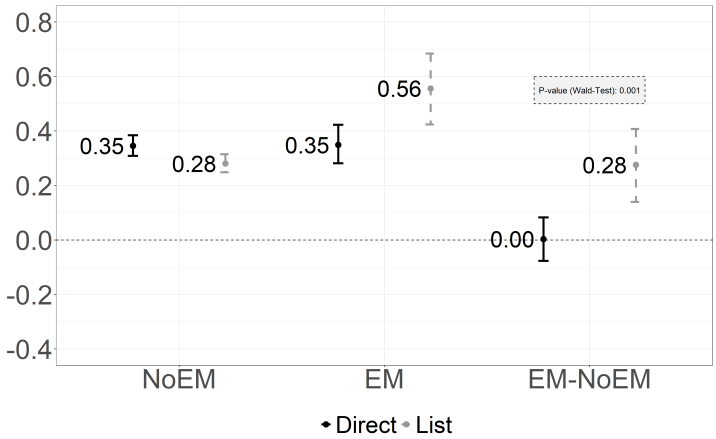
Note: The figures display linear estimators of the prevalence of earnings management, reported separately for experimental groups and conditional on four firm characteristics: (a) firm size, (b) disclosure obligation, (c) local business tax rate, and (d) leverage. In Panel (a), firms are classified by size based on whether their revenue is above or below the sample median (50,000 euros). In Panel (b), a firm is considered to have a disclosure obligation if it is required to publish annual financial statements in the Federal Gazette (Question: *Do disclosure obligations apply to your company (e.g. publication of the annual financial statements in the Federal Gazette)?*). Panel (c) defines the variable *LBTR_high* as indicating that a firm's municipality imposes a local business tax rate above the sample median. Leverage is defined as the ratio of current liabilities to total assets. In Panel (d), firms with leverage above the sample median are categorized as *leverage_high*. The indirect(share) group receives the question "In past years, what percentage of companies in your industry (in %) have taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?". *p*-values from T-tests to test for the differences between the Direct and the LIST group are reported in the figures. If applicable, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Figure 3: Relevance of Perceived Motives by Experimental Groups and EM Practices

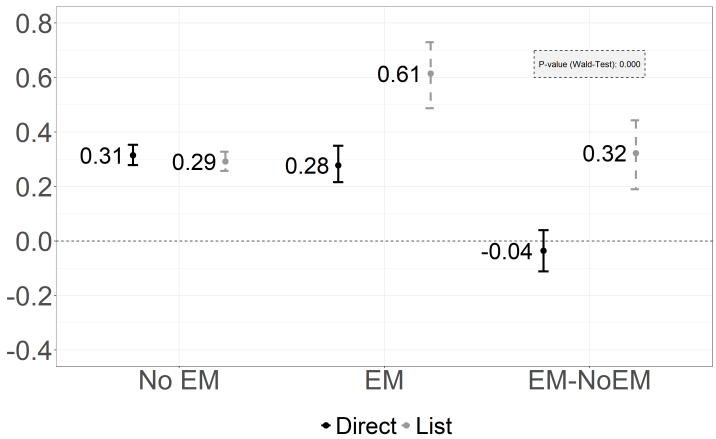
Panel A: Save taxes



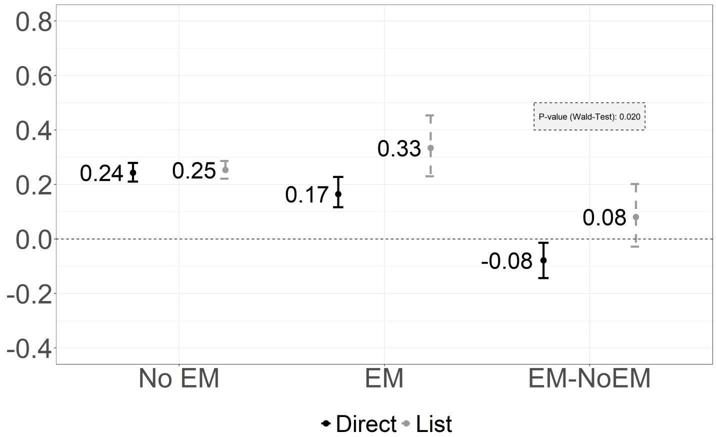
Panel B: Ensure better credit terms



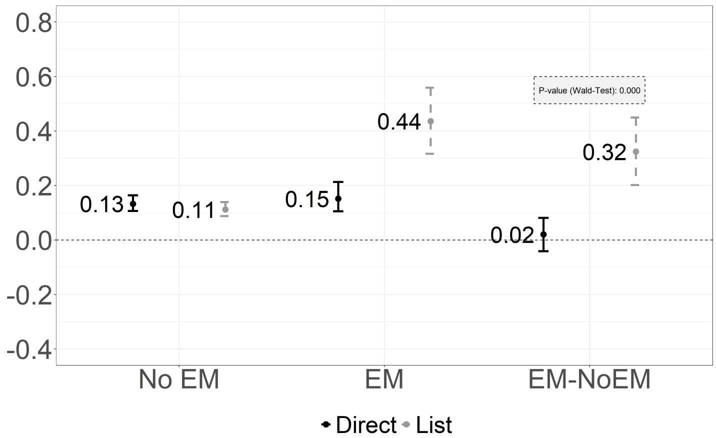
Panel C: Give customers/suppliers more certainty about the stable development of business



Panel D: Give higher compensation for employees (e.g., higher bonuses)

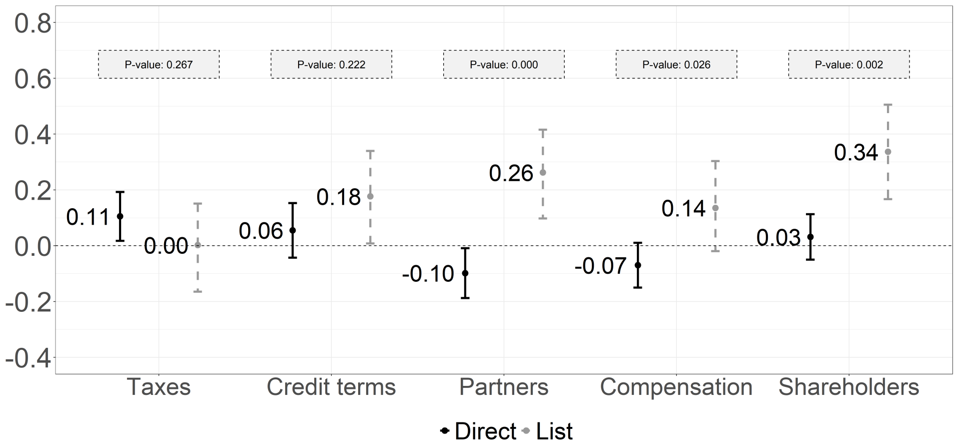


Panel E: Give the owners or investors more security about the stable development of business



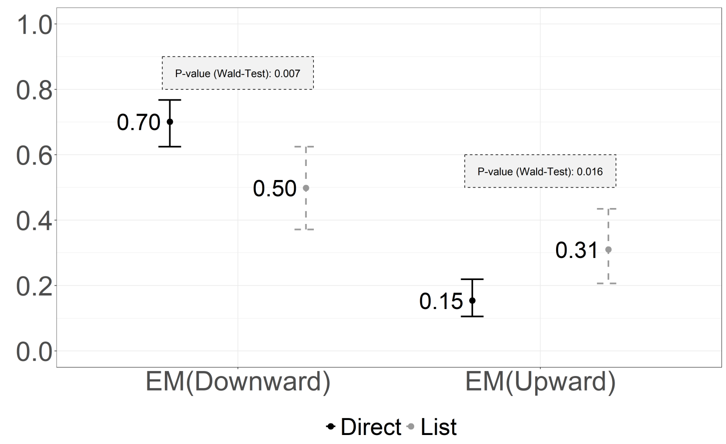
Note: The figures illustrate the relevance of perceived earnings management (EM) motives by experimental group (List vs. Direct) and disclosed (DIRECT group) or predicted (LIST group) EM practices. Panels are presented in order of the overall prevalence of the motives indicated by participants. Each panel shows the prevalence of perceived EM motives by experimental group and firms' EM practice for each motive. For the LIST group, the predicted responses from the list experiments are used as predictors (EM(Yes/No)) for the selection of motives. The dependent variable is a binary indicator equal to one if a participant selected a given motive. Differences in outcomes between EM and non-EM firms are reported separately for both the Direct and LIST groups. Additionally, Wald test *p*-values for differences between the two experimental groups are presented.

Figure 4: Relevance of Perceived Motives of Firms with Disclosure Obligation

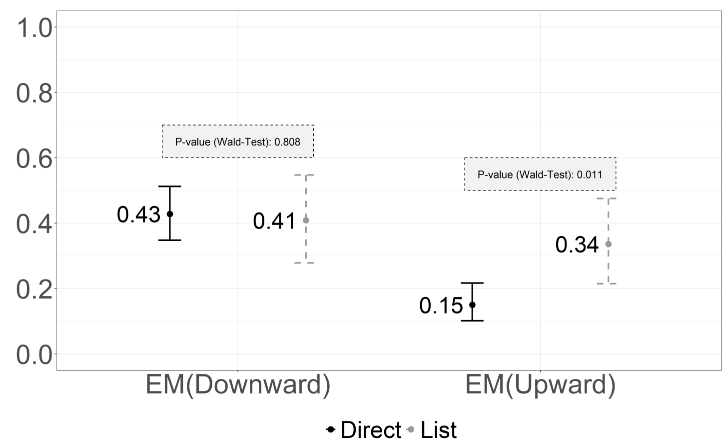


Note: This figure explores potential under-reporting behavior by comparing the differences between List experiment and DIRECT group in motive selection for firms with disclosure obligation (Survey Question: *Do disclosure obligations apply to your company (e.g. publication of the annual financial statements in the Federal Gazette)?*). The vertical lines in the graph correspond to the difference in indicated relevance of each motive between the EM and non-EM firms. The dependent variable is a binary indicator equal to one if a participant selected a given motive. For the LIST group, the predicted responses from the list experiments are used as predictors (EM(Yes/No)) for the selection of motives. Wald test *p*-values for differences between the two experimental groups are presented.

Figure 5: Directions of EM



(a) Past years



(b) Current year

Note: The figures focus on firms identified as engaging in EM and show the disclosed EM directions (Survey Question: *In which direction did the exercised accounting options and managers' discretion in the past years/ in the current year influence your company's earnings?*) for both past and current years by experimental group (List vs. Direct). EM firms in the DIRECT group are identified through self-reporting in the survey, while in the LIST group, EM practice is predicted based on responses to the list experiment. The outcome is a binary indicator reflecting the direction of EM selected by these firms (Answer options: *Increase in earnings; Neutral; Decrease in earnings; No accounting discretion exercised*). All models control for firm size, legal form, disclosure obligation, and the respondent's gender. *P*-values from Wald tests comparing the responses from the Direct and LIST groups are reported.

Table 1: Standard and Double List Experiment Design

Panel A: Standard List Experiment	
Treatment Group	**Control Group**
1. We have outsourced activities to third-party service providers to reduce costs.	1. We have outsourced activities to third-party service providers to reduce costs.
2. We have reincorporated outsourced activities to reduce supply-chain risks.	2. We have reincorporated outsourced activities to reduce supply-chain risks.
3. We have intensively communicated with our customers to enhance our revenues.	3. We have intensively communicated with our customers to enhance our revenues.
4. We have invested in risk management to increase internal efficiency.	4. We have invested in risk management to increase internal efficiency.
5. We have taken advantage of available accounting choices and managers‘ discretion for own benefit to misrepresent earnings.	
Panel B: Double List Experiment	
List Experiment Group 1	**List Experiment Group 2**
Treatment List A	**Control List A**
1. We have outsourced activities to third-party service providers to reduce costs.	1. We have outsourced activities to third-party service providers to reduce costs.
2. We have reincorporated outsourced activities to reduce supply-chain risks.	2. We have reincorporated outsourced activities to reduce supply-chain risks.
3. We have intensively communicated with our customers to enhance our revenues.	3. We have intensively communicated with our customers to enhance our revenues.
4. We have invested in risk management to increase internal efficiency.	4. We have invested in risk management to increase internal efficiency.
5. We have taken advantage of available accounting choices and managers‘ discretion for own benefit to misrepresent earnings.	
Control List B	**Treatment List B**
1. We have restructured business units to reduce costs.	1. We have restructured business units to reduce costs.
2. We have reduced the number of suppliers to reduce delivery risks.	2. We have reduced the number of suppliers to reduce delivery risks.
3. We have adjusted our project management to increase internal efficiency.	3. We have adjusted our project management to increase internal efficiency.
4. We have invested in product marketing to improve sales.	4. We have invested in product marketing to improve sales.
	5. We have taken advantage of available accounting choices and managers‘ discretion for own benefit to misrepresent earnings.

Note: This table provides examples of a standard list experiment (Panel A) and a double list experiment (Panel B). In a standard list experiment, participants are randomly assigned to either a treatment group (shown on the left) or a control group (shown on the right). Each participant receives the corresponding list and is asked to report the total number of statements that apply to them, without specifying which statements apply (Question: *How many of the following measures has your company taken in recent years?* Answer: *We have taken __ of measures.*). The difference between the mean responses of the two group provides an estimate of the prevalence of the sensitive behavior described in statement 5 in the treatment group. In our double list experiment, participants are randomly assigned to either List Experiment Group 1 (shown on the left) or List Experiment Group 2 (shown on the right). Each participant receives two different lists. Participants assigned to Group 1 serve as treatment observations for List A and control observations for List B, and vice versa for Group 2. Both treatment lists, A and B, contain the same sensitive statement (statement 5) but different control statements. In this setup, participants read through both lists and report the number of true statements for each (Question: *How many of the following measures has your company taken in recent years?* Answer: *We have taken _ of measures.*). The estimate of the prevalence of the sensitive behavior is derived by averaging the responses from both lists. To mitigate the order effect, the display order of items in every list is randomized in our experiments.

Table 2: Firm Characteristics

	Statistical Register (2022) (1)	Our Sample (2)	LIST (3)	DIRECT (4)	INDIRECT (5)	P values across groups (6)
Number of employees						
Median	NA	4	4	4	4	0.534
Mean	NA	48	79	16	18	0.352
0 - 9	0.868	0.703	0.712	0.703	0.688	0.387
10 - 49	0.105	0.229	0.221	0.235	0.237	0.528
≥ 50	0.027	0.068	0.067	0.062	0.076	0.460
Revenue						
Median	NA	550,000	500,000	520,000	600,000	0.681
Mean	NA	3,005,469	2,831,810	3,111,950	3,240,131	0.352
< 2 million	0.919	0.559	0.566	0.562	0.541	0.428
2 - 10 million	0.060	0.394	0.388	0.385	0.413	0.337
> 10 million	0.022	0.047	0.046	0.053	0.045	0.665
Legal Form						
Sole Proprietorship	0.592	0.214	0.204	0.227	0.222	0.263
Business Partnerships	0.121	0.152	0.154	0.151	0.150	0.935
Corporations	0.237	0.581	0.591	0.576	0.565	0.355
Other legal forms	0.052	0.053	0.051	0.046	0.063	0.200
Disclosure Obligation						
Yes	NA	0.662	0.672	0.638	0.666	0.241
N	3,435,478	4,086	2,034	1,019	1,033	NA

Note: This table shows the share of firms by employment size class, revenue class, and legal form across the statistical register data, our sample, and the experimental groups. For the main analyses, we restrict the sample to firm participants who are CEOs or owners and who do not classify "backdating invoices" as GAAP-compliant. The statistical register data was obtained from the German Federal Statistical Office (Statistisches Bundesamt (Destatis)) for the 2022 reporting period. The INDIRECT group comprises both the indirect (Yes/No) group ($N = 437$) and the indirect (Share) group ($N = 596$). Column (6) presents the p-values from statistical tests assessing equality across experimental groups. The Kruskal-Wallis test is used for comparing median values, while one-way ANOVA is employed for mean values. For testing the distribution of categorical variables across experimental groups, chi-squared test is applied. If applicable, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 3: Respondents Characteristics

	Our Sample	Revenue (Median)	Revenue (Mean)	Employee Number (Median)	Employee Number (Mean)	LIST	DIRECT	INDIRECT	P values across groups
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education									
University degree	0.633	600,000	3,682,156	5	19	0.626	0.628	0.651	0.442
Master craftsperson, technician	0.158	500,000	1,377,413	4	10	0.162	0.153	0.153	0.760
Apprenticeship/Other	0.199	500,000	2,520,461	3	28	0.199	0.212	0.188	0.463
Position									
Owner/CEO	1	550,000	3,005,469	4	48	1	1	1	NA
Gender									
Male	0.852	600,000	3,151,679	4	56	0.852	0.847	0.857	0.809
N	4,086	4,086	4,086	4,086	4,086	2,034	1,019	1,033	NA

Note: This table presents the respondents characteristics by education, position, and gender in our sample and across experimental groups. For the main analyses, we restrict the sample to firm participants who are CEOs or owners and who do not classify "backdating invoices" as GAAP-compliant. Column (2) to (5) report the median and mean revenue and employee number of the corresponding firms of each category. The INDIRECT group includes both the indirect(Yes/No) group ($N = 437$) and the indirect(Share) group ($N = 596$). Column (6) presents the p-values from statistical tests assessing equality across experimental groups (6) to (8). The Kruskal-Wallis test is used for comparing median values, while one-way ANOVA is employed for mean values. For testing the distribution of categorical variables across experimental groups, chi-square test is applied. If applicable, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 4: Prevalence of EM

	Experimental Groups				Previous Studies		
	List	Direct	Indirect(Share)	Indirect(Yes/No)	Graham et al. (2005)	Dichev et al. (2013)	Cade et al. (2024)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant(Prevalence)	0.214***	0.226 ***	0.457***	0.643***	0.080	0.246	0.205
	(0.025)	(0.013)	(0.013)	(0.023)			
T-test	Column(1) = Column(2) (Diff = 0.012, *p*-value = 0.687)						
Observations	2,034	1,019	596	437	401	375	738

Note: This table shows the linear estimators for the prevalence of earnings management from experimental groups without controls in column (1) to (4). Column (1) reports the linear estimate derived from the double list experiment. Column (2) reports the prevalence when respondents are directly asked about EM activities in their own firms. Column (3) reports the perceived prevalence (in %) of EM in other firms that operate in the same industry as the responding firms (Survey question: *In past years, what percentage of companies in your industry (in %) have taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?*). Column (4) also reports the perceived prevalence (in %) of EM in other firms that operate in the same industry as the responding firms, based on a similar question: *In past years, have companies in your industry taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?* Standard errors are reported. Column (5) to column (7) report the estimates from prior published survey studies. The prevalence estimates derived from Dichev et al. (2013) and Cade et al. (2024) are both based on an indirect question that is similar to our Indirect(Share) question in Column (3). For survey questions used in those studies in detail, please refer to the Appendix Table A.1. The prevalence from Dichev et al. (2013) is weighted with the number of public/ private firms in their sample. If applicable, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 5: Perceived EM Motives

	Tax savings	Credit terms	Suppliers/ Customers	Employee compensation	Owner/ Investors	N
	(1)	(2)	(3)	(4)	(5)	(6)
Total	0.695	0.375	0.354	0.249	0.202	4,037
DIRECT	0.699	0.317	0.304	0.217	0.159	1,019
LIST	0.666	0.331	0.370	0.263	0.180	2,034
P-value	0.067*	0.422	0.000***	0.005***	0.142	NA
Small	0.708	0.351	0.335	0.220	0.192	1,810
Large	0.683	0.396	0.371	0.276	0.209	2,227
P-value	0.082*	0.003***	0.017**	0.000***	0.191	NA
DisclObli_no	0.729	0.319	0.294	0.229	0.149	1,142
DisclObli_yes	0.671	0.408	0.385	0.266	0.230	2,237
P-value	0.000***	0.000***	0.000***	0.017***	0.000***	NA
Tax_low	0.692	0.364	0.375	0.265	0.209	1,313
Tax_high	0.678	0.402	0.362	0.249	0.225	1,315
P-value	0.441	0.044*	0.473	0.360	0.331	NA
Leverage_low	0.667	0.364	0.373	0.255	0.236	679
Leverage_high	0.677	0.427	0.434	0.288	0.228	705
P-value	0.709	0.016**	0.020**	0.166	0.749	NA

Note: This table reports which perceived earnings management (EM) motives are selected by firm respondents in the total sample and across subgroups: experimental groups (List experiment or DIRECT group), firm size, disclosure obligation, local business tax rate, and leverage. Column (6) reports the number of responses. The relevant survey question is: *"There are different reasons why a company might take advantage of accounting choices and discretion for their own benefits to misrepresent earnings. Which of the following reasons are relevant for companies in your industry?"* The answer options are: *Better lending opportunities. Higher compensation for employees (e.g., higher bonuses). Give customers/suppliers more certainty about the stable development of business. Give the owners or investors more security about the stable development of business. To save taxes.* Motives are ordered by their overall prevalence across columns. We test whether firms assigned to the List Experiment group differ in their motive selections from those in the DIRECT group. Firms are classified as small or large based on whether their revenue is above or below the sample median (50,000 euros). *DisclObli_yes* indicates firms that have disclosure obligation to publish annual financial statements in the Federal Gazette (Survey Question: *Do disclosure obligations apply to your company (e.g. publication of the annual financial statements in the Federal Gazette)?*). *Tax_high* indicates that a firm's municipality imposes a local business tax rate above the sample median. Leverage is defined as the ratio of current liabilities to total assets. Firms with above-median leverage are labeled as *leverage_high*. We use T-tests to test the distribution across groups. The resulting *p*-values are reported. If applicable, ***, **, * denote statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Appendix A. PRIOR SURVEY LITERATURE

The following table presents estimates of the prevalence of EM from prior survey research. It is worth noting that the mentioned studies vary in terms of sample, EM definition, and research methods.

Table A.1: Prevalence of Earnings Management from Prior Survey Literature

Study	Sample	Definition of EM	Estimates
Graham, Harvey, and Rajgopal (2005)	401 U.S. financial executives (46 of them are private firms)	"Near the end of the quarter, it looks like your company might come in below the desired earnings target. Within what is permitted by GAAP, which of the following choices might your company make?"	8% of respondents report being willing to "alter accounting estimates"; 40% of respondents report being willing to "book revenues now rather than next quarter"
Dichev et al. (2013)	CFOs of public (169) and private (206) U.S. firms	"From your impressions of companies in general, in any given year, what percentage of companies use discretion within GAAP to report earnings which misrepresent the economic performance of the business?"	Mean response: 18% (Public firms); 30% (Private firms)
Cade, Gunn, and Vandenberg (2024)	971 listed firms in the Russell 3000 Index	"From your impressions of companies in general, at any point in a five-year period, what percentage of companies use discretion allowed by GAAP to report earnings that misrepresent the economic performance of the business?"	Mean response: 21%; Median response: 10%
		"I am aware of a time in the past five years when my company used discretion allowed by GAAP to report earnings that misrepresent the economic performance of the business."	Direct question: 6.6%; List experiment: 5.4%

Note: This table presents an overview of estimates of the prevalence of earnings management from prior research studies that we consider most comparable.

Appendix B. LIST EXPERIMENT METHOD

In this section, we first formally derive the prevalence estimator from our list experiment design. Then, we demonstrate how predicted responses from the list experiment can be used as explanatory variables in regression models.

List Experiment Prevalence Estimates

Following the notations of Imai (2011), let's first consider the standard design for a single list experiment where there are J control items and one sensitive item. Let T_i denote the treatment status of the respondent i. $T_i = 0$ means that the respondent views only the list of J control items whereas $T_i = 1$ indicates that the respondent views the list of J control items and one sensitive item, i.e., $J+1$ items in total.

Suppose that each respondent has a latent response to each control item $j = 1, ..., J$, which may depend on the treatment status. $Z_{ij}(t)$ denotes the latent response for respondent i to item j under treatment status t. For example, $Z_{12}(1) = 1$ means that the latent answer of respondent 1 to the second control item is affirmative under the treatment condition. According to this logic, $Z_{i,J+1}(1)$ denotes the latent response of the respondent i to the sensitive statement under the treatment condition. We use $Z^*_{i,J+1}$ to denote respondent i's truthful response to the sensitive item. Under the *no liars* assumption, for each respondent i, we assume $Z_{i,J+1}(1) = Z^*_{i,J+1}$. $Z_{i,J+1}(0)$ is undefined because the sensitive item is only presented in the treatment list.

Since respondents are asked to provide the total number of items on the list that apply, their responses can be defined as $Y_i(1) = \sum_{j=1}^{J+1} Z_{ij}(1)$ for treatment lists and $Y_i(0) = \sum_{j=1}^{J} Z_{ij}(0)$ for control lists. When the treatment status is randomly assigned and there is no design effect[14], the quantity that interests us, $Z^*_{i,J+1}$, is

$$Z^*_{i,J+1} = Z_{i,J+1}(1) = Y_i(1) - Y_i(0). \tag{6}$$

Empirically, the standard difference-in-means estimator that is commonly used to analyze

14. Under the *no design effect* assumption, adding sensitive item does not change respondents' responses to the control items. Formally: $\sum_{j=1}^{J} Z_{ij}(0) = \sum_{j=1}^{J} Z_{ij}(1)$.

the list experiment is

$$\hat{\tau} = \frac{1}{N_1}\sum_{i=1}^{N} T_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)Y_i, \tag{7}$$

where $N_1 = \sum_{i=1}^{N} T_i$ and $N_0 = N - N_1$. $\hat{\tau}$ is an unbiased estimate of the population average response to the sensitive item. Formally, $\mathbb{E}(\hat{\tau}) = Pr(Z_{i,J+1}(1) = 1)$. Assuming linear relationships, $\hat{\tau}$ is identical to the slope coefficient of a simple linear regression of Y_i on T_i:

$$Y_i = \lambda + \tau T_i + \varepsilon_i, \tag{8}$$

where λ is the average number of items answered affirmatively in the control group.

Different from a standard list experiment design, we conducted a double list experiment to estimate the prevalence for accrual-based earnings management. As illustrated in Table 1, firm participants assigned to Group 1 view the sensitive statement in conjunction with the innocuous control list A, whereas participants assigned to Group 2 receive the sensitive item regarding earnings management in conjunction with the innocuous control list B. The standard difference-in-means estimator has to be slightly adjusted. Specifically, we apply the difference-in-means estimator separately to each of the list and compute the average between the two. Formally we write:

$$\hat{\tau} = \left[\left\{\frac{1}{N_1}\sum_{i=1}^{N} T_i Y_i^A - \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)Y_i^A\right\} + \left\{\frac{1}{N_1}\sum_{i=1}^{N} T_i Y_i^B - \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)Y_i^B\right\}\right]/2, \tag{9}$$

The adapted regression is then:

$$Y_i = \lambda + \tau T_i + \beta List_i + \varepsilon_i. \tag{10}$$

Y_i is a discrete variable that ranges from 0 to 5 for treatment lists and from 0 to 4 for control lists. T_i is a binary variable equal to one if the response comes from a treatment list, and zero otherwise. *List* is an indicator variable equal to one for list A, and zero for List B. This variable controls for the different innocuous items in the two treatment lists. Since each firm participant views two lists, we cluster standard errors at the individual level. The main parameter of interest is τ, which corresponds to the difference between the average number of items reported in the treatment list and the average number of items reported in the control list. This difference indicates the prevalence of the sensitive statement, which is accrual-based earnings management in our case.

Using Predicted Responses from List Experiment

We are further interested in using the latent responses from the list experiment as predictors in regression models for the outcome variable, earnings management motives. If we assume linear relationships, we have

$$Motiv_i = \alpha + \beta^T X_i + \gamma Z^*_{i,J+1} + \zeta Y^*_i + \varepsilon_i, \tag{11}$$

where X_i is a vector of respondents' characteristics. The main parameter of interest is the coefficient for the latent response to earnings management, γ. However, we do not directly observe the answer to the sensitive item $Z^*_{i,J+1}$ and the number of control items answered affirmatively given the response to the sensitive item and characteristics of the respondents Y^*_i. The next sections explains the methodology proposed by Imai et al. (2015) to estimate these outcome regression models.

Multivariate Regression Models

We first review the multivariate regression models developed by Imai (2011) and Blair and Imai (2012), as they are essential to derive the estimators. Their methodology is represented by the following two submodels:

$$g_\delta(x) = Pr(Z^*_{i,J+1} = 1|X_i = x; \delta) \tag{12}$$

$$h_\psi(y|x,z) = Pr(Y^*_i = y|X_i = x, Z^*_{i,J+1} = z; \psi) \tag{13}$$

for $z = 0,1, y = 0,1,...,J$, and $x \in \chi$, where δ and ψ are vectors of unknown parameters. Equation (12) models the probability of an affirmative response to the sensitive item, given a vector of characteristics, whereas Equation (13) models the response to the J control items given the response to the sensitive item and a vector of characteristics.

The naive two-step estimator

To estimate γ, we proceed in the following two steps: First, we fit the multivariate regression models in Section B.2.1. Then, for each respondent, we compute the predicted probability of affirmatively answering the sensitive item as a function of respondents' characteristics. This quantity is denoted by $\hat{Z}^*_{i,J+1} = g_{\hat{\delta}}(x_i)$. Similarly, we also estimate the expected response to the control items $\hat{Y}^*_i$. Second, we include these predicted values in the outcome regression. That is, we fit the linear regression model given in equation (11) by replacing $Z^*_{i,J+1}$ and Y^*_i with their predicted values, $\hat{Z}^*_{i,J+1}$ and $\hat{Y}^*_i$.

As discussed in Imai et al. (2015), this naive two-step estimator suffers from several problems. First, the estimator will be biased if the outcome regression is nonlinear. Second, the standard error will be underestimated because the first-stage estimation uncertainty is not taken into account. Third, even if the outcome regression is linear, the naive two-step estimator is statistically inefficient because the predicted values $\hat{Z}^*_{i,J+1}$ and $\hat{Y}^*_i$ are highly correlated with X_i.

The improved two-step estimator

Imai et al. (2015) propose an improved two-step estimator by utilizing all information from the list experiment. While the naive two-step estimator is based on the conditional expectation of the outcome given X_i, the improved version conditions on all information from the list experiment, (Y_i, T_i, X_i). This yields the following regression equation:

$$\mathbb{E}(Motiv_i|X_i,T_i,Y_i) = \alpha + \sum \beta X_i + \gamma Pr(Z^*_{i,J+1} = 1|X_i,T_i,Y_i) + \zeta\{Y_i - T_i Pr(Z^*_{i,J+1} = 1|X_i,T_i,Y_i)\} \tag{14}$$

This equation could be fitted via a two-step procedure similar to the naive two-step estimator. This improved two-step estimator incorporates the first-stage estimation uncertainty, but it is only applicable when the outcome model is linear. It is also not statistically efficient because the first step does not use the information contained in the outcome regression.

The general full maximum likelihood estimator

Imai et al. (2015) develop the general full maximum likelihood estimator that incorporates all the information from the data in the likelihood framework and therefore is fully efficient. The estimator can also be extended to a wide range of outcome models. Consider the general outcome model defined as

$$f_\theta(Motiv_i|X_i, Y^*_i, Z^*_{i,J+1}). \tag{15}$$

The observed-data likelihood function can be defined based on joint distribution of $(Y_i(0), Z^*_{i,J+1})$, as shown by Glynn (2013). Under the standard list experiment design, this joint distribution completely characterizes each respondent's type. To illustrate this, Table B.1 shows the identification of the population proportion of each respondent type with four control items.

Table B.1: Identification under the standard design with four control items

Response (Y_i)	Treatment group ($T_i = 1$)	Control group ($T_i = 0$)
5	(4,1)	
4	(3,1), (4,0)	(4,1), (4,0)
3	(2,1), (3,0)	(3,1), (3,0)
2	(1,1), (2,0)	(2,1), (2,0)
1	(0,1), (1,0)	(1,1), (1,0)
0	(0,0)	(0,1), (0,0)

Note: This table shows the potential respondent type $(Y_i(0), Z^*_{i,J+1})$ when the response and the treatment status is observed (Y_i, T_i). $Y_i(0)$ indicates the total number of affirmative answers for J control items and $Z^*_{i,J+1}$ represents the truthful response to the sensitive item. In our example, the total number of control items J is four.

For example, $(Y_i(0), Z^*_{i,J+1}) = (3,1)$ means that respondent i selects three control items affirmatively as well as the sensitive item. There exists a total of $2 \times (J+1) = 2 \times 5 = 10$ possible types of respondents. Specifically, the respondents in the control group who provides a response of 2 ($Y_i = 2$) is either of type (2,1) or (2,0). Similarly, the respondents in the treatment group who gives an answer of 2 is either of type (1,1) or (2,0). Note that the respondent types for $Y_i = 4$ and $Y_i = 0$ can be unambiguously determined. Therefore, the population proportion for $Y_i = 4$ and $Y_i = 0$ could also be determined. Because we assume random allocation of the treatment group and that the inclusion of sensitive item does not alter respondents' answer to the control items, the population proportion of certain type is independent of the treatment status. That is, the population proportion of the type (2,1) is the same for respondents in the treatment or control group. Then, the population proportion of each type could be derived from the observed data. Formally,

$$Pr(Y_i(0) = y, Z^*_{i,J+1} = 1) = Pr(Y_i \leq y | T_i = 0) - Pr(Y_i \leq y | T_i = 1) \tag{16}$$

$$Pr(Y_i(0) = y, Z^*_{i,J+1} = 0) = Pr(Y_i \leq y | T_i = 1) - Pr(Y_i \leq y - 1 | T_i = 0) \tag{17}$$

Given this setup, the observed-data likelihood function can be derived. In practice, the Expectation-Maximization (EM) algorithm is useful to find the maximum likelihood estimates of the parameters.

Appendix C. ADDITIONAL ANALYSES

Table C.1: Responses to the List Experiment

List A	Group		0	1	2	3	4	5	Mean Reported Items	Total
	Control	N	167	439	293	93	35	-	1.41	1,027
		%	16.26	42.75	28.53	9.06	3.41	-		100
	Treatment	N	143	359	326	139	30	10	1.59	1,007
		%	14.20	35.65	32.37	13.80	2.98	0.99		100
List B	Group		0	1	2	3	4	5	Mean Reported Items	Total
	Control	N	179	325	332	124	47	-	1.54	1,007
		%	17.78	32.27	32.97	12.31	4.67	-		100
	Treatment	N	141	276	356	184	57	13	1.78	1,027
		%	13.73	26.87	34.66	17.92	5.55	1.27		100

Note: This table shows the distribution of the responses (aggregated number of items that apply) to the list experiment. Only completed responses to both lists are considered. There are two lists included, and for each list there is a treatment and a control group. Both lists have 4 control statements. Thus, the answers of the control (treatment) groups from both lists have a range from 0 to 4 (5).

Figure C.1: Prevalence of EM from Other List Experiment Estimators

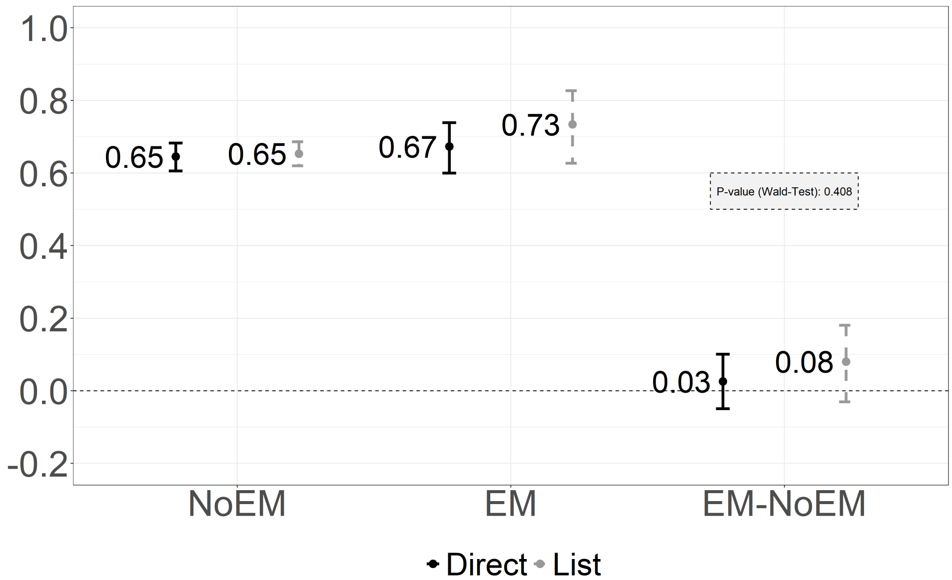
Note: This figure shows the regression results on list A and B using various estimators, which include (1) The linear model, (2) the nonlinear least squares estimation proposed in Imai (2011), (3) the maximum likelihood estimator corrected for top-biased error, and (4) the maximum likelihood estimator corrected for uniform error. Top-biased error occurs when a random subset of respondents chooses the maximal (ceiling) response value, regardless of their truthful response (Blair, Chou, and Imai 2019). Uniform error occurs when a subset of respondents choose their responses at random (Blair, Chou, and Imai 2019).

Table C.2: Firm and Executive Characteristics as Potential Moderators

	List		Direct	
	(1)	(2)	(3)	(4)
Treat	0.209	0.195		
List	0.164**	0.197***		
Small	0.207*	0.268*	0.036	0.046
Medium_Large	0.203	0.194	0.129	0.058
Partnerships	−0.415		0.375**	
Corporations	−0.212	0.188	0.227*	0.069
Discl	0.136	0.180	−0.037	−0.267
TechFirms	−0.002	−0.021	0.020	0.011
ManuFirms	−0.191	−0.228*	0.007	−0.037
Leverage_high	0.078	0.085	−0.064	−0.094
Tax_high	−0.020	−0.047	0.075	0.076
Male		0.048		0.025
University		0.066		−0.067
Treat x Small	0.241	0.333*		
Treat x Medium_Large	−0.061	0.066		
Treat x Partnerships	0.161			
Treat x Corporations	0.127	−0.015		
Treat x Discl	−0.284	−0.139		
Treat x TechFirms	−0.181	−0.078		
Treat x ManuFirms	0.334**	0.386**		
Treat x Leverage_high	0.017	0.075		
Treat x Tax_high	0.088	0.009		
Treat x Male		0.066		
Treat x University		−0.205		
Constant	1.390***	0.826**	−0.028	0.409
Observations	716	570	176	133
Adjusted R^2	0.030	0.044	−0.023	−0.037

Note: This table presents the results of the multivariate cross-sectional analyses. Column (1) and (2) relates to our list experiment approach, and column (3) and (4) relates to our direct questioning approach. Standard errors are clustered at the individual firm level. The dependent variable used in the first two models is the reported number of participants in the list experiment. It is a discrete variable that ranges from 0 to 5 for treatment lists and from 0 to 4 for control lists. *Treat* is a binary variable equal to one if response is from a treatment list, and zero otherwise. *List* is a binary variable equal to one for the first list and two for the second list. The size groups are defined according to the SME- EU Definition 2003/361: Very small ($\leq$ 9 employees & $\leq$ 2 mio. revenues), Small ($\leq$ 49 employees & $\leq$ 10 mio. revenues), Medium ($\leq$ 249 employees & $\leq$ 50 mio. revenues), Large ($>$ 249 employees or $>$ 50 mio. revenues). *Discl* is a binary variable indicating whether the firm is obliged to disclose financial statements in Federal Gazette. *TechFirms* and *ManuFirms* are binary variables equal to one for observations where the participants report working for a company in the technology or manufacturing industry. *Leverage_high* is a binary variable equal to one if the firm has higher than median leverage levels, calculated by the ratio of current liabilities to total assets. *Tax_high* is a binary variable equal to one if the local business tax rate of the firm's municipality is larger than median tax rates in the sample. *Male* is a binary variable equal to one if the participant identifies as male, and zero otherwise. *University* is a binary variable equal to one if the participant reports having completed a university degree, and zero otherwise.

Figure C.2: Disclosure Obligation by Experimental Groups and Earnings Management Practices



Note: This figure examines whether participants' self-reported existence of disclosure obligation is correlated with experimental group (LIST vs. DIRECT) and disclosed (DIRECT group) or predicted (LIST group) EM practices. For the LIST group, the predicted responses from the list experiments are used as predictors (EM(Yes/No)) for self-reported disclosure obligations. The dependent variable is a binary indicator equal to one if a participant reported that disclosure obligations apply to their company (Survey question: *Do disclosure obligations apply to your company (e.g. publication of the annual financial statements in the Federal Gazette)?*). Differences in outcomes between EM and non-EM firms are reported separately for the Direct and LIST groups. Additionally, Wald test *p*-values for differences between the two experimental groups are presented.

ONLINE APPENDIX: EXPERIMENTAL SETUP

Section C presents screenshots of the original survey questions as they appeared in the survey in German. English translations are also provided.

We survey firms using a double list experiment, a direct question, and an indirect question. Figure C.3 illustrates the survey design. Firms are randomly assigned to these three groups. Though list experiment is designed to reduce response bias, it comes at the cost of increased variance compared to direct questioning. To reduce the variance of a list experiment design, we design a double list experiment (DLE), which is first introduced by Miller (1984) as an alternative approach to balance the bias-variance trade-off. DLEs involve conducting two parallel list experiments simultaneously, using the average of the treatment effects from each experiment to estimate the prevalence of the sensitive item. Since each respondent receives the sensitive item only once in this design, the variance of the combined DLE estimate is expected to be reduced by half. Therefore, list experiment has double observations than the other groups.

Figure C.3: Experimental Design

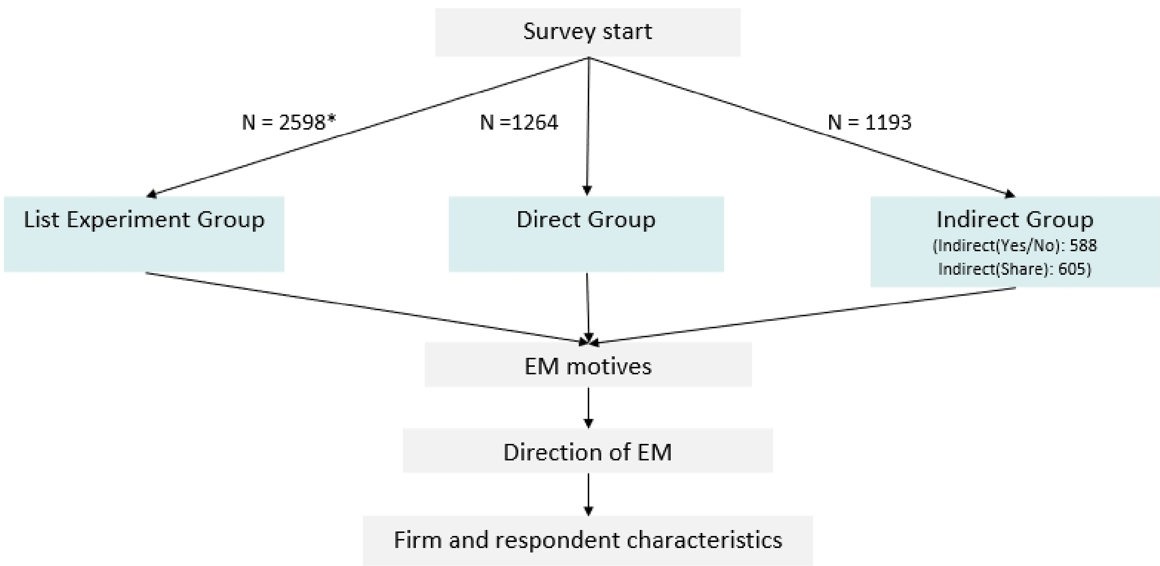


Figure C.4: Introductory Text

Bei der Gewinnermittlung kann die Geschäftsführung Einschätzungen vornehmen und den Gewinn in eine bestimmte Richtung steuern. Die legale Ausübung von Wahlrechten und Ermessensspielräumen kann zu informativerer Darstellung des finanziellen Erfolgs beitragen, jedoch auch von persönlichen Motiven geleitet sein und den Gewinn verzerren. Dies wird oft als Bilanzpolitik bezeichnet.

Im Folgenden möchten wir Ihre Sicht auf legale Bilanzpolitik erfahren.

Note. This figure presents the introductory text that each participant reads at the beginning of the experiment.

English translation:
When determining profit, management can make estimates and steer profit in a certain direction. The legal exercise of options and discretion can contribute to a more informative presentation of financial performance, but it can also be influenced by personal motives, which may distort profits. This is often referred to as earnings management.

In the following, we would like to understand your views on legal earnings management practices.

Figure C.5: Comprehension Question

Welche der folgenden Maßnahmen fallen Ihrer Ansicht nach unter legale Ausübung von Bilanzpolitik?

Hinweis: Bitte wählen Sie alle relevanten Optionen aus.

- Verzögerung oder Beschleunigung von Umsätzen und Ausgaben
- Über- oder Unterbewertung von Kosten
- Wahl zwischen degressiver oder linearer Abschreibung
- Überweisung von privatem auf das betriebliche Bankkonto
- Außerplanmäßige Abschreibung bei voraussichtlich dauernder Wertminderung
- Rückdatieren von Rechnungen
- Andere

English translation:
Which of the following accounting measures can be used for earnings management in compliance with the GAAP?

Answer options:
Acceleration or delay of revenue and expense recognition
Over/Understatement of cost (Cost allocation)
Choice of depreciation methods
Transfer from private to company bank account
Impairments (Unplanned depreciation when an unexpected permanent reduction in the value of an asset)
Backdating invoices
Others

Figure C.6: List Experiment Group 1: Treatment Group

Wie viele der folgenden Maßnahmen hat Ihr Unternehmen in den letzten Jahren ergriffen?

- Ausgelagerte Tätigkeiten wieder eingegliedert, um Lieferrisiken zu reduzieren.
- Wahlrechte und Ermessenspielräume bei der Gewinnermittlung ausgeübt, um den Unternehmensgewinn zum eigenen Nutzen zu verzerren.
- Tätigkeiten an externe Dienstleister ausgelagert, um Kosten zu reduzieren.
- Intensiv mit unseren Kunden kommuniziert, um die Umsätze zu verbessern.
- In Risikomanagement investiert, um interne Effizienz zu erhöhen.

Wir haben [ ] der genannten Maßnahmen ergriffen.

English translation:
How many of the following measures has your company taken in recent years?

We have reincorporated outsourced activities to reduce supply-chain risks.
We have taken advantage of available accounting choices and managers‘ discretion for own benefit to misrepresent earnings.
We have outsourced activities to third-party service providers to reduce costs.
We have intensively communicated with our customers to enhance our revenues.
We have invested in risk management to increase internal efficiency.

We have taken _ of the measures.

Figure C.7: List Experiment Group 1: Control Group

Wie viele der folgenden Maßnahmen hat Ihr Unternehmen in den letzten Jahren ergriffen?

- Geschäftseinheiten umstrukturiert, um Kosten zu reduzieren.
- Unser Projektmanagement angepasst, um interne Effizienz zu erhöhen.
- In Marketing von Produkten investiert, um die Umsätze zu verbessern.
- Die Zahl der Lieferanten reduziert, um Lieferrisiken zu reduzieren.

Wir haben [ ] der genannten Maßnahmen ergriffen.

English translation:
How many of the following measures has your company taken in recent years?

We have restructured business units to reduce costs.
We have adjusted our project management to increase internal efficiency.
We have invested in product marketing to improve sales.
We have reduced the number of suppliers to reduce delivery risks.

We have taken _ of the measures.

Figure C.8: List Experiment Group 2: Treatment Group

Wie viele der folgenden Maßnahmen hat Ihr Unternehmen in den letzten Jahren ergriffen?

- Wahlrechte und Ermessenspielräume bei der Gewinnermittlung ausgeübt, um den Unternehmensgewinn zum eigenen Nutzen zu verzerren.
- Die Zahl der Lieferanten reduziert, um Lieferrisiken zu reduzieren.
- Geschäftseinheiten umstrukturiert, um Kosten zu reduzieren.
- Unser Projektmanagement angepasst, um interne Effizienz zu erhöhen.
- In Marketing von Produkten investiert, um die Umsätze zu verbessern.

Wir haben [ ] der genannten Maßnahmen ergriffen.

English translation:
How many of the following measures has your company taken in recent years?

We have taken advantage of available accounting choices and managers‘ discretion for own benefit to misrepresent earnings.
We have reduced the number of suppliers to reduce delivery risks. We have restructured business units to reduce costs.
We have adjusted our project management to increase internal efficiency.
We have invested in product marketing to improve sales.

We have taken _ of the measures.

Figure C.9: List Experiment Group 2: Control Group

Wie viele der folgenden Maßnahmen hat Ihr Unternehmen in den letzten Jahren ergriffen?

- Tätigkeiten an externe Dienstleister ausgelagert, um Kosten zu reduzieren.
- Intensiv mit unseren Kunden kommuniziert, um die Umsätze zu verbessern.
- In Risikomanagement investiert, um interne Effizienz zu erhöhen.
- Ausgelagerte Tätigkeiten wieder eingegliedert, um Lieferrisiken zu reduzieren.

Wir haben [ ] der genannten Maßnahmen ergriffen.

English translation:
How many of the following measures has your company taken in recent years?

We have outsourced activities to third-party service providers to reduce costs.
We have intensively communicated with our customers to enhance our revenues.
We have invested in risk management to increase internal efficiency.
We have reincorporated outsourced activities to reduce supply-chain risks.

We have taken _ of the measures.

Figure C.10: Direct Question on Earnings Management

Hat Ihr Unternehmen in den letzten Jahren Wahlrechte und Ermessenspielräume bei der Gewinnermittlung ausgeübt, um den Unternehmensgewinn zum eigenen Nutzen zu verzerren?

Ja

Nein

English translation:
In past years, has your company taken advantage of accounting choices and discretion for own benefit to misrepresent earnings?

Answer options:
Yes
No

Figure C.11: Indirect Question(Share) on Earnings Management

Was schätzen Sie: Welcher Anteil der Unternehmen Ihrer Branche (in %) haben in den letzten Jahren Wahlrechte und Ermessenspielräume bei der Gewinnermittlung ausgeübt, um den Unternehmensgewinn zum eigenen Nutzen zu verzerren?

%

English translation:
In past years, what percentage of companies in your industry (in %) have taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?

Figure C.12: Indirect Question(Yes/No) on Earnings Management

Was schätzen Sie: Haben Unternehmen **Ihrer Branche** in den letzten Jahren Wahlrechte und Ermessenspielräume bei der Gewinnermittlung ausgeübt, um den Unternehmensgewinn zum eigenen Nutzen zu verzerren?

Ja

Nein

English translation:
In past years, have companies in your industry taken advantage of accounting choices and discretion for their own benefits to misrepresent earnings?

Figure C.13: Earnings Management Motives

Es gibt verschiedene Gründe, warum ein Unternehmen Wahlrechte und Ermessenspielräume bei der Gewinnermittlung ausübt, um den Unternehmensgewinn zum eigenen Nutzen zu verzerren. Welche der folgenden Gründe sind relevant für Unternehmen **Ihrer Branche**?

Steuern sparen

Den Investoren mehr Sicherheit über eine stabile Geschäftsentwicklung geben

Den Kunden/Lieferanten mehr Sicherheit über eine stabile Geschäftsentwicklung geben

Höhere Vergütung (z.B. Boni) für Mitarbeiter

Bessere Chancen bei Kreditgebern

Andere

English translation:
There are several potential reasons for why a company might take advantage of accounting choices and managers' discretion for its own benefit to misrepresent earnings. Which of the following reasons are most relevant for such practices in your industry?

Answer options:
Save taxes
Give the owners or investors more security about the stable development of business
Give customers/suppliers more certainty about the stable development of business
Give higher compensation for employees (e.g., higher bonuses)
Ensure better credit terms
Others

Figure C.14: Earnings Management Directions

Im Vergleich zum **wahren** Gewinn Ihres Unternehmens, in welche Richtung wirkten die ausgeübten Wahlrechte und Ermessensspielräume bei der Gewinnermittlung überwiegend?

	In den letzten Jahren	**Im aktuellen Geschäftsjahr**
Gewinnerhöhung	○	○
Neutral	○	○
Gewinnminderung	○	○
Keine Wahlrechte und Ermessensspielräume genutzt	○	○

English translation:
In which direction did the exercised accounting options and managers' discretion in the past years/ in the current year influence your company's earnings?

Answer options:
Increase in earnings
Neutral
Decrease in earnings
No accounting discretion exercised