Management Practices, ESG Incentives, and Corporate Communication in Organizations

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Stefanie Anna Ehmann
aus Friedrichshafen

Tübingen
2026

1. Betreuer:	Prof. Dr. Patrick Kampkötter
2. Betreuer:	Prof. Dr. Martin Ruf
Tag der mündlichen Prüfung:	25.06.2026
Dekanin/Dekan:	Prof. Dr. Taiga Brahm
	Prof. Dr. Dominik Papies
1. Gutachter:	Prof. Dr. Patrick Kampkötter
2. Gutachter:	Prof. Dr. Martin Ruf

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Organizations today operate in environments with fast-changing dynamics, pervasive uncertainty, growing pressure for transparency, and complex interdependencies. Demands and expectations from stakeholders inside and outside the organization, alongside crises, require organizations to constantly monitor and, if necessary, also adapt their strategies and organizational practices. These demands and expectations may, for example, be driven by global health crises like the Covid pandemic, climate-related developments, and concerns related to employee engagement and retention. Taken together, these factors create a challenging environment that organizations must navigate to remain effective and resilient. At its core, this dissertation examines how organizations build such resilience by adapting internal practices and strategic responses to various stakeholder demands. It highlights how management systems, incentive structures, and communication mechanisms shape employee behavior, decision-making, and external stakeholder perceptions. To address these themes, the dissertation is organized into four analytically distinct yet thematically connected empirical chapters.

First, Chapter 2 investigates the effect of structured management practices on employee-level work engagement. Specifically, the analysis focuses on three distinct and commonly used management practices: appraisal interviews, formal target setting, and performance-based pay. The findings highlight that each of the structured management practices examined positively affects employee work engagement. However, there is some context-specific variation in the magnitude of this relationship. Specifically, the effect, while still positive, is found to be slightly weaker in larger firms and in sectors such as IT, Communication & Other Services. Thus, although the magnitude of the positive effect slightly varies across organizations, these results underscore the substantial importance of formal performance management systems.

Extending the focus on structured management practices, Chapter 3 explores their implementation in family firms, highlighting how both ownership structures and management com-

position shape these practices and their effects. The study reveals that management, rather than ownership alone, plays a decisive role in shaping the adoption of formal management practices. Specifically, family-owned firms led by family members implement significantly fewer formal practices, while those managed by non-family managers adopt more structured practices, though still below the levels observed in firms with dispersed ownership. Despite these differences, employees in family-owned firms report levels of job satisfaction, procedural fairness, and perceived leadership quality that are comparable to those in dispersed-ownership firms when the firm is family-managed, and even higher when it is managed by non-family executives. These results suggest that informal coordination mechanisms embedded in family-firm cultures may partly substitute for formal management structures. However, their effectiveness is found to be limited for higher-skilled and supervisory roles. Thus, the study contributes to a more nuanced understanding of how formal (and informal) management practices shape organizational outcomes in family firms.

Shifting the level of analysis from employees to executives, Chapter 4 examines how environmental, social, and governance (ESG) considerations are incorporated into a key structured management practice: executive compensation systems. This approach reflects the growing attention to the role of climate-related considerations in corporate policies and incentive design. Specifically, the chapter investigates whether ESG metrics in compensation contracts are structured in a way that can plausibly generate incentives to improve ESG outcomes. The findings indicate that the necessary conditions for such incentive mechanisms are often violated in practice. ESG metrics are shown to either play a negligible role in pay calculations or to be largely insensitive to ESG performance. The chapter further illustrates that ESG contracting varies across executive roles and industries, with firms rarely tailoring ESG targets to the functional roles of executives and stronger, binding ESG metrics, i.e., metrics with an ex-ante specified weight, concentrated in sectors with direct financial exposure to ESG outcomes. Overall, this study contributes to the literature on corporate governance and sustainability by shedding light on current practices of ESG-linked executive incentives and their implications for ESG-related risk.

Finally, after examining formal management and incentive systems, the analysis turns to how organizations respond to acute and unforeseen challenges in real time. Specifically, Chapter 5 investigates managerial communication and market reactions during a particular health crisis, the Covid pandemic. The chapter shows that a firm's crisis exposure correlates positively with nonanswers, i.e., responses that evade investor queries. Furthermore, results also indicate that nonanswers continue to be negatively associated with capital market reaction, even in crisis conditions. This negative relationship appears to be even stronger when management otherwise presents an optimistic or opportunity-oriented narrative of the crisis. Thus, the

findings suggest that periods of uncertainty, such as the Covid pandemic, are associated with changes in communication, and that credibility plays a key role with respect to market reactions.

Collectively, these four chapters provide a detailed perspective on organizational practices and explore how organizations design, guide, and communicate behavior. By integrating evidence on internal management practices, executive ESG incentive design, and crisis-related communication, the dissertation deepens the understanding of how organizations coordinate internal actors and manage external expectations. Throughout the chapters, the formulation of the hypotheses relies on established theoretical arguments as well as documented empirical patterns in previous literature. Additionally, the diversity of settings allows for a comprehensive examination of how managerial practices and communication strategies interact with organizational characteristics and external pressures to shape individual and collective outcomes. Accordingly, the empirical analyses rely on multiple diverse datasets. Chapters 2 and 3 draw on longitudinal matched employer–employee data representative of German private-sector establishments with more than 50 employees, covering a range of industries and firm sizes, and including detailed ownership and/or management information. Chapter 4 explores a European dataset combining quantitative and qualitative data on executive compensation, with a particular focus on ESG metrics and associated ESG-related risk. Finally, Chapter 5 draws on textual data obtained from transcriptions of conference calls from the US and Europe, providing insights into international managerial communication during crises.

The thesis also pays close attention to practical applicability, ensuring that the findings are informative for managers seeking to apply relevant mechanisms. Firstly, as shown in Chapter 2, managers aiming to foster employee engagement benefit from the consistent use of structured performance management practices. All examined structured management practices have positive effects on employee engagement, indicating their broad value across organizational settings. Nevertheless, it needs to be noted that the strength of these positive effects differs slightly across firm sizes and industries. Secondly, Chapter 3 highlights that family-owned firms can actively leverage their distinctive culture to enhance intrinsic employee motivation and partly compensate for fewer formal management practices. At the same time, integrating non-family managers into leadership roles can strengthen work quality and help balance informal and formal structures. Importantly, firms should recognize that higher-skilled and supervisory employees typically require more formalized management systems to foster engagement. Thirdly, Chapter 4 suggests that weak, performance-insensitive ESG metrics that carry negligible weight expose firms to concerns about greenwashing and weaken the credibility of sustainability efforts. Stronger, tailored ESG measures that carry material weight, particularly where environmental risks are high, are better suited to provide meaningful commitments. With rising stakeholder scrutiny, improving the structure of ESG-linked pay, which currently

seems mostly symbolic, may become increasingly critical. Lastly, Chapter 5 highlights that crisis situations may heighten investor sensitivity to the clarity of managerial communication. When firms appear to be coping relatively well, nonanswers may be viewed particularly critically, suggesting that transparent and straightforward communication is especially valuable in such contexts. Thus, in short, the results highlight that effective management relies on a dynamic balance between formal structures, informal organizational features, and transparent communication, with their relative importance contingent on organizational and environmental conditions.

Chapter 2

Performance Management and Work Engagement - New Evidence Using Longitudinal Data*

The value and the impact of performance management systems on employee-level and firm-level outcomes is a widely discussed topic in both academia and practice. In this paper, we investigate the impact of performance management and evaluation practices (PMEPs), such as performance appraisal interviews, formal target agreements and performance-related pay schemes, on employee-level work engagement. Building on management theories, we argue that PMEPs relate positively to employee-level work engagement. By applying a contingency perspective, we further explore the context dependence of this relationship with respect to firm size, industry, and ownership type. Using a longitudinal, linked employer-employee data set in four waves; and by applying various fixed effects regressions to account for unobserved heterogeneity, our results imply a positive and statistically significant impact of PMEPs on work engagement. We further find that the relationship between PMEPs and work engagement is smaller for larger firms and for firms in the IT, Communication & Other Services sector.

Key Words: Performance Management; Work Engagement; Performance Evaluation; Target Agreements; Performance Appraisals; Performance Feedback; Performance-related Pay

JEL Classification: D23, J01, J33, M41, M52

*This chapter is based on joint work with Patrick Kampkötter, Patrick Maier, and Philip Yang and published in Management Accounting Research (see Ehmann et al. (2024)).

2.1 Introduction

The value and the impact of performance management systems (PMSs) on employee-level and firm-level outcomes is a widely discussed topic in both academia and practice (Otley (1999); Ferreira and Otley (2009); Franco-Santos et al. (2012); Aguinis et al. (2013); Armstrong (2022)). Research on this topic encompasses a broad array of definitions and conceptions of PMSs. Otley (1999) and Ferreira and Otley (2009) propose a multi-dimensional performance management framework that consists of a variety of key aspects. In the World Management Survey, for instance, a large-scale management survey developed by Bloom and Van Reenen (2007, 2010), information on the use of multiple performance management practices from the areas of monitoring, targets, and incentives is elicited from industrial-firm managers. Although some research on PMSs suggest positive management quality and firm-performance implications (Bloom et al. (2019)), the overall empirical evidence concerning its general effectiveness is mixed (e.g., Franco-Santos et al. (2012)).

In this study, we focus on performance management and evaluation practices (PMEPs) as specific tools of a PMS. PMEPs are a system of management practices of different degrees of formality with the aim of incentivizing, evaluating, and rewarding past and future performance of employees in alignment with organizational objectives. We particularly focus on three specific PMEPs: (1) performance appraisal interviews (PAIs, also called performance feedback interviews or performance reviews (Armstrong (2022)), i.e., pre-scheduled conversations between supervisors and subordinates, during which the latter's past performance is assessed, (future) potential and training needs are identified, and career and personal development plans are discussed, (2) formal target agreements (TAs), i.e., written agreements specifying the targets the employee needs to meet over the course of the year, and (3) performance-related pay (PRP) schemes, i.e., compensation structures that make part of the remuneration dependent on the performance of the employee, her team/division, and the firm.[1]

We study the effect of PMEPs on employee work engagement, which is "a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption" (Schaufeli et al. (2002, p. 74)). In previous research, work engagement has shown to be a key factor influencing individual outcomes such as employee effort (Presslee et al. (2023)), job performance (Salanova et al. (2005); Halbesleben (2010); Rich et al. (2010); Christian et al. (2011)), and innovativeness at work (Schaufeli et al. (2006b)). Furthermore, it has also found to be a relevant

[1]Representative evidence from private-sector firms in Germany (data used in this paper) shows that PMEPs are actually of high practical relevance. In 2018, 85% of firms reported using PAIs, while 81% reported using formal TAs and 74% PRP schemes. Taking the employee perspective, 64% of all employees working in these firms reported being covered by a PAI, 42% reported having both a PAI and a formal TA, and 62% stated that they received PRP.

antecedent to organizational outcomes such as customer loyalty (Salanova et al. (2005)), firm productivity (Harter et al. (2002)), and financial returns (Xanthopoulou et al. (2009)).

Building on key overarching management theories that are also applied in the context of management accounting (goal-setting theory, organizational justice theory, expectancy theory, and principal-agent theory)[2], we argue that PMEPs relate positively to work engagement due to a stronger feedback link, enhanced goal and target clarity, the specificity and development of performance-related rewards, and the closer alignment of interests between employees and the firm. Furthermore, we explore the context dependence of this relationship by applying a contingency perspective. Because the extent to which PMSs and PMEPs are used varies among firms of different sizes, industries, and ownership patterns, we analyze whether the relationship between PMEPs and work engagement is moderated by these firm contingencies.

We then test these relationships empirically using data from the Linked Personnel Panel (LPP), a longitudinal, matched employer-employee data set (Kampkötter et al. (2016)). This data set is representative for German private-sector establishments with more than 50 employees and their respective workforce.[3] The LPP combines an employer survey (between 769 and 1,219 firms per wave) with an employee survey, consisting of a random sample of employees working within the surveyed firms (roughly between 6,500 and 7,500 individuals per wave). We observe four waves - 2012, 2014, 2016/17 and 2018/19 - of the LPP resulting in a total of around 16,500 employee-firm-year observations.

Using a multitude of fixed effect regressions, we find a positive and statistically significant overall effect of PMEPs on work engagement, even controlling for time-constant heterogeneity at the firm (e.g., leadership culture) or individual (e.g., ability or talent) level. We also find positive engagement implications for each of the three specific PMEPs. However, the positive effect is stronger for PAIs and TAs than it is for PRP schemes. With respect to contingencies, our empirical results provide tentative evidence that this positive direct effect decreases slightly for larger firms and firms in the IT, Communication, & Other Services sector. Despite these heterogeneities our results suggest that the overall positive effect of PMEPs on work engagement persists across the analyzed contingencies. Hence, we conclude that the usefulness of PMEPs as a tool to increase work engagement applies to firms of different sizes, industries, and ownership patterns.

[2]For reviews of the theories in the context of accounting research, see, for instance, Birnberg et al. (2006), Lambert (2006) or Wibbeke and Lachmann (2020).

[3]An establishment denotes a regionally and economically separate unit. Two types of firms exist: 1) multiple-establishment firms, which comprise several establishments, and 2) single-establishment firms, where firm and establishment denote the same entity. For the purpose of readability, we only use the term "firm" throughout the paper. However, in all regressions, we control for these two types of firms.

We contribute to the literature in several ways. First, our study complements existing research on the value of PMSs for employee-level outcomes by analyzing the implications of PMEPs, a combination of three established performance management practices – performance appraisal interviews, written target agreements, and performance-related pay schemes (see, for instance, Otley (1999) and Ferreira and Otley (2009)). Building on management theories, we then develop a positive link between PMEPs and employee-level work engagement. Empirically, we show that PMEPs relate positively to work engagement in a larger-scale linked employer-employee data set. However, this positive effect is of different magnitude when distinguishing the effect of the specific PMEPs. This empirical analysis benefits from our linked employer-employee data structure, since usually firm-level information on the design and coverage of PMSs combined with data on employee-level outcomes are rarely available, especially with regard to middle management and across firms (Indjejikian and Nanda (2002); Dekker et al. (2012); Qiu et al. (2022)).

Second, our paper complements existing literature with respect to the incidence and effectiveness of PMSs in a wide range of firm environments. Our descriptive results show that PMEPs are used differently across firms of various sizes, industries, and ownership patterns. For this reason, we apply a contingency perspective and explore whether the effect of PMEPs on work engagement differs with respect to these firm contexts. Our empirical results show that the effect of PMEPs on work engagement is overall positive. Nonetheless, our findings also point to tentative differences in the magnitude of the effect when inspecting various firm sizes and industries. Thus, our study emphasizes the need to consider contextual factors on the firm and individual level when studying the implications of PMSs and PMEPs (Abdel-Kader and Luther (2008); Sholihin et al. (2011); Messner (2016)).

Finally, we also complement previous literature from a methodological perspective. Our longitudinal data structure allows for an estimation of a variety of different regression specifications employing a diverse set of fixed effects such as year and firm fixed effects. Firm fixed effects, in particular, provide relevant insights to previous between-firm studies, as they allow for within-firm comparisons of a larger number of firms. Thus, we are able to account for unobserved time-constant heterogeneity on the firm level that likely determines both the level of work engagement in a given firm and the existence or extent of PMEPs. With this empirical design, we aim at addressing potential sources of endogeneity (Latham and Locke (2007)) and at mitigating "statistical artefacts, such as sampling error" (Derfuss (2009), p. 203), which might explain heterogeneous results so far.

The remainder of the study is structured as follows: In Section 2.2, we define our main con-

structs and discuss theoretical considerations. Section 2.3 describes the data and the empirical strategy. In Section 2.4, we present our results, and Section 2.5 concludes.

2.2 PMEPs and Work Engagement

2.2.1 Definition of Main Constructs

There are many different definitions of PMSs in previous research. While some of them define performance management in a rather narrow sense (e.g., focusing on performance measurement only), others define it more broadly (see, for instance, Ferreira and Otley (2009); Broadbent and Laughlin (2009); Franco-Santos et al. (2012); Abernethy et al. (2021)). To date, no universal definition of the term has emerged in the literature due to its complex nature and the variety in perspectives. In some cases, research emphasizes variations in distinct features of PMSs (Cheng et al. (2007); Lau (2011); Franco-Santos et al. (2012)). Other definitions attach more importance to variations concerning the purpose of the systems (Hall (2008); Franco-Santos et al. (2012); Micheli and Mari (2014)). Building on the established (extended) performance management framework of Otley (1999) and Ferreira and Otley (2009), we study the implications of PMEPs as a system of management practices of different degrees of formality with the aim of incentivizing, evaluating, and rewarding past and future performance of employees in alignment with organizational objectives. Our research focuses on three specific practices: 1) annual, pre-scheduled performance appraisal interviews, 2) formal target-setting between supervisors and subordinates, and 3) performance-related compensation schemes.

PAIs are pre-scheduled, regularly-held appointments in which supervisors meet up with their subordinates (Burke et al. (1978); Cederblom (1982)). These PAIs, which comprise both an evaluative and developmental dimension (Cummings and Schwab (1978); Asmuß (2008); Boswell and Boudreau (2002)), serve a variety of purposes. Firstly, they enable supervisors to evaluate their employees' current performance and to provide them with performance feedback on their past achievements (Kay and Meyer (1965); Ivancevich (1982); Pitkänen and Lukka (2011); Jacobs et al. (2014); Pichler et al. (2020)). Additionally, these meetings offer the opportunity to discuss potential training needs and career ambitions of employees and thus, to assess the future potential of subordinates (Gordon and Stewart (2009); Meinecke et al. (2017); Armstrong (2022)). These sessions can either have a formal or a non-formal character. In this paper, we focus on pre-scheduled and, thus, formal PAIs. The effectiveness of PAIs has shown to depend on factors such as, for instance, the level of support and criticism of supervisors (Burke et al. (1978); Cederblom (1982); Silverman and Wexley (1984)).

Written TAs are often negotiated during annual PAIs (Burke et al. (1978); Cederblom

(1982)). They specify and formalize the goals that employees are expected to achieve during a prescribed time period. Thus, TAs can contain transparent quantitative as well as qualitative goals for employees that should foster the understanding of their job-specific tasks and help them to make sense of their superiors' expectations (Kampkötter et al. (2017); Martin and Thomas (2022)). Whereas much of the prior literature focuses on unique aspects of the target setting process (e.g., Indjejikian et al. (2014); Ioannou et al. (2016); Voußem et al. (2016); Feichter et al. (2018); Yitzhaky et al. (2021); Martin and Thomas (2022); Maas and Shi (2023); Shang et al. (2023); Feichter (2023)), in this paper we focus on the implications of the introduction of formally-specified targets as a performance management instrument in firms generally.

PRP schemes cover all remuneration systems, in which the extent of variable pay of employees depends on individual, divisional, and organizational performance outcomes (Murphy (2000); Bloom and Van Reenen (2010); Luft (2016); Unger et al. (2020); Bandiera et al. (2021)). They can be structured along various elements (Liu and Leitch (2013)). Firstly, they may contain some sort of performance metric, usually subjective or objective indicators or a combination of these two (Murphy (2000); Lazear (2000); Manthei et al. (2023); Wood et al. (2023)). Secondly, a reference standard (Murphy (2000); Indjejikian and Nanda (2002); Wood et al. (2023)) can be used to evaluate performance, which is typically linked to various different internal or external factors (Gibbons and Murphy (1990); Murphy (2000); Holzhacker et al. (2019); Yatsenko (2022)). Thirdly, the relationship between performance and pay is of significance (Murphy (2000); Manthei et al. (2023)). It could be specified in a way such that individual and organizational performance outcomes are aligned via an employee's compensation contract.

In this paper, we relate PMEPs to work engagement of employees. Work engagement is grounded in the literature on management and organizational psychology and reflects work-related motivation, enthusiasm, and personal investment in work and job-related tasks (Kahn (1990); Schaufeli and Bakker (2003); Bakker and Leiter (2010); Bakker (2017)). Schaufeli et al. (2002) define work engagement as "a positive, fulfilling, work-related state of mind" (p. 74). According to Schaufeli and Bakker (2004) and Schaufeli et al. (2002), work engagement consists of three different dimensions: vigor, dedication, and absorption. Employees scoring high in the vigor dimension can be characterized by high energy levels, resilience, and perseverance. Thus, vigor captures the degree of activation of an employee. Strong dedication illustrates, amongst others, employees feeling enthusiastic and proud of their work. Hence, this dimension depicts how strongly employees identify with their work. An advanced level in the absorption dimension relates to employees being fully immersed in their work and losing track of time over work. This is sometimes also described as a particular kind of flow state in which employees are very focused and fixated on their work.

2.2.2 The Relationship between PMEPs and Work Engagement

Several management theories, which have also been applied in the management accounting context, are informative about the relationship between PMEPs and work engagement (for an extensive review, see Birnberg et al. (2006), Lambert (2006) or Wibbeke and Lachmann (2020)).

Firstly, goal-setting theory, put forward by Locke and Latham (1990), states that more specific and concise goals are associated with a higher level of performance and motivation than general and vague "do your best" goals. Given that performance is controllable, an increased goal specificity focuses the individuals' attention on what should be accomplished and, thus, is more effective, particularly in the presence of feedback systems (Locke and Latham (1990, 2002); Lunenburg (2011); Locke and Latham (2019)).

Secondly, organizational justice theory, going back to Greenberg (1987), emphasizes the motivational potential of, amongst others, procedural justice. Procedural justice refers to the fairness perceptions of individuals concerning the processes in which outcomes are established. Generally, organizational justice theory assumes that individuals experiencing higher levels of procedural justice will also react in a more positive way, thus evoking better outcomes. Processes in which individuals are given the opportunity to actively participate in and voice their thoughts are typically perceived as fairer (Greenberg (1987, 1990)).

Thirdly, expectancy theory postulates that the degree of individual motivation depends on three aspects, namely their belief that their performance can be influenced by their effort level (expectancy), that their performance is rewarded with corresponding outcomes (instrumentality), and that those outcomes are valuable to them (valence). These three aspects are connected multiplicatively and, thus, all together determine the level of individual motivation (Vroom (1964); Ferris (1977))[4].

Finally, principal-agent theory emphasizes the typical divergence between the interests of employees and those of their firm, particularly in cases where employee effort is non-contractible and moral hazard problems arise. To address these issues, management practices are implemented to better align these diverging interests and through that motivate their workforce to act on behalf of their shareholders (Prendergast (1999)).

[4]There are variations in expectancy theory that, for instance, forego the differentiation between *performance* and *outcome* and thus, sum up expectancy and instrumentality or focus on an additive connection between these elements (Ferris (1977); Kominis and Emmanuel (2007)).

2.2.2.1 Performance Appraisal Interviews

PAIs provide a structured framework for supervisors to communicate formal performance feedback concerning their subordinates' actual and potential contribution to firm value (Cederblom (1982); Pichler et al. (2020)). According to goal-setting theory, performance feedback is an essential factor that ties the specificity of goals to the motivation of employees (Locke and Latham (2002); Budworth et al. (2015); Locke and Latham (2019)). Beyond this contextual role, feedback can also directly affect the motivation of employees as it informs them about their progress made so far and gives them insights into potential future adjustments needed concerning the level or the direction of their effort (Bakker and Demerouti (2007); Pitkänen and Lukka (2011); Gruman and Saks (2011)). Thus, PAIs contribute to greater clarity on the side of employees concerning their goals and corresponding achievement strategies. Because of the feedback component inherent in PAIs, these can hold work engagement-enhancing implications. Also, during PAIs, supervisors and their subordinates typically engage in a two-way communication about performance, potential, and future development to which both parties can contribute (Cawley et al. (1998); Sumelius et al. (2014)). Ideally, employees are thereby given the opportunity to participate in the process to a certain extent and make their voice heard (Holbrook (2002); Thurston and McNall (2010); Pichler (2012); Jacobs et al. (2014)). Taking into account the key concepts of organizational justice theory, the setting of PAIs can be associated with a comparatively strong perception of procedural fairness. Thus, PAIs should provoke positive reactions of employees' and foster their involvement.

Relatedly, formal PAIs may also foster an employee's feeling of procedural justice, as the higher degree of formalization can increase the perceptions of employees that the rules and procedures to make decisions in their company are more fair, as compared to situations where supervisors follow their ad hoc judgments and strongly differ in the way they conduct more informal conversations in the absence of formalized, organizational procedures (Folger et al. (1992); Gibbs et al. (2004); Lau and Moser (2008); Hartmann and Slapničar (2012)).

2.2.2.2 Formal Target Agreements

TAs consist of an employee's objectives that are formalized and put into writing. Thus, they represent a valuable resource for employees which specifies their individual goals and provides them with guidance concerning the direction of their effort. As outlined above, goal specificity is a key determinant in goal setting theory (Locke and Latham (1990, 2002)). Therefore, formal TAs with their - usually by design - higher degrees of goal specificity are particularly suited to positively impact employee motivation and thus engagement (Dekker et al. (2012); Kampkötter et al. (2017); Feichter et al. (2018)).

Furthermore, the implications of the already described common features of TAs can even be more extensive. Expectancy theory assumes that individuals are more motivated when instrumentality is stronger, thus, when they are convinced that superior, individual performance will be rewarded by higher outcomes. In general, TAs explicitly specify the valued achievements ex ante, while at the same time, employees are more likely to be able to rely on the previously agreed objectives. Therefore, the use of TAs should strengthen the aspect of instrumentality and thus enhance employee engagement (Ferris (1977); Kominis and Emmanuel (2007)).

2.2.2.3 Performance-related Pay

In case of performance-related pay schemes, principal-agent theory delivers arguments for an increased level of employee engagement. As performance-related pay schemes typically tie individual bonuses to individual, divisional, and organizational outcomes, this closer alignment of firm and individual interests and their increased responsibility towards results should foster employees' effort, motivation, and engagement (Prendergast (1999); Lazear (2000)).

Additionally, the implications of expectancy theory need to be considered. The motivation of individuals also is higher, the more value they attribute to the potential outcomes they can reach. With performance-related pay, employees benefit monetarily from superior performance. Following expectancy theory and assuming that employees typically ascribe at least some value to financial compensation, the use of the PMEP element performance pay should be connected to higher levels of employee engagement (Ferris (1977); Kominis and Emmanuel (2007)).

Building on the previous discussion, all three PMEPs should thus be positively associated with work engagement.

2.2.3 A Contingency Perspective on the Link between PMEPs and Work Engagement

The arguments outlined above suggest a generally positive influence of PMEPs on work engagement (i.e., a positive average effect). Yet, there is a large body of literature, also in the field of management accounting, that emphasizes the need to consider the context in which practices are embedded (Otley (1980); Abdel-Kader and Luther (2008); Dekker et al. (2012); Messner (2016); Otley (2016)). Based on contingency theory, this strand of research claims that management systems are not equally beneficial in all firms. Rather, in order to be effective,

management systems have to fit the particular features in which they are used.

Some research is already beginning to record differences in antecedents of PMSs, i.e., in the distribution of their incidence across various firm sizes, industries, and ownership patterns. With respect to firm size, different studies have shown that larger firms make use of performance appraisals (Grund and Sliwka (2009)) and management practices (Bloom and Van Reenen (2007, 2010)) significantly more often compared to smaller firms. Regarding industry-specific differences, Grund and Sliwka (2009) describe that performance appraisals are rather prevalent in the financial and corporate-services industries, but are only rarely implemented in the construction sector. Concerning the ownership status and, in particular, the particularities of family firms, prior literature finds tentative differences in terms of the extent of performance management coverage (Bloom and Van Reenen (2007, 2010); Broszeit et al. (2019)).

The difference in the distribution of PMSs across firms described in the literature could have many reasons. Among other things, a higher prevalence of PMSs may be the result of differences in the firms' resources and capabilities. Alternatively, a firm could purposely choose (not) to implement PMSs if their effectiveness varies across organizations. Therefore, in an exploratory analysis, we will investigate potential heterogeneities regarding the effectiveness of PMEPs for work engagement in different contextual settings. We hereby explore the moderating role of firm size, industry, and ownership status in particular.

2.3 Data

2.3.1 Dependent and Independent Variables

In order to examine the nexus between PMEPs and work engagement, we use the Linked Personnel Panel (LPP), a longitudinal, matched employer-employee data set.[5] The LPP is representative of German private-sector establishments with more than 50 employees covered by social security (for a detailed description of the design of the data set and the applied constructs, see Kampkötter et al. (2016)). Response rates are comparatively high, amounting to, for instance, 75% for the employer survey and 34% for the employee survey for the fourth wave. Overall, we find no significant selectivity effects on panel participation. Surveyed establishments are randomly drawn from the IAB establishment panel, an annual survey of nearly 16,000 German establishments. In order to consider establishments of all sectors and

[5]The data set is open to any researcher and is available via the Research Data Centre (FDZ) of the German Federal Employment Agency at the Institute for Employment Research (IAB). The DOI is: 10.5164/IAB.LPP1819.de.en.v1. For more details, see Haylock and Kampkötter (2019) and Ruf et al. (2020).

size classes, the sample is randomly drawn in a disproportionate and stratified manner by establishment size, federal state, and business sector.

We can make use of four waves – 2012, 2014, 2016/17 and 2018/19 – of the LPP employer-employee data set. The employer survey covers between 769 and 1,219 firms per wave. Firm managers provided information on management practices and other firm characteristics. From these firms, a random sample of employees working within the surveyed firms (roughly between 6,500 and 7,500 individuals per wave) were interviewed at home via telephone (CATI) or web interface (CAWI) about job characteristics and perceptions, personal characteristics, attitudes towards their organization, and behavioral variables. This feature of the data enables us to examine the link between the presence of PMEPs and work engagement on the individual level, while simultaneously being able to control for organizational characteristics at firm level. Furthermore, the longitudinal structure of the data enables us to employ panel data methods, allowing us to move closer towards causality.

Dependent variable. We operationalize work engagement via the widely used, internationally validated nine-item work engagement scale UWES-9 (Schaufeli et al. (2006a)). Respondents are asked to indicate to which extent they agree with nine statements regarding their job on a five-point Likert scale, such as the following: *At my work, I feel bursting with energy.*[6] Higher scores on this scale indicate higher levels of work engagement and vice versa. The reported scores of every single item of the work engagement scale are added up and divided by 9, with the aggregated *work engagement score* representing an equally weighted average with values between 1 and 5 (Sonnentag (2003)). We further standardize this work engagement score to zero mean and unit variance prior to entering the regressions. Cronbach's Alpha is 0.915, indicating a high degree of internal consistency of this construct. In 2019, the latest year included in our data, the mean (unstandardized) engagement score is 3.49, while the median is 3.56. Therefore, we observe a higher probability mass at larger values of work engagement, indicating that employees in our sample are, on average, considerably engaged. The distribution is also very stable over time.

Independent variables. Our analysis of PMEPs is based on three items from the LPP employee survey and is in line with recent large-scale management surveys such as the World Management Survey or the Management and Organizational Practices Survey (Bloom and Van Reenen (2007, 2010); Bender et al. (2018); Bloom et al. (2019)). We first measure the presence of a PAI via the following question: *"Did you have an appraisal interview with your direct supervisor last year about issues related to, for instance, your professional development or personnel evaluation? Please consider only appraisal interviews for which an appointment*

[6] A complete list of the items used can be found in Appendix A.1.

was made in advance." This question is then used as a filter question for the item measuring the incidence of a formal TA, implying that an employee can only be covered by a TA if she is also covered by a PAI. It is important to stress the formal character of PAIs, since respondents should only consider meetings with their supervisor for which a formal appointment was made. The item measuring the presence of a TA (in conjunction with a PAI) is based on the following question: *"Has your supervisor agreed on written targets with you during the PAI?"*. To measure the existence of a PRP scheme, we make use of the following item: *"Do you receive performance-related pay in addition to your fixed base salary? This refers, for example, to bonuses, spot bonuses, commissions, etc., regardless of whether they are contractually agreed or not."*

We generate several indicators of PMEPs. First, we construct a composite score variable, *PMEP score*, adding up the total number of the three specific performance management practices by which an employee i in firm j is covered at date t. This variable ranges from 0 to 3, with zero stating that an employee is not covered by any of these practices, and three stating that the employee is simultaneously covered by all three performance management practices considered. The mean of our PMEP score amounts to 1.47 and the median to 1. In total, our data comprise 16,498 non-missing employee-firm-year observations.

Since this PMEP score assumes a linear relationship with work engagement when increasing the number of PMEPs, we also use an alternative, non-linear measure. In detail, we subdivide the PMEP score measure into three dummy variables indicating whether employees are simultaneously covered by one, two, or three practices, regardless of the exact composition of PMEPs. For all three dummies, employees not covered by *any* of these three practices constitute the reference category. 29% of employee-firm-year observations are covered by a single practice, 19% by two practices, and 27% by three practices. Third, we include one distinct dummy variable for each practice in order to analyze the isolated effects of the different types of PMEPs.

Importantly, the data allow us to account for a rich set of control variables on the firm and employee level. Firm-level controls comprise industry (5 categories), region (north, east, south, west), firm size (4 dummies), a set of dummies capturing majority ownership of the firm (e.g., family firm, dispersed ownership, financial investor), and a dummy capturing a single-establishment firm vs. a firm that operates multiple establishments (e.g., multiple production sites). Employee-level controls include gender (with 1 being female and 0 male), age (4 dummies), leadership position (0/1), full-time position (0/1), white-collar employee (0/1), monthly net income, type of employment contract (fixed term/permanent), permanent relationship (0/1), highest level of school education (6 dummies), highest level of further education (7 dummies), household size, and CAWI survey method (0/1).

Empirical strategy. We estimate a variety of different OLS regression specifications employing various fixed effects in order to be able to make more causal statements. In our main analysis, we regress work engagement on different measures of PMEPs, including all controls on the firm and individual level as well as year fixed effects to make between-firm comparisons. Hence, our baseline specification can be expressed by the following equation:

$$y_{ijt} = \beta \times PMEP_{ijt} + \delta \times Controls_{it} + \eta \times Controls_{jt} + \gamma_t + \epsilon_{ijt} \tag{2.1}$$

The dependent variable y_{ijt} is standardized work engagement of individual i in firm j in time t, $PMEP_{ijt}$ is our key independent variable of interest (measured by either the PMEP score, the three PMEP dummies or the isolated practices), $Controls_{it}$ and $Controls_{jt}$ are vectors of time-varying controls on the employee (employee i) and firm level (firm j), while γ_t denotes year fixed effects and ϵ_{ijt} the error term.

We complement these regressions with a firm fixed effects specification in order to make additional within-firm comparisons. We account for unobserved time-constant heterogeneity on the firm level that determines both the level of work engagement in a given organization and the existence or extent of performance management. In supplementary analyses, we also run specifications including individual fixed effects and firm-year fixed effects. This enables us to explore within-person variation and to take into account unobserved firm-year specific characteristics or trends. In all regressions, standard errors are clustered on the firm level. A descriptive overview of our main dependent and independent variables as well as on all firm-level and employee-level controls and contextual factors is provided in Tables A.1, A.2, and A.3 in Appendix A.2.

2.3.2 The Distribution of PMEPs

The data set examined in this paper comprises a broad set of contextual variables. These variables are not only beneficial for control purposes in analyses. Beyond that, they provide the opportunity to gather insights into the use of PMEPs over time and across different firm characteristics. Table 2.1 first provides an overview of the average value of the PMEP score across time and firm-level contextual factors. Additionally, it shows the respective percentage of employees being covered by the individual PMEPs; thus, by a PAI, a TA, or a PRP scheme across these variables.

Table 2.1: Distribution of PMEPs

	Variables	PMEP Score	PAI[a]	TA[a]	PRP[a]
Year	Wave 2012	1.37	45.76	31.84	59.26
	Wave 2014	1.39	48.12	34.10	56.50
	Wave 2016/2017	1.50	55.63	36.69	58.05
	Wave 2018/2019	1.76	66.66	43.55	66.21
Firm Size	Fewer Than 100 Employees	0.97	30.85	17.29	48.76
	100 - 249 Employees	1.15	39.73	25.74	49.23
	250 - 499 Employees	1.47	52.64	36.60	57.77
	More Than 500 Employees	1.81	65.63	46.22	69.43
Industry	Manufacturing	1.48	51.86	37.76	58.59
	Metal, Electrical, & Automotive	1.57	54.06	35.90	67.52
	Retail, Logistics, & Media	1.22	43.96	26.66	51.42
	Company-Related & Financial Services	1.45	53.69	35.91	55.05
	IT, Communication, & Other Services	1.27	52.87	37.10	36.93
Ownership	Family, Founder	1.26	43.35	27.80	54.82
	Management	1.38	49.82	33.63	54.77
	Financial Investor	1.76	62.00	45.80	67.86
	Dispersed Ownership/Widely held	1.82	65.28	48.37	68.60
	Government/Public Sector	1.90	72.73	51.43	65.45
	Other	1.63	59.24	40.35	63.40

Notes: All numbers represent averages. [a] In %. PMEP Score = Sum of performance management and evaluation practices per individual; PAI = Performance appraisal interview; TA = Written target agreement; PRP = Performance-related pay scheme.

The results show that, on average, the use of PMEPs increases over time. While the mean of the PMEP score amounts to 1.37 in 2012, it increases to 1.76 PMEPs per employee in 2018/2019. While the increasing trend is present for each PMEP individually, it is especially pronounced for the usage of PAIs and TAs, which go up by 46% and 37%, respectively.

On the firm level, a considerable heterogeneity exists with respect to firm size. On average, the incidence of PMEPs strongly increases with a firm's size. Individuals working in firms with more than 500 employees are covered, on average, by 1.81 PMEPs, whereas individuals working in organizations with fewer than 100 employees benefit from approximately half that amount. Again, this difference is especially noticeable for PAIs and TAs. This variation in PMEP incidence is not as pronounced when evaluating industry or ownership patterns. Industry-wise, merely the comparatively low implementation of TAs in Retail, Logistics, & Media and of variable pay in the IT, Communication, & Other Services industry is striking. These numbers drive the relatively lower average PMEP score compared to other industries. When looking at different ownership structures, PMEPs are more frequently used in firms with dispersed ownership

and financial investors as majority owners, as compared to family firms. The divergence is once more driven by PAIs and TAs.[7]

2.4 Results

2.4.1 The Impact of PMEPs on Work Engagement

In Table 2.2, we estimate the average impact of the PMEP variables (score and dummy variables) on work engagement. Column 1 is based on a regression of work engagement on the PMEP score as the main independent variable including all controls on the firm and employee level as well as year fixed effects. The positive PMEP coefficient is statistically and economically significant: If the PMEP score increases by one unit (i.e., if employees are covered by an additional PMEP), the work engagement score is, on average, approximately 10% of a standard deviation higher. Next, column 2 shows results when using our non-linear PMEP measures as the key independent variables. In this specification, we compare employees who are simultaneously covered by either one, two, or three PMEPs to those that are not covered by any of these practices (reference group). The estimated coefficients in column 2 of Table 2.2 indicate that employees exposed to a larger number of practices at the same time exhibit statistically and economically significant higher levels of work engagement. The magnitude ranges from 6.7% of a standard deviation in case of one PMEP to 28.3% of a standard deviation when simultaneously being covered by three PMEPs, compared to employees not covered by any PMEP in a given firm and year. Since all these coefficients are significantly different from each other (all Wald tests with $p < 0.01$), an increase in the number of practices is significantly associated with increasing levels of work engagement.

Finally, we focus on the isolated effects of the different types of PMEPs. To do so, we include one distinct dummy variable for each practice in the specification reported in column 3. By splitting up the PMEP variable into its distinctive components, we can evaluate if one of them specifically drives the results or if all of them actually are linked to work engagement. The results indicate that, indeed, all three coefficients in column 3 are statistically and economically significant. Being covered by a PAI is, on average, associated with an increased level of work engagement of 12.4% of a standard deviation, holding everything else constant. The effect considering the presence of a TA is quantitatively similar in size (and not significantly different from PAI at conventional levels). Being subject to a PRP scheme is also positively connected to work engagement, even though the magnitude is smaller. The PRP coefficient is

[7] Additional descriptive statistics for firm-level competitive strategy and individual-level workforce characteristics (employee age, leadership responsibility, work contract, and gender) are provided in Table A.4 in Appendix A.2.

also significantly different from the PAI and the TA coefficient (both Wald tests with $p < 0.05$).[8]

Besides these between-firm estimations, it is also important to account for unobserved, time-constant heterogeneity on the firm level such as organizational culture, which might determine both an employee's work engagement level and coverage by PMEP measures. Table 2.3 shows these firm fixed effects regressions. The results are broadly in line with our baseline estimates in Table 2.2 with slightly larger coefficients when taking the within-firm rather than the between-firm perspective. Hence, even when comparing two identical employees in the same firm, we still find that (an increasing number of) PMEPs are significantly positively related to employees' work engagement.[9]

To conclude, our results suggest that the presence and the extent of PMEPs have a positive impact on work engagement on aggregate, even controlling for time-constant firm heterogeneity. This effect is mainly driven by the presence of PAIs and TAs. These individual-level results are in line with larger-scale firm-survey findings of a positive link between management quality (with performance management being an integral part of it) and firm-level outcomes such as productivity or profitability (Bloom et al. (2019)).

[8]When comparing the single coefficients, one should recall that the survey item on TAs is, by design, a filter question, i.e., firms with no PAIs in place were not eligible to answer the TA question. The rationale here is that negotiations about formalized performance targets between subordinates and supervisors usually take place during pre-scheduled PAIs, at least in German firms. The survey item on PRPs, however, applies to all firms.

[9]Previous literature has argued that performance management practices need to be considered as a set of complements, i.e., the effectiveness of performance management practices should vary depending on their respective combination and connectedness (Milgrom and Roberts (1995); Widener (2007); Malmi and Brown (2008); Ferreira and Otley (2009); Grabner and Moers (2013); Brickley et al. (2017); Qiu et al. (2022)). Table A.5 in Appendix A.3 shows regression results where we include interaction terms between PAI & PRP schemes and between TA & PRP schemes (the interaction term between PAI & TA is ruled out by survey design, because having a TA is only possible with a pre-scheduled PAI). While the main effects of the individual dummies for PAI, TA, and PRP schemes remain significant and quantitatively similar, none of the interaction terms is significantly different from zero, indicating that there are no complementarities. This supports recent evidence by Manthei et al. (2023), who do not find complementarities between feedback about performance metrics and performance-related pay, but significant effects of both isolated practices.

Table 2.2: PMEPs and Work Engagement

Dependent Variable:	Work Engagement (std.)		
Variables	(1)	(2)	(3)
PMEP Score	0.0979***		
	(0.0086)		
PMEP: One Practice		0.0667***	
		(0.0244)	
PMEP: Two Practices		0.200***	
		(0.0301)	
PMEP: Three Practices		0.283***	
		(0.0272)	
Appraisal Interview			0.124***
			(0.0253)
Target Agreement			0.111***
			(0.0238)
PRP Scheme			0.0483**
			(0.0196)
Firm Controls	yes	yes	yes
Employee Controls	yes	yes	yes
Year FE	yes	yes	yes
Observations	16,498	16,498	16,498
R-squared	0.078	0.078	0.078

Notes: The dependent variable is a score containing the equally-weighted average of nine items and is standardized. In column 1, the independent variable is a score representing how many performance management and evaluation practices (PMEPs) an individual is covered by in a given firm and year. In column 2, the PMEP score is subdivided into three dummy variables indicating whether employees are covered by one, two, or three practices simultaneously. For all three dummies, employees not covered by any of the three practices constitute the reference category. In column 3, the three PMEPs (appraisal interviews, target agreements, performance-related pay) are included as distinct dummy variables. Employee controls comprise female (0/1), age (4 dummies), supervisory position (0/1), white-collar employee (0/1), full-time position (0/1), monthly net income, type of employment contract (fixed term/permanent), permanent relationship (0/1), highest level of school education (6 dummies), highest level of further education (7 dummies), household size, and survey method (CAWI/CATI). Firm-level controls include industry (5 dummies), regional area (4 dummies), ownership type (6 dummies), firm size (4 dummies), and single-establishment firm. Standard errors clustered on the firm level are reported in parentheses. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively. Constant not displayed.

Table 2.3: PMEPs and Work Engagement – Firm Fixed Effects

Dependent Variable:	Work Engagement (std.)		
Variables	(1)	(2)	(3)
PMEP Score	0.129***		
	(0.0094)		
PMEP: One Practice		0.0797***	
		(0.0255)	
PMEP: Two Practices		0.243***	
		(0.0299)	
PMEP: Three Practices		0.374***	
		(0.0298)	
Appraisal Interview			0.144***
			(0.0257)
Target Agreement			0.162***
			(0.0254)
PRP Scheme			0.0675***
			(0.0211)
Firm Controls	yes	yes	yes
Employee Controls	yes	yes	yes
Year FE	yes	yes	yes
Firm FE	yes	yes	yes
Observations	16,347	16,347	16,347
Number of Firms	1,148	1,148	1,148
R-squared (within)	0.057	0.057	0.058

Notes: For a detailed description of dependent and independent variables, as well as employee-level and firm-level controls, see Table 2.2. Standard errors clustered on the firm level are reported in parentheses. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively. Constant not displayed.

We conduct various robustness checks employing different fixed effects specifications to verify the consistency of our results. First, we replicate our baseline specifications by controlling for individual fixed effects and firm-year fixed effects. The inclusion of interaction terms between firm and year fixed effects is an even more conservative and strict approach, as it allows unobserved, firm-specific characteristics to vary over time. As can be seen in Table IA.1 in the Internet Appendix A.4, the results remain unchanged, both with respect to the magnitude and to the significance of the coefficients, which are still positive and statistically significant. We also employ fixed effects on the individual level. They allow us to control for any time-constant individual heterogeneity that might affect the level of engagement and the existence or extent of PMEPs, such as, for example, talent or ability. As can be seen in Table IA.2 in the Internet Appendix A.4, our results in columns 1 and 2 largely confirm those from the baseline estimations, whereas in column 3 only PAIs are significantly correlated with work engagement. This is likely driven by the fact that we do not have sufficient within-person variation across years.[10] We conduct further (unreported) robustness checks. In order to account for concerns regarding the ordinal nature of the dependent variable, we apply ordered logistic regression models. To tackle issues related to selection bias and reverse causality, in addition, we estimate the instrumental variable method proposed by Lewbel (2012).[11] These issues might arise in case engaged individuals self-select into PMEPs or in case they are chosen for having a PMEP based on their previous level of engagement. In all robustness checks, our results remain qualitatively the same. Furthermore, all results are robust to clustering standard errors on the individual and firm-year level. To conclude, our results remain robust across a series of additional specifications, which supports a strong positive impact of PMEPs on work engagement.

2.4.2 The Influence of Contextual Factors

In Section 2.3.2, we descriptively shed light on heterogeneities in the incidence of PMEPs across our sample. In this section, we go one step further and analyze if the previously documented positive effect of PAIs, TAs, and PRP schemes on work engagement is moderated by specific firm characteristics, or if this effect is the same in all types of organizations.

[10] This is a common problem of individual fixed effects analyses in management practices research, namely that there is "not enough real time series variation (given measurement error) to identify any significant relationships" (Bender et al. (2018)).

[11] We use Stata's ivreg2h command developed by Baum and Schaffer (2012).

Table 2.4: Effect Heterogeneity

Dependent Variable:	Work Engagement (std.)			
Specific PMEP:	(1) PMEP Score	(2) Appraisal Interview	(3) Target Agreement	(4) PRP Scheme
PMEP	0.151***	0.171***	0.155**	0.205***
	(0.0283)	(0.0607)	(0.0683)	(0.0607)
Moderator: Firm Size				
PMEP ×	-0.0084	0.0631	0.0365	-0.107*
100 - 249 employees	(0.0273)	(0.0598)	(0.0658)	(0.0596)
PMEP×	-0.0609**	-0.0459	-0.0518	-0.189***
250 - 499 employees	(0.0294)	(0.0657)	(0.0679)	(0.0641)
PMEP×	-0.0512*	-0.0403	-0.0340	-0.163***
> 500 employees	(0.0273)	(0.0614)	(0.0626)	(0.0597)
Moderator: Industry				
PMEP×	-0.0137	-0.0295	-0.0224	-0.0318
Metal, Electrical & Automotive	(0.0221)	(0.0487)	(0.0477)	(0.0485)
PMEP×	-0.0193	-0.0643	0.0002	-0.0276
Retail, Logistics & Media	(0.0295)	(0.0640)	(0.0694)	(0.0627)
PMEP×	0.0021	-0.0226	-0.0674	0.105
Company-Related & Financial Serv.	(0.0268)	(0.0608)	(0.0550)	(0.0646)
PMEP×	-0.0770**	-0.151**	-0.197***	-0.0617
IT, Communication & Other Serv.	(0.0315)	(0.0765)	(0.0689)	(0.0867)
Moderator: Ownership Type				
PMEP×	-0.0160	-0.0005	-0.0189	-0.0690
Management	(0.0261)	(0.0564)	(0.0600)	(0.0551)
PMEP×	0.0058	0.0056	0.0651	-0.0555
Financial Investor	(0.0268)	(0.0620)	(0.0636)	(0.0572)
PMEP×	-0.0205	-0.0795	-0.0190	-0.0099
Dispersed Ownership	(0.0256)	(0.0631)	(0.0507)	(0.0711)
PMEP×	0.0396	0.0765	0.111	0.0207
Government/Public Sector	(0.0398)	(0.0879)	(0.0880)	(0.101)
PMEP×	0.0083	0.0176	0.0289	-0.0060
Other	(0.0255)	(0.0534)	(0.0484)	(0.0569)
Firm Controls	yes	yes	yes	yes
Employee Controls	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Moderator Base Terms	yes	yes	yes	yes
Other PMEPs		yes	yes	yes
Observations	16,498	16,498	16,498	16,498
R-squared	0.079	0.079	0.079	0.080

Notes: The table presents interaction terms between our performance management and evaluation practices (PMEPs) proxies (PMEP score and three distinct dummy variables), as referenced in the respective column header and the set of contextual factors (moderators). In columns 2 to 4, the other two PMEPs are controlled for, but not reported. A detailed description of the dependent variable, employee controls, and firm-level controls can be found in Table 2.2. Standard errors clustered on the firm level in parentheses. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively. Constant not displayed.

Table 2.4 shows interaction terms between the set of contextual factors and our PMEP proxies, namely PMEP score (column 1) and our three dummy variables for PAIs, TAs, and PRP schemes (columns 2 to 4). Note that the PMEP coefficients in the first row in this table always reflect the impact of the considered PMEP proxy on work engagement for the respective reference categories, i.e., fewer than 100 employees in case of firm size, the manufacturing sector in case of industry, and family firms and firms owned by their founder in case of ownership status. In columns 2 to 4, we also control for the occurrence of the other two PMEPs.

The first line of Table 2.4 confirms the positive impact of a higher PMEP score and the use of the three PMEPs on work engagement. All coefficients are economically and statistically significant, which supports our positive effects documented in the baseline specifications in Tables 2.2 and 2.3. Yet, the results further indicate that the overall effects of PMEPs on work engagement are moderated by firm size and industry. Concerning firm size, the results in column 1 of Table 2.4 show that the positive link between the number of PMEPs and work engagement is significantly weaker in firms with more than 250 employees compared to firms with fewer than 100 employees. Nevertheless, the overall impact of PMEPs on work engagement remains positive, as the sum of the PMEP coefficient and the respective interaction terms is greater than zero. Columns 2 to 4 confirm the negative coefficients for larger firms for all three practices; however, only the results for PRP schemes are statistically significant. But also in case of PRP schemes, the overall impact remains positive even in the largest firms. The result that employees covered by PRP schemes show higher levels of work engagement when working in smaller firms is in line with the findings from our distributional analysis in Table 2.1, which has shown that the heterogeneity across firm sizes, i.e., the distance between small and large firms, in the use of PRP schemes is smallest compared to the two other PMEPs. A potential explanation could be that managers in larger firms have difficulties observing their employees' actions (e.g., because they are located elsewhere) and, hence, have incomplete information about their subordinate's contributions (Bol (2011)), which could render PMEPs less effective. Additionally, in larger firms, a higher proportion of job tasks are cross-functional or team-based, which makes performance management more difficult and prone to biases. This is particularly true in case of subjective performance evaluations or when job tasks are mainly composed of creative and leadership tasks. If PRP schemes in larger firms are based on group outcomes to a larger extent, a free-rider problem might occur, which weakens the link between PMEP and work engagement even further (Holmström (1982); Sprinkle and Williamson (2006)). In smaller firms with smaller and more decentralized teams, formal performance management practices such as PAIs and TAs might be substituted by a closer interaction between supervisors and subordinates. PRPs, however, can be considered more complementary to induce individual effort and, hence, should be used relatively more often in smaller firms, as shown in Table 2.1. These firms might anticipate the positive impact on engagement and are able to set themselves

apart from other small firms by implementing PRP schemes, whereas in larger firms, everyone expects PMEP practices to be present.[12]

We next turn to industry as a moderating variable. In all columns, a negative interaction term for firms in the IT, Communication, & Other Services industry is observable (compared to the manufacturing industry), which is also significant for the PMEP score and for PAIs and TAs, but not for PRP schemes. This significantly weaker link between PMEPs and work engagement is intuitive, because many jobs in this industry are characterized by R&D-intensive, more creative, and mainly team-based tasks, for which performance measurement, target setting, and monitoring are likely to be more challenging. Furthermore, setting up targets in written form might hinder creativity and decrease motivation, particularly when output is only measurable at the team rather than the individual level. Again, the overall effect (sum of PMEP coefficient and respective interaction terms) is still positive in three out of four specifications. Recall that we observe a lower incidence of PRP schemes in firms in the IT, Communications, & Other Services sector, as shown in Table 2.1. Hence, although PRP schemes are used to a lower extent in this industry, the impact of PRP schemes on work engagement does not significantly differ between sectors.

For majority ownership as a moderator, we do not find significant interaction terms, indicating that PMEPs work very similarly across firms with different ownership patterns. To conclude, the main take-away of our study is that the overall effects of PMEPs on work engagement are moderated by firm size and industry, whereas for the type of ownership we do not see any significant interaction terms. Still, even though we document some effect heterogeneity, the overall effect of PMEPs on work engagement remains strongly positive across the vast majority of specifications. Hence, our results indicate that the usefulness of PMEPs as a tool to increase work engagement applies to firms of different sizes, industries, and ownership patterns.

2.5 Conclusion

In this paper, we analyzed PMEPs as a combination of three well-established performance management practices (performance appraisal interviews, written target agreements, and performance-related pay schemes) and related it to employee work engagement. In line with our theoretical arguments, our empirical results based on linked employer-employee data document positive, statistically significant work engagement-implications of PMEPs in general and for each of its specific PMEPs individually, i.e., PAIs, TAs, and PRP schemes. Further-

[12] We thank an anonymous reviewer for pointing this out.

more, we analyzed the context dependence of this relationship and found a smaller effect for larger firms and firms in the IT, Communication, & Other Services sector.

Our descriptive results have shown that the distribution of PMEPs varies across firms, which is in line with the findings of Grund and Sliwka (2009) and Broszeit et al. (2019). Similarly, our multivariate results indicate contextual heterogeneity of the PMEP-work engagement relationship with respect to firm size and industry. However, these moderating effects are small and only marginally affect the overall positive effect of PMEPs on work engagement. Hence, given this evident positive effect of PMEPs on work engagement, firms should consider a more prevalent use of PMEPs to enhance employee motivation and subsequent behavior, irrespectively of their size, industry, or ownership pattern.

Our empirical results indicate that all specific PMEPs have positive effects on employee engagement. However, additional analyses show that certain PMEPs (TAs and PAIs) affect work engagement more than others (PRP schemes). At the same time, we also find that employees covered by a greater number of PMEPs show higher levels of work engagement regardless of the specific PMEPs. Thus, from a theoretical perspective, our study suggests future research should consider both, a combined measurement of PMEPs as well as more fine grained measurements of its sub-dimensions that take into account the individual facets of the specific PMEPs. Moreover, our findings highlight the need to further explore contingencies when studying the implications of performance management systems (Dekker et al. (2012); Messner (2016); Otley (2016)).

This study is not without weaknesses. First of all, as described above, the main advantages of our data are the larger sample size and the longitudinal structure, which allow us to analyze the effectiveness of PMEPs across firms and its context dependence. As a consequence of this, however, we lack evidence on the exact specificities of the PMEPs within firms (such as a distinction between subjective and objective performance measurement or the contractual details of the target agreements and incentive schemes for each employee, see, e.g., Widener (2006); Hartmann and Slapničar (2009); Bol (2011); Voußem et al. (2016)). Future research could fill this void, by further exploring employee outcomes from these specific aspects of PMEPs. Second, we focus on formal PMEPs, as we lack evidence of informal controls. Hence, our results can be seen as the lower boundary of the true effect of the considered PMEPs. In the cases where informal management practices do also have a positive impact on work engagement, our documented effect sizes would be even higher. As these informal controls are considered equally important (Ferreira and Otley (2009)), future research should analyze employee outcomes of these tools in firms. A further issue of our study is the focus on linking PMEPs to work engagement. Despite empirical evidence on the relevance of work engagement for employee effort and

productivity (Christian et al. (2011); Presslee et al. (2023)), as well as firm outcomes (Salanova et al. (2005); Xanthopoulou et al. (2009)), our study is not informative about the direct effect of PMEPs on these employee-level and firm-level measures. These relationships could be investigated in future research.

Our study should not be viewed in isolation, but adds to the body of research on the implications of performance management. We concur with Falk and Heckman (2009) and Kampkötter and Sliwka (2016), who highlight that compiling evidence from a multitude of data sources and empirical methods – each of them with its strengths and weaknesses – is necessary to advancing the scientific state of knowledge. A hierarchy among different data set-ups or scientific approaches of analyses does not exist, as they act as "complements, not substitutes" (Falk and Heckman, 2009, p. 537; Schäffer (2018)). Through a plurality of data and methods, single-firm case studies with detailed firm-level information about internal characteristics of PMSs (see, e.g., Anderson et al. (2010); Sholihin et al. (2011); Bol (2011); Holzhacker et al. (2019)), survey data research with comprehensive responses from single firms or from business managers across different firms and industries (see, e.g., Tuomela (2005); Dekker et al. (2012); Arnold and Artz (2015); Bedford et al. (2016); Voußem et al. (2016); Groen et al. (2017); Arnold and Artz (2019); Guenther and Heinicke (2019); Demartini and Otley (2020)), and field experiments through causal analysis in controlled settings (see, e.g., Berger et al. (2013); Liu and Zhang (2015); Casas-Arce et al. (2017); Li and Sandino (2018); Casas-Arce and Martínez-Jerez (2022); Manthei et al. (2023)), have all contributed to a better understanding of the impact of performance management from different perspectives and levels. We believe that our longitudinal, linked employer-employee survey design adds a new layer to the existing literature and offers promising avenues for future research in this field.

Appendix A

A.1 UWES-9 Work Engagement Scale

The following nine statements are about how you feel at work. Please read each statement carefully and decide how often you feel this way about your work. Each of the following items should be answered on a five-point Likert scale from 1 (never) to 5 (daily):

1. At my work, I feel bursting with energy. (VI1)

2. At my job, I feel strong and vigorous. (VI2)

3. I am enthusiastic about my job. (DE1)

4. My job inspires me. (DE2)

5. When I get up in the morning, I feel like going to work. (VI3)

6. I feel happy when I am working intensely. (AB1)

7. I am proud of the work that I do. (DE3)

8. I am immersed in my work. (AB2)

9. I get carried away when I am working. (AB3)

Note: VI = Vigor scale; DE = Dedication scale; AB = Absorption scale.
Source: Schaufeli and Bakker (2003)

A.2 Descriptive Statistics

Table A.1: Descriptive Statistics – Main Dependent and Independent Variables

Variables	(1) Mean	(2) Std. Dev.	(3) Min.	(4) Max.
PMEP Score	1.473	1.135	0	3
PMEP: One Practice	0.290	0.454	0	1
PMEP: Two Practices	0.191	0.393	0	1
PMEP: Three Practices	0.267	0.442	0	1
Appraisal Interview	0.522	0.500	0	1
Target Agreement	0.356	0.479	0	1
PRP Scheme	0.594	0.491	0	1
Work Engagement (standardized)	-0.0073	1.003	-3.206	1.588
Work Engagement (non-standardized)	3.669	0.837	1	5
Work Engagement Items				
Engagement - Energy	3.388	1.006	1	5
Engagement - Strong	4.019	0.860	1	5
Engagement - Enthusiastic	3.755	1.027	1	5
Engagement - Inspiring	3.370	1.262	1	5
Engagement - Feel Like Working	3.499	1.132	1	5
Engagement - Happy	3.819	1.083	1	5
Engagement - Proud	4.088	1.006	1	5
Engagement - Immersed	3.623	1.155	1	5
Engagement - Carried Away	3.460	1.164	1	5

Notes: $N = 16,498$.

Table A.2: Descriptive Statistics – Firm-Level Controls and Contextual Factors

Variables	(1) N	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.
Industry - Manufacturing (0/1)	16,498	0.299	0.458	0	1
Industry - Metal, Electronics, & Automotive (0/1)	16,498	0.402	0.490	0	1
Industry - Retail, Logistics, & Media (0/1)	16,498	0.102	0.303	0	1
Industry - Company-Related & Financial Serv. (0/1)	16,498	0.125	0.330	0	1
Industry - IT, Communication, & Other Serv. (0/1)	16,498	0.0719	0.258	0	1
Region - North (0/1)	16,498	0.192	0.394	0	1
Region - East (0/1)	16,498	0.251	0.434	0	1
Region - South (0/1)	16,498	0.276	0.447	0	1
Region - West (0/1)	16,498	0.281	0.449	0	1
Size - Less Than 100 Employees (0/1)	16,498	0.127	0.333	0	1
Size - 100 to 249 Employees (0/1)	16,498	0.231	0.421	0	1
Size - 250 to 499 Employees (0/1)	16,498	0.231	0.421	0	1
Size - More Than 500 Employees (0/1)	16,498	0.412	0.492	0	1
Strategy - Cost Leadership (0/1)	11,084	0.0533	0.225	0	1
Strategy - Benefit Leadership (0/1)	11,084	0.430	0.495	0	1
Strategy - Both Equally (0/1)	11,084	0.517	0.500	0	1
Ownership - Family/Founder (0/1)	16,498	0.429	0.495	0	1
Ownership - Management (0/1)	16,498	0.151	0.358	0	1
Ownership - Financial Investor (0/1)	16,498	0.0932	0.291	0	1
Ownership - Dispersed Ownership	16,498	0.110	0.312	0	1
Ownership - Government/ Public Sector (0/1)	16,498	0.0233	0.151	0	1
Ownership - Other types (0/1)	16,498	0.194	0.396	0	1
Single-establishment Firm (0/1)	16,498	0.693	0.461	0	1

Table A.3: Descriptive Statistics – Employee-Level Controls

Variables	(1) N	(2) Mean	(3) Std. Dev.	(4) Min	(5) Max
Age Category - under 25 (0/1)	16,498	0.0385	0.192	0	1
Age Category - 25 to 39 (0/1)	16,498	0.233	0.423	0	1
Age Category - 40 to 54 (0/1)	16,498	0.521	0.500	0	1
Age Category - over 55 (0/1)	16,498	0.207	0.405	0	1
Education - None (0/1)	16,498	0.0042	0.0645	0	1
Education - Lower Secondary School (0/1)	16,498	0.219	0.414	0	1
Education - Intermediate Secondary School (0/1)	16,498	0.424	0.494	0	1
Education - University of Applied Sciences Entrance Qualification (0/1)	16,498	0.110	0.313	0	1
Education - General Higher Education Entrance Qualification (0/1)	16,498	0.235	0.424	0	1
Education - Other (0/1)	16,498	0.0069	0.0825	0	1
Female (0/1)	16,498	0.272	0.445	0	1
Fixed-Term Contract (0/1)	16,498	0.0450	0.207	0	1
Full Time/Part Time (0/1)	16,498	0.127	0.333	0	1
Monthly Net Income (in Euros)	16,498	2,419	1,842	1	74,221
Household Size	16,498	2.777	1.228	1	14
Permanent Relationship (0/1)	16,498	0.841	0.366	0	1
Leadership Position (0/1)	16,498	0.290	0.454	0	1
Further Education - None (0/1)	16,498	0.0209	0.143	0	1
Further Education - Apprenticeship (0/1)	16,498	0.462	0.499	0	1
Further Education - Vocational/ Business School (0/1)	16,498	0.0936	0.291	0	1
Further Education - Master Craftsman/ Technical College (0/1)	16,498	0.206	0.404	0	1
Further Education - University of Applied Sciences (0/1)	16,498	0.0992	0.299	0	1
Further Education - University (0/1)	16,498	0.114	0.318	0	1
Further Education - Other (0/1)	16,498	0.0044	0.0659	0	1
White Collar (0/1)	16,498	0.625	0.484	0	1

Table A.4: Distribution of PMEPs – Extended

	Variables	PMEP Score	PAI[a]	TA[a]	PRP[a]
Strategy	Cost Leader	1.38	49.74	36.04	51.78
	Benefit Leader	1.75	63.16	44.30	67.63
	Both Strategies	1.34	49.21	31.50	53.40
Workforce	Age $<$ 25 years	1.49	52.44	33.39	63.46
	Age 25-39 years	1.56	56.01	37.36	62.90
	Age 40-54 years	1.49	52.56	36.62	59.51
	Age $>$ 55 years	1.33	47.13	31.52	54.72
	No Leadership Position	1.39	49.47	32.75	56.62
	Leadership Position	1.68	58.99	42.56	66.36
	Full-time Employee	1.51	52.73	36.27	61.67
	Part-time Employee	1.24	48.83	30.97	44.21
	Female Employee	1.35	51.12	34.70	49.68
	Male Employee	1.52	52.65	35.94	63.10

Notes: [a] in %; PMEP Score = Sum of performance management and evaluation practices per individual, PAI = Performance appraisal interview, TA = Written target agreement, PRP = Performance-related pay scheme.

A.3 Interactions between PMEPs

Table A.5: Interactions between PMEPs

Dependent Variable:	Work Engagement (std.)			
Variables	(1)	(2)	(3)	(4)
Appraisal Interview	0.124***	0.132***	0.124***	0.145***
	(0.0310)	(0.0317)	(0.0252)	(0.0257)
Target Agreement	0.111***	0.160***	0.110***	0.145***
	(0.0243)	(0.0257)	(0.0382)	(0.0388)
PRP Scheme	0.0490*	0.0569**	0.0479**	0.0600**
	(0.0267)	(0.0285)	(0.0226)	(0.0246)
Appraisal Interview X	-0.0014	0.0220		
PRP Scheme	(0.0380)	(0.0375)		
Target Agreement X			0.0013	0.0245
PRP Scheme			(0.0432)	(0.0427)
Firm Controls	yes	yes	yes	yes
Employee Controls	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Firm FE		yes		yes
Observations	16,498	16,347	16,498	16,347
R-squared (within)	0.078	0.058	0.078	0.058

Notes: For a detailed description of dependent and independent variables as well as employee-level and firm-level controls, see Table 2.2. Standard errors clustered on the firm level are reported in parentheses. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively. Constant not displayed.

A.4 Internet Appendix

Table IA.1: PMEPs and Work Engagement – Year X Firm Fixed Effects

Dependent Variable:	Work Engagement (std.)		
Variables	(1)	(2)	(3)
PMEP Score	0.132***		
	(0.00988)		
PMEP: One Practice		0.0743***	
		(0.0273)	
PMEP: Two Practices		0.244***	
		(0.0316)	
PMEP: Three Practices		0.379***	
		(0.0315)	
Appraisal Interview			0.150***
			(0.0272)
Target Agreement			0.169***
			(0.0267)
PRP Scheme			0.0603***
			(0.0227)
Firm Controls	yes	yes	yes
Employee Controls	yes	yes	yes
Year × Firm FE	yes	yes	yes
Observations	16,026	16,026	16,026
R-squared (within)	0.058	0.059	0.059

Notes: For a detailed description of dependent and independent variables as well as employee-level and firm-level controls, see Table 2.2. Standard errors clustered on the firm level are reported in parentheses. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively. Constant not displayed.

Table IA.2: PMEPs and Work Engagement – Individual Fixed Effects

Dependent Variable:	Work Engagement (std.)		
Variables	(1)	(2)	(3)
PMEP Score	0.0349***		
	(0.0126)		
PMEP: One Practice		0.00734	
		(0.0314)	
PMEP: Two Practices		0.0983***	
		(0.0374)	
PMEP: Three Practices		0.0889**	
		(0.0388)	
Appraisal Interview			0.0959***
			(0.0301)
Target Agreement			0.0100
			(0.0269)
PRP Scheme			-0.0152
			(0.0253)
Firm Controls	yes	yes	yes
Employee Controls	yes	yes	yes
Individual FE	yes	yes	yes
Observations	16,498	16,498	16,498
Number of Employees	12,057	12,057	12,057
R-squared (within)	0.028	0.029	0.030

Notes: For a detailed description of dependent and independent variables as well as employee-level and firm-level controls, see Table 2.2. Standard errors clustered on the firm level are reported in parentheses. The symbols *,**, and *** represent significance levels of 10%, 5%, and 1%, respectively. Constant not displayed.

Chapter 3

In the Hand of the Family: Management Practices and Perceived Job Quality[†]

This paper explores the use and implications of management practices in family-owned firms compared to firms with dispersed ownership. We make use of a longitudinal, representative employer-employee dataset with detailed data on firm-level management practices and family firm management types. The analysis reveals that differences in the adoption of structured management practices are predominantly driven by management type rather than ownership. Family-owned firms led by family members implement significantly fewer formal management practices, whereas those managed by non-family managers adopt more structured practices, though still below the levels observed in firms with dispersed ownership. Yet, employees in family-owned firms, particularly those with non-family managers, rate job quality (e.g., job satisfaction, procedural fairness, leadership quality) similarly or even higher despite fewer formal practices. These findings suggest that informal practices and a distinctive firm culture in family-owned firms may foster employee motivation and partially substitute for formal management structures. Importantly, additional heterogeneity analyses reveal that this substitution is only effective for lower-skilled employees and those in non-managerial positions, while formal management practices remain critical for higher-skilled and supervisory roles.

Key Words: Corporate Ownership Structure; Management Practices; Family Firms; Job Satisfaction; Job Quality

JEL Classification: G32, M21, M50

[†]This chapter is based on joint work with Patrick Kampkötter, Julian Wenzel, and Stefanie Wolter and published in the Journal of Economic Behavior & Organization (see Ehmann et al. (2025)).

3.1 Introduction

Earlier studies have indicated that substantial variations in productivity can be observed between firms within the same country and industry, even when technology differences are taken into account. This unexplained part of productivity differences is usually attributed to varying levels of management quality, particularly structured management practices (Bloom and Van Reenen (2007, 2010); Bloom et al. (2012); Englmaier et al. (2022)). A key trait in this context is the ownership structure of a firm, which has been identified as a significant precursor to the quality of management. In many countries, family ownership is and has long been an important and widespread form of business ownership, both among private and publicly traded firms (La Porta et al. (1999); Burkart et al. (2003); Bjuggren (2015)). Recent empirical evidence from Aminadav and Papaioannou (2020) shows that about 17.5% of companies around the world are family-controlled. This figure even doubles when including firms controlled by an unidentified private owner, which are mostly also controlled by families or individuals. Internationally, this fraction of family-owned firms or firms controlled by an unidentified private owner ranges from only 12% in Japan to 52% in Germany and even 73% in Estonia.

In the literature on corporate ownership, a frequently studied subject is the disparity in productivity between family-owned businesses and those with a more dispersed ownership structure (non-family-owned firms). Previous studies often focus on the direct link between ownership structure and organizational performance (e.g., labor productivity, ROA, ROS, sales growth, market-to-book-ratio). Although a multitude of these studies document that family-owned firms perform poorly in terms of productivity compared to firms with a dispersed ownership structure, research on the topic is still not unequivocal (Anderson and Reeb (2003); Wagner et al. (2015); Neckebrouck et al. (2018); Bandiera et al. (2018); Christensen-Salem et al. (2021); Peláez-León and Sánchez-Marín (2022); Creemers et al. (2023)). A complementary research stream also examines the difference in a firm's non-financial performance (e.g., environmental performance, internal/external social performance, innovation input/output). Results are mixed and depend on the country and the specific non-financial component analyzed (Dyer and Whetten (2006); Berrone et al. (2010); Cruz et al. (2014); Block and Wagner (2014); Matzler et al. (2015); Rees and Rodionova (2015); Duran et al. (2016); El Ghoul et al. (2016); Cabeza-García et al. (2017); Canavati (2018); Dal Maso et al. (2020); Abeysekera and Fernando (2020); García-Sánchez et al. (2021); Block et al. (2023); Miroshnychenko et al. (2025)).

Potential explanations for this observed heterogeneity in performance and productivity between family-owned and non-family-owned firms refer to the use of management practices and the participation of the family in the top management team (Reid and Adams (2001); De Kok et al. (2006); Cruz et al. (2011); Bloom et al. (2012); Neckebrouck et al. (2018); Sánchez-

Marín et al. (2019); Dal Maso et al. (2020); Peláez-León and Sánchez-Marín (2023)). This is highlighted, for instance, by findings based on the World Management Survey (WMS), which reveal that family-owned firms generally are not significantly worse managed than non-family-owned firms (Bloom and Van Reenen (2007, 2010); Broszeit et al. (2019)). Yet, on average, family-owned firms with family members being involved in the top management team exhibit significantly lower management scores than non-family-owned firms. In line, a review of the aggregate survey data on manufacturing and retail companies conducted between 2004 and 2010 reveals that the management score of non-family-managed family-owned businesses is comparable to that of companies with dispersed shareholders (Bloom et al. (2012); Lemos and Scur (2019)).

We contribute to the literature in two ways. First, we complement prior research by shedding light on the specific role of structured, i.e., formalized, management practices that are common in family- and non-family-owned firms in various industries. Whereas previous papers have mainly analyzed the adoption of operational-focused management practices tailored to production environments, particularly in medium-sized manufacturing firms (see, for instance, Bloom and Van Reenen (2007, 2010); Bloom et al. (2012); Broszeit et al. (2019); Lemos and Scur (2019)), we are able to go beyond the examination of operational-focused management practices and also include strategic management practices from a broader range of the value chain. We thereby also analyze how management formalization in family-owned firms is shaped by leadership structures beyond the CEO. Although the CEO undeniably holds significant power in strategic decision-making, we propose that the entire management board is crucial in the formalization process. This involves differentiating between family-owned firms that have a management board made up exclusively of family members, solely of non-family managers, or a combination of both.

Importantly, our second and more substantive contribution lies in investigating the relationship between family ownership, the type of management, and the perceived job quality of employees. So far, previous literature has investigated individual-level differences in wages, employee mobility, and job security between family-owned and non-family-owned firms based on administrative data sets (see, for instance, Bassanini et al. (2013); Ellul et al. (2018); Pagano et al. (2024)). It has been documented that employees in family-owned firms are willing to accept lower wages (which could also indicate that the quality of the workforce is more compressed, see e.g., Bassanini et al. (2013) and Bender et al. (2018)) and that those employees are exposed to lower management quality and less formalized management practices. The idea is that of a compensating wage differential: as family-owned firms offer higher job security, they can pay lower wages than the market wage and employees will not quit their employer immediately. In addition to the focus on job security, more detailed analyses of potential differences in perceived

work quality and corporate culture between these types of firms are scarce. Previous literature relied on proxies such as absenteeism (Bennedsen et al. (2019)) as an indirect representation rather than a direct metric of job satisfaction or working conditions. Furthermore, assessments of work quality are often based on managerial assessments measured at the firm level rather than direct feedback from employees (Bloom et al. (2011); Pittino et al. (2016); Kang and Kim (2020)) or use rather crude employee-level measures within a selective sample, not addressing firm-specific characteristics (Huang et al. (2015)). We complement previous literature by extending the outcome variables to capture the broader concept of work quality and integrate individual-level outcomes such as employees' global job satisfaction, procedural fairness, goal clarity, and leadership quality. Hence, we aim at investigating whether family-owned firms are able to substitute the lack in formal management practices with a more superior workplace or (informal) management culture which, in the end, makes their employees rate workplace quality higher or at least comparable to that of firms with dispersed ownership. We further contribute by examining potential effect heterogeneity with respect to the composition of the workforce in terms of supervisory responsibilities and the level of qualification.

This combined analysis of formal management practices drawn from employer surveys paired with individual-level perceptions of work quality is based on a new linked employer-employee dataset, the Linked Personnel Panel, which is representative of private sector establishments with more than 50 employees liable to social security in Germany and its workforce.[13] For our analyses, we can use five waves running from 2012 to 2021. The analyses confirm that differences in the adoption of management practices between family-owned and non-family-owned firms are driven primarily by the type of management rather than the ownership structure. Family-managed family-owned firms tend to implement fewer structured management practices. In contrast, exclusively non-family-managed family-owned companies more readily adopt structured practices, aligning their management approaches more closely with those of their widely-held counterparts. We further show that perceived job quality in family-owned firms, such as, for instance, job satisfaction, is not significantly negatively affected by the relative lack of formal management practices. This indicates that informal practices, such as trust-based relationships and personalized interactions, play an important role in maintaining a positive work environment. Furthermore, we find tentative evidence that exclusively non-family-managed family-owned firms may achieve even higher levels of perceived job quality. This could result from a combination of professional management practices and the relational strengths and superior firm culture often found in family-owned businesses, creating a workplace that employees perceive as especially positive. Importantly, this aggregate result is driven by family-owned

[13] In our data, we observe establishments, which are regionally and economically separate units. In principle, a firm might therefore consist of several establishments. For simplicity and whenever there is no possibility of confusion, we use the terms "firm" and "establishment" interchangeably.

firms with a relatively small share of highly-qualified workforce and employees working at lower hierarchical levels. In family-owned enterprises with a higher percentage of qualified employees, specifically for those in supervisory positions, this positive association with workplace quality is reduced or even turns negative, highlighting the importance of workforce heterogeneity.

Our paper is organized as follows. Section 3.2 reviews the theoretical background and the literature and develops hypotheses. Section 3.3 describes our data and Section 3.4 outlines the statistical analyses. The main results are presented in Section 3.5. Finally, Section 3.6 concludes.

3.2 Theoretical Background and Related Literature

Ownership and management shape the structures and dynamics of a company, and thus also the management practices in place. Previous literature highlights that control, rather than ownership alone, is the key driver of managerial behavior. Hence, it is important to consider not only the implications of family ownership but also variations in management types within family-owned firms, i.e., to differentiate between family-owned companies that are exclusively managed by family members and family-owned companies that are exclusively run by non-family managers (Bennedsen et al. (2007); Amore et al. (2011); Bloom et al. (2012); Cucculelli et al. (2014); Lemos and Scur (2019); Amore et al. (2021); Pagano et al. (2024)).[14] Multiple theoretical frameworks are informative. Agency theory, one of the most prominent theories in corporate governance, highlights how ownership and control affect managerial incentives (Jensen and Meckling (1979); Fama and Jensen (1983)). Family-owned firms with family managers experience fewer agency conflicts due to the alignment of ownership and management interests, reducing the need for formalized control mechanisms such as performance monitoring and incentive-based compensation. However, family-owned enterprises with non-family management are more similar to widely-held firms, as ownership and control are separated. Comparatively, this increases the need for structured management practices to align interests.

A second stream of literature builds on the idea that family-owned companies often prioritize socioemotional wealth, emphasizing long-term control, family legacy, and non-financial goals (Gómez-Mejía et al. (2007); Berrone et al. (2010)). This focus can be associated with a lower reliance on formalized management practices, particularly if they are perceived as potentially undermining family control or fostering external influence (Cruz et al. (2011); Bjuggren (2015); Sánchez-Marín et al. (2019); Peláez-León and Sánchez-Marín (2022, 2023)). Hernández-Linares et al. (2023) further support this argument, suggesting that commitment to preserving socioe-

[14] For a further discussion of recent literature, see Tsoutsoura (2021).

motional wealth and the adoption of management practices are more likely substitutes than complements. In addition, the pursuit of socioemotional goals is likely to be more prominent the stronger the family influence, particularly also in terms of management control. Consequently, family-led family-owned firms are more likely to scale back the adoption of management practices, whereas non-family-managed family-owned companies tend to use practices similar to those of non-family-owned enterprises.

Third, stewardship theory (Davis et al. (1997, 2010)) suggests that family members act as stewards, prioritizing the firm's long-term success over short-term personal gains (Corbetta and Salvato (2004); Miller and Le Breton-Miller (2006); Miller et al. (2008); Le Breton-Miller and Miller (2009); Davis et al. (2010); Miller et al. (2013)). This stewardship culture may reduce the need for rigid monitoring mechanisms, favoring informal control and relational governance over formalized management structures (Davis et al. (2010); Miller et al. (2013); Hoffmann et al. (2016); Chrisman (2019); Afonso Alves et al. (2021)).

Lastly, talent constraints in family-managed family-owned firms can influence management formalization. Nepotism and asymmetric altruism may lead to suboptimal hiring decisions, resulting in lower average management quality and resistance to professionalized practices (De Kok et al. (2006); Bloom and Van Reenen (2007); Bennedsen et al. (2007); Bertrand et al. (2008); Bloom and Van Reenen (2010); Firfiray et al. (2018)). This issue is particularly pronounced in family-led companies, especially those with rigid succession norms (Bloom and Van Reenen (2010)). Conversely, non-family-managed family-owned firms, with access to a broader talent pool, may be more inclined to implement structured management practices.

Previous empirical studies support the theoretical argument that family-owned enterprises tend to exhibit lower levels of formalized management practices (Speckbacher and Wentges (2012)). Again, family firm control appears to be more influential than ownership alone. For instance, Bennedsen et al. (2007) find that family-owned firms with non-family CEOs show better firm performance than family-owned firms with family CEOs. In line, Pagano et al. (2024) document that family-owned enterprises face greater difficulties in managing the sudden death of their CEO compared to non-family-owned enterprises, with these challenges being particularly pronounced for family-owned companies with a family CEO. The same pattern tends to hold true with respect to the degree of formalization. Bloom et al. (2012) and Lemos and Scur (2019) find that family-owned firms with a family CEO exhibit substantially weaker management practices, whereas companies that appoint "professional" (i.e., non-family) CEOs perform similarly compared to non-family-owned firms.

All in all, in line with the majority of the theoretical arguments and empirical evidence, we propose the following hypothesis:

Hypothesis 1: Family-owned firms with exclusively family members on the management board are less likely to adopt formalized (professional) management practices compared to family-owned firms with exclusively non-family managers on the management board and firms with dispersed ownership.

Importantly, we not only focus on the application of management practices, but on the nexus between the use of management practices, corporate ownership, type of management, and perceived workplace quality. A review of the literature by Sageder et al. (2018) shows that family-owned firms enjoy a positive reputation among stakeholders and customers with respect to different factors such as social responsibility and reliability compared to non-family-owned firms. Considering these findings on external reputation, one could also expect that internal reputation, i.e., a higher satisfaction of employees or higher scores on employee attitudes, is also related to corporate ownership status. A prevalent issue frequently highlighted in this discourse is the potential impact of a reduced degree of formalization on employees. For instance, previous findings by Bloom et al. (2011) indicate that, on average, formalized management practices can improve perceived work-life balance through the provision of family-friendly work practices. However, some (even though not all) of the particularities of family-owned firms described above suggest a generally lower need for the implementation of strategic management practices compared to companies with dispersed ownership, particularly in case of family-owned enterprises managed by family members. Therefore, while traditional economic theory suggests that formalized management practices generally enhance job quality by improving working conditions and career development (Bloom and Van Reenen (2007); Bloom et al. (2011, 2019)), family-owned firms may achieve similar or even better perceived outcomes through informal mechanisms such as trust, loyalty, and implicit contracts. Hence, independent of the amount of structured management policies implemented, job quality in family-owned firms might be driven by other more intangible factors such as, for instance, a superior firm culture.

Theories in organizational economics and behavioral management suggest that family ownership can influence employee outcomes in several ways. First, Shleifer and Summers (1988) suggested that shared trust and long-term commitment between shareholders and stakeholders are essential to impose implicit contracts. This is argued to be particularly strong in family-owned firms which helps to foster a greater sense of security among employees (e.g., see Anderson and Reeb (2003); Bjuggren (2015); Bennedsen et al. (2019); Lemos and Scur (2019); Amore et al. (2022)), as they tend to emphasize stability over short-term financial gains (Ellul et al. (2018)). This aligns with empirical findings, which highlight the role of family ownership in

creating long-lasting employment relationships. Sraer and Thesmar (2007) and Bassanini et al. (2013) indicate that employees in family-owned firms, despite earning lower wages, benefit from greater job stability which is likely driven by stronger, trust-based ties between the owning family and employees. Specifically, Bassanini et al. (2013) show significantly lower dismissal rates for family-owned companies, suggesting that long-term employment security and non-monetary benefits help offset lower salaries. In this case, fewer management practices align with a reduced need for management practices and therefore do not necessarily compromise job quality.[15]

Second, employees in family-owned firms may also benefit from a more personalized and cohesive workplace culture, associated with a higher job quality. For instance, stewardship behavior of family members can be a crucial building block for a firm-wide stewardship culture that can be reflected in high levels of organizational identification, prosocial motivation, and the pursuit of cooperative goals. Therefore, while the rationale for stewardship is particularly prevalent for family employees, it can also be transferred to non-family employees (Miller et al. (2008); Afonso Alves et al. (2021); Bormann et al. (2021)). This point is implicitly supported by Bennedsen et al. (2019), who find that employees in family-owned enterprises show, on average, lower levels of absenteeism, suggesting a stronger work ethic or job commitment. An additional explanation might be that through the ability of family-owned firms to maintain implicit contracts, employees reciprocate this trust and commitment with cooperative behavior (Sraer and Thesmar (2007); Ellul et al. (2018)). This reinforces the idea that employee outcomes in family-owned companies extend beyond formal incentive contracts, fostering job quality through implicit relational contracts. The results of Kang and Kim (2020) support this argument by showing that family-owned firms generally have lower levels of employee treatment concerns, suggesting that they are more likely to invest in worker relations. The findings of a related study by Pittino et al. (2016) point in a similar direction. Evaluating a sample of 232 SMEs in Austria and Hungary, they show that the adoption of specific management practices as such is not the decisive factor impacting employee retention rates in family-owned firms, indicating that a distinct firm culture may play an important role. Regardless, the retention rates of family-owned companies are comparable to or higher than those of non-family-owned companies. Conversely, Neckebrouck et al. (2018) claim that the behavior of family-owned firms toward employees ("organizational stewardship") can also be associated with unfavorable outcomes, since they offer less employee training and experience greater voluntary turnover by employees.

For both rationales, the degree to which family members are involved in management has

[15]In contrast, Neckebrouck et al. (2018) and Pagano et al. (2024) show diverging results. Furthermore, Mueller and Philippon (2011) find that family ownership is more common in countries with rigid labor laws, suggesting that these firms navigate complex labor relations more effectively, absorb higher labor costs, and offer stronger implicit job protections.

implications for job quality. For example, as stressed by Sraer and Thesmar (2007) and Lemos and Scur (2019), family-owned firms with a family CEO seem to be more credible than family-owned firms with a non-family CEO when it comes to long-term commitment and consequently the ability to impose implicit contracts, i.e., due to higher reputation costs. Kang and Kim (2020) confirm this claim and show that family-owned companies tend to invest more in worker relations when family members are involved in management. Thus, it may be more feasible for them to forgo formal management practices. The overall impact of decreased formal, yet increased informal, practices may result in similar job quality levels as seen in non-family-owned companies. On the other hand, non-family-managed family-owned firms might be able to combine the best of both settings. They adopt professional management practices similar to widely-held enterprises and additionally benefit from a family workplace culture inherent to family-owned firms. Consequently, the ultimate effect of this governance structure on job quality might even be positive, leading to higher levels of job quality than in widely-held companies. The study by Huang et al. (2015) provides preliminary insights into the dependence of job quality in family-owned enterprises on the management structure. They analyze employee-survey data collected from Glassdoor and highlight that compared to non-family-owned firms, family-owned firms run by founders have a higher level of employee satisfaction, whereas those managed by relatives of the founder have a similar (relatives not being the CEO) or even lower level of employee satisfaction (one of them is the CEO). Nevertheless, the latter result should be interpreted with caution, as these companies represent only 1% of the sample.

In summary, some studies suggest that formalized structures improve efficiency and job quality, while others argue that the informal governance of family-owned firms can substitute for such practices. These findings reinforce the idea that this substitution can lead to distinct but often positive employee experiences. Although several studies focus on evaluating the trade-offs between job security and wages in family-owned firms (e.g., Sraer and Thesmar (2007); Bassanini et al. (2013); Ellul et al. (2018); Pagano et al. (2024)), they only partially address aggregate job quality.

Building on the previous discussion, we form the following hypothesis:

Hypothesis 2: Employees in family-owned firms show similar or even higher perceptions of job quality than employees in widely-held firms, with the strength of the relationship varying by the respective management structure in family-owned firms.

While the arguments above suggest a potentially positive effect of family ownership on individual-level outcomes, there might be heterogeneity across different levels of the organizational hierarchy. Both Pagano et al. (2024) and Bennedsen et al. (2019) emphasize that

employees in family-owned firms, particularly those in managerial or leadership roles, may experience workplace conditions differently to non-managerial employees. They attribute this to a distinctive characteristic of family-owned companies, a tendency to retain control within their inner circle. As a result, individuals from outside the family are less likely to advance to senior positions and face limited opportunities to climb the organizational hierarchy, as typically preference is given to family members. Pagano et al. (2024) refer to this phenomenon as a family firm glass ceiling. The outcomes of their study also provide supporting evidence showing that family-owned enterprises on average offer lower top-end wages, fewer promotions, and smaller salary increases upon promotion compared to non-family enterprises. The previously mentioned findings by Bennedsen et al. (2019) further enrich this line of research. While, ceteris paribus, family ownership is associated with lower absenteeism, this relationship is primarily driven by non-managerial employees and is assumed to be offset by a higher likelihood of nepotism at upper hierarchical levels. Yet, the implementation of structured management practices at upper organizational levels may mitigate these barriers by providing all employees with greater opportunities for professional development. Thus, within the broader employee population, structured management practices are particularly likely to enhance individual-level outcomes for those in senior positions. Consequently, these types of employees might even rate job quality lower in family-owned firms than in widely-held enterprises. Again, management in family-owned companies matters. The argument is especially compelling for family-led family-owned firms, where strict control is already maintained within the exclusive inner circle of family members in management.

Another important dimension to consider is the heterogeneity in employee ability. Previous research suggests that family firms may systematically attract or retain less-qualified workers than their non-family counterparts (e.g., see Bennedsen et al. (2007), Bassanini et al. (2013), and Bender et al. (2018)). Such employees may rely less on formal management practices for their daily work and career development, which could further explain the lower level of structured management in some family-owned firm settings. In contrast, highly-skilled employees typically benefit more from well-defined career structures, performance incentives, and professional development frameworks. Therefore, the relationship between family ownership and employee-level outcomes may depend not only on hierarchical position, but also on the skill composition of the workforce.

Based on the raised arguments, we pose the following hypothesis:

Hypothesis 3: The proposed relationship between family ownership and perceived job quality is weaker among employees in higher hierarchical levels and with higher average employee skill levels.

3.3 Data and Empirical Strategy

3.3.1 Data

Our empirical analysis is based on the Linked Personnel Panel (LPP1221_v2, in the following denoted as LPP), a linked employer-employee panel data set that is representative of German private sector establishments with more than 50 employees (Kampkötter et al. (2016); Ruf et al. (2020)). The employer survey provides detailed information about a variety of management practices along the dimensions of recruitment, selection, performance management, talent management, employee development, retention management, and corporate culture. Additionally, a number of organizational characteristics are available, such as ownership structure, number of hierarchical levels, and business strategy. The first wave of the employer survey was administered in 2012 and included 1,219 establishments randomly drawn from the IAB establishment panel, a representative annual panel study of nearly 16,000 German establishments representing all industries and establishment sizes (Fischer et al. (2009); Ellguth et al. (2014)). The subsequent waves of the LPP employer survey were carried out every two years, most recently in 2020, and contained about 800 firms in each wave. Of the surveyed establishments, between 6,500 and 7,500 employees working in these establishments were randomly drawn and interviewed outside of their workplace again every two years by telephone (CATI) and online (CAWI, starting 2019) about job and workplace characteristics, perceptions about job and management practices, personal attitudes, and attitudes towards their organization. The most recent employee survey was conducted in 2021.

3.3.2 Measures and Descriptive Statistics

Ownership and management structure. Our main explanatory variable contains information on the ownership structure of the surveyed establishment. The LPP asks for the main owner of the establishment and, in case the establishment is part of a company or holding, information on the main owner of the parent company.[16] The available categories are *family-owned and/or founder-owned*[17], *management-owned*, *financial investor*, and *widely held stock/stock market*. Importantly, firm respondents can choose only exactly one response category. This should minimize the risk of potential overlaps between different ownership types.

[16]The original wording of the question is as follows: "Who is the *ultimate owner* of your establishment? If your establishment is not independent, who is the owner of the parent company?".

[17]It is not possible to distinguish between companies owned by the family and/or the founder, which might qualify as a limitation. However, previous literature shows that founder-owned firms play only a minor role (Sraer and Thesmar (2007)). In Germany, only 5% of firms are controlled by the individual who founded the firm (Bloom and Van Reenen (2007); Bender et al. (2018)). Hence, we are confident that this limitation might not be a major one, at least in our sample of German firms, and denote this group as "family-owned firms" in the remainder of the paper.

Thus, in line with the survey question, our definition of family-owned firms targets businesses that are characterized by a controlling interest or a significant stake that enables a family to substantially influence or control operational and strategic business decisions.[18]

Finally, we pool financial investor and stock market ownership into one category called *dispersed ownership/widely-held firms*, as these two are closely related for the purpose of our analysis. Also, we exclude the other two selectable categories of ownership: enterprises with the state as a significant shareholder (about 1.8% of the cases), which are a specific type of company, and other forms of ownership (about 15.8% cases), as these types are not specified in more detail in the survey.[19]

Table 3.1 shows the (weighted) distribution of different ownership types throughout the years. Across waves, 62% to 73% of all establishments are family-owned. Furthermore, 18% to 28% are management-owned, and around 9% to 13% are classified as dispersed ownership (financial investor and stock market). Furthermore, family ownership is very stable over time, as can be seen in Table B.1 in Appendix B.1. Panel A shows that almost 92% of those firms that were family-owned in 2012 were still owned by a family in the year 2020. Furthermore, transitions from family-owned firms to firms with dispersed ownership are very rare, with only 1.8% of family-owned firms in 2012 changing ownership status to dispersed ownership in 2020. Similarly, companies with dispersed ownership also very rarely transition to family-owned companies (2.7%), but much more often to management-owned companies (18.9%).

To validate our ownership information derived from the LPP employer survey, we merge Bureau van Dijk's (BvD) ORBIS database to a subset of LPP firms that can be identified in both data sets (n=217).[20] The ORBIS data set includes complete information on a firm's direct owner and historical linkages to a potential global ultimate owner (GUO) mainly sourced from company records and stock exchange releases and has been used in previous research on identifying control rights (Aminadav and Papaioannou (2020)).

We apply the definition of a global ultimate owner with a minimum ownership of 25% to denote significant ownership, which is very closely related to the widely-used definition of a controlled firm in the literature (e.g., see La Porta et al. (1999) and Aminadav and Papaioannou

[18] The definitions in the literature are quite heterogeneous. They range from different voting right cut-offs, governance characteristics, self-reports via survey items, to techniques that use Artificial Intelligence (see, e.g., La Porta et al. (1999); De Kok et al. (2006); Bloom and Van Reenen (2007); Cruz et al. (2011); Neckebrouck et al. (2018); Sánchez-Marín et al. (2019); Aminadav and Papaioannou (2020); Dal Maso et al. (2020); Peláez-León and Sánchez-Marín (2022); Amore et al. (2024)). Nonetheless, they all agree that a significant share of decision rights and the governance system play an important role, which matches our definition.

[19] All results remain robust when including other forms of ownership in the regressions.

[20] Also see Appendix B.2.

(2020)). The resulting share of family ownership amounts to 59% of the cases, which validates the use of our ownership variable from the LPP employer survey.

A unique feature of the LPP is the possibility to link both surveys to the annual IAB establishment panel and, as a result, retrieve information on the composition of the management (top executive) board of each establishment. Thus, we are able to focus on the management board as a whole, and not solely on the position of the CEO. Table 3.1 shows that in the case of family-owned enterprises, the management board exclusively consists of family representatives in the majority of cases (*Family management*), again with an increasing trend over time. However, the share of family-owned firms run exclusively by non-family (i.e., external) managers (*Non-family management*), is quite stable over time, amounting to 21.2% in 2020. At the same time, the share of companies with a management board run by a combination of family members and non-family managers (*Mixed family and non-family management*) decreases.[21] Our descriptive results on the types of corporate ownership and management board composition are broadly in line with previous literature, although the exact definition of family-owned firms and the type of management might differ across studies. La Porta et al. (1999), for instance, report that 40% of medium-sized companies in Europe are family-owned. More recently, Aminadav and Papaioannou (2020) report family ownership (families and unidentified private owners, which are likely also families or individuals) for approximately 51.5% of their German sample firms. In addition, Broszeit et al. (2019) analyze a sample in the German manufacturing industry and identify about 61% of the enterprises as family-owned. Using WMS data from 2004 to 2010 (Bloom et al. (2012)) and applying a similar classification procedure to ours reveal that family ownership in Germany amounts to approximately 58%, of which 44% of family-owned firms also have a CEO of the same family and 20% an external CEO.

[21] All in all, changes in the type of management in family-owned companies are also quite stable over time, particularly for family-owned companies with family management (Panel B of Table B.1 in Appendix B.1). For instance, in roughly 95% of the cases, family-managed family firms do not change their status throughout our eight-year time window, and only 1.8% of family-managed family firms switch to non-family-managed family firms and 3.4% switch to a mixed management board type. These numbers are consistent with prior research. Kölling (2020), for instance, shows that changes to family-managed firms and to non-family-managed firms (although slightly differently measured as in our case) also occur very rarely using German administrative data., i.e., occur in less than three percent of the observations. Similarly, Franks et al. (2012) emphasize the high persistence of family control over time, particularly in the case of German firms. Sraer and Thesmar (2007) identify only 21 observations for French family enterprises whose control is transmitted to a professional CEO, corresponding to 8 such transition cases from family management to non-family management in our paper.

Table 3.1: Distribution of Ownership and Management Structure (in Percent)

Ownership and management structure	2012	2014	2016	2018	2020
Family-owned firm	64.7	61.6	69.6	72.0	73.0
thereof with					
Family management	55.5	60.7	61.1	63.6	65.3
Non-family management	21.3	20.7	17.4	20.9	21.2
Mixed family and non-family management	23.2	18.6	21.6	15.5	13.5
Management-owned firm	22.7	27.5	18.9	17.5	18.3
Dispersed ownership (widely-held firms)	12.6	10.9	11.5	10.4	8.8
Number of firms	788	566	589	525	476

Notes: Weighted results. State-owned enterprises and other forms of ownership excluded. $N = 2,944$.

Management practices. We can differentiate between a series of (strategic) management practices along the value chain, which are grouped into three distinct management sub-scores (management indices): *personnel planning and recruiting*, *performance management and monitoring*, and *employee development and participation*. This classification of management practices has been widely used in past research and is closely related to the ability-motivation-opportunity model of high-performance/involvement work systems (e.g., Boselie et al. (2005); Subramony (2009); Wu and Chaturvedi (2009); Sánchez-Marín et al. (2019); Hauff et al. (2022); Peláez-León and Sánchez-Marín (2022)). The first sub-score, *personnel planning and recruiting*, includes the use of personnel forecast plans, a systematic analysis of the age structure of the workforce, the use of headhunters or consultancies for poaching employees from competitors, and the use of social networks for addressing applicants from competitors. The second sub-score involves items that deal with incentivizing and monitoring the workforce. It includes the use of performance-related (variable) pay schemes, performance appraisal systems, recommended distributions, and formal (i.e., written) target agreements. Finally, the third sub-score summarizes items that are important for the development and participation of incumbent workers. It includes the use of annual, pre-scheduled employee feedback talks, career development plans, and regular employee surveys.

Each sub-index is determined by calculating the total sum of all applied management practices per establishment, i.e., the total sum of all sub-index-specific dummy variables. Hence, all management practices are equally weighted in the calculation of overall scores. To make the results for the different sub-scores comparable to each other in the empirical analyses, we computed the z-score for each sub-index with zero mean and standard deviation of one. Table 3.2 gives an overview of the different management practices, including the original wording of the survey item, the number of observations, the mean, and the standard deviation.

Table 3.2: Categorization of Management Practices (Unstandardized, Establishment level)

Practice	Questionnaire Item	Mean	Std. dev.
	In your **establishment**,...		
Planning and recruiting		**1.992**	**1.185**
Personnel forecast	... have you implemented a personnel forecasting plan/staffing plan?	0.658	0.474
Age structure analysis	... do you systematically analyze the age structure of your workforce?	0.634	0.482
Recruiting via headhunter	... have you recruited applicants in the past two years, who were employed by another company, with the help of a private recruitment agency or HR consultancy?	0.301	0.459
Address via social networks	... have you directly addressed applicants employed by another company via social networks such as Xing, LinkedIn etc. in the past two years?	0.399	0.490
Performance management		**1.862**	**1.150**
Variable pay scheme	... have you implemented a variable compensation scheme?	0.577	0.494
Performance appraisals	... have you implemented performance appraisals by the respective supervisor(s) which take place at least once a year?	0.614	0.487
Recommended distributions	... have you implemented recommended distributions, which contain information about the proportion of employees who should receive the best grade, the second-best grade, etc. (conditional on having performance appraisals in place)	0.072	0.259
Target agreements	... have you implemented written target agreements with your employees?	0.598	0.490
Development and participation		**1.487**	**1.078**
Employee feedback talks	... have you implemented formalized employee feedback talks which take place at least once a year?	0.682	0.466
Personnel development plan	... have you implemented personnel development plans?	0.429	0.495
Employee surveys	... have you implemented employee surveys on a regular basis?	0.376	0.484

Notes: The number of observations for each item and sub-score is 2,944.

Perceived job quality. To test our second and third hypotheses, we use the different items of employees' perceived job quality available via the LPP employee survey.[22]

The first group of variables includes two widely applied *global* measures of job quality, namely *job satisfaction* and *turnover intention* (Hamermesh (1977); Freeman (1978); Tett and Meyer (1993)). *Job satisfaction* reflects the current well-being of an employee and depends on the subjective evaluation of various aspects of their work, e.g., assigned tasks, perceived management practices, or interactions with colleagues (e.g., see Ganzach (1998); Gaertner (1999); Prottas (2013); Inceoglu et al. (2018); Clark et al. (2021); Artz et al. (2022)). It is a single-item construct measured on an 11-point Likert scale, where 0 means totally unsatisfied with the job and 10 corresponds to totally satisfied (the wording of the item is equivalent to the German SOEP survey, see Richter et al. (2013)). *Turnover intention* reflects an employee's willingness to leave the organization. It is elicited through a single item measured on a 5-point Likert scale ranging from 1 "I never think about leaving my current employer" to 5 "I think about leaving every day" (Tett and Meyer (1993)).

Compared to global measures, the second set of metrics consists of job quality items at the *facet level*, i.e., with a more narrow focus on specific aspects of the job. These domain-specific elements of job quality are likely to be directly influenced by formal and informal management practices. We focus on constructs such as *leadership quality*, *procedural fairness*, *goal clarity*, and *professional development*.

Procedural fairness characterizes employees' perceptions about the fairness and transparency of internal rules and regulations (Kim and Leung (2007); Wu and Chaturvedi (2009)) such as formalized performance evaluations (Frederiksen et al. (2017)), recommended distributions, and feedback talks. This single-item measure reads as follows: "The rules and procedures to make decisions are fair". It is measured on a 5-point Likert scale, where 1 corresponds to "Does not apply at all" and 5 to "Fully applies".

Secondly, we focus on the dyadic relationship between supervisors and subordinates (Prottas (2013); Moorman et al. (2013); Inceoglu et al. (2018); Hauff et al. (2022)). This relationship might be enforced in non-family-owned firms in a rather formal way, e.g., via annual employee feedback talks and performance appraisals. However, family-owned companies might lean more towards their stewardship culture and/or ability to impose implicit relational contracts, and instead use informal practices to deepen the supervisor-employee relationship. Then, formal practices may not necessarily be needed to the same extent. We construct the metric *leader-*

[22]For a detailed description of the applied scales (including internal consistency), see Kampkötter et al. (2016).

ship quality to pick up this association by taking the average of four items derived from the Organizational Climate Questionnaire by Patterson et al. (2005). These items reflect whether employees perceive that their supervisors have confidence in them, provide good guidance, show understanding of their subordinates, and clearly communicate goals and requirements to them. Each item is evaluated on a 5-point Likert scale, where 1 corresponds to "Does not apply at all" and 5 to "Fully applies". The internal consistency of the four-item construct measured via Cronbach's alpha is high (0.83).

Thirdly, from a performance management perspective, it appears vital that employees understand the company's goals and that firms define their targets accordingly. This can be achieved by rather formal means to communicate targets or by developing informal practices that help achieve the objectives (e.g., Locke and Latham (1990, 2002, 2019)). Again, the latter case of informal mechanisms might be especially prominent in family-owned firms, e.g., as implicit contracts and a prevailing firm culture can help align goals for all employees. Hence, we collect information on whether employees perceive the firm's overall goals as clear and if they are aware of these goals. We take two items (awareness and a good understanding of long-term goals) of the Clarity of Organizational Goals scale borrowing from the Organizational Climate Questionnaire by Patterson et al. (2005), measured on the same 5-point Likert scale as *procedural fairness*. These two items are then averaged to an index of *goal clarity*. Cronbach's alpha reports an internal reliability of 0.72.

Lastly, we investigate the employees' assessment of personnel advancement opportunities. Formal employee development and participation practices, such as personnel development plans, might be particularly beneficial in this respect. In case non-family-owned firms more commonly implement these practices, they might exhibit improved scores. However, family-owned companies could perceive informal and flexible training for staff members as more suitable (Kotey and Folker (2007)). Therefore, we analyze the single item *professional development*, based on the survey question "Our establishment is palpably interested in further development of my professional knowledge and competencies". Again, similar to the previous items, it is measured on a 5-point Likert scale.

3.4 Statistical Analysis

To test hypothesis 1, we present the results of OLS regressions at the establishment level using all survey waves from 2012 to 2020. In our baseline specification, the key dependent variable is the z-score of each of the three management indices *planning and recruiting*, *performance management*, as well as *development and participation*. Our main independent variable is the ownership and/or management structure of the firm which will be analyzed in three steps.

For each step, we follow previous studies and define firms with dispersed ownership, i.e., financial investor- and stock market-owned firms, as reference group (see, for instance, Sraer and Thesmar (2007); Bassanini et al. (2013); Lemos and Scur (2019)).[23] In the first and second step, we will make use of our three distinct ownership categories (family-owned firms, management-owned firms, widely-held firms), which we collapse into different dummy variables. Step one consists of a single dummy variable that compares family-/management-owned firms, i.e., closely-held enterprises, against the reference category. We first summarize these two types of ownership as they share a closer distance to the business and its employees compared to widely-held firms. In step two, we differentiate between family-owned and management-owned firms to determine if family-owned companies genuinely differ from other firms, including those where owners are close to the business. Finally, step three categorizes the group of family-owned businesses into three types: those managed by family members, those led by non-family members, and those with a combination of family and non-family management, potentially emphasizing the previously discussed management variations.

We include a rich set of establishment-level control variables: information on the workforce composition measured by the proportion of female and part-time employees, the use of limited work contracts (1/0), establishment age, single-establishment (vs. being a multi-establishment firm, 1/0), competitive strategy (cost leadership, benefit leadership, both strategies), a dummy whether a change in the CEO position has occurred within the previous two years, whether the management of the HR function is represented in the top executive board of the company (vs. HR management is represented below the top executive board, 1/0), as well as industry sector (5 dummies), firm size (5 dummies), region (4 dummies), and year fixed effects. As previous studies found differences in organizational performance between family-owned firms and firms with dispersed ownership, we also control for the annual profit situation of the establishment in the previous fiscal year (i.e., whether net profits were positive, negative, or balanced). As some management practices are more targeted on certain types of workers, such as higher-skilled employees, we also control for the skill composition of the workforce by including the proportion of qualified and highly qualified employees in an establishment.

To test hypotheses 2 and 3, we use combined employer-employee level data from 2012 to 2021. Our OLS regressions use z-scores for *job satisfaction*, *turnover intention*, *leadership quality*, *procedural fairness*, *goal clarity*, and *professional development*, with mean 0 and standard deviation 1, as dependent variables. The principal independent variable remains unchanged from that used in hypothesis 1, including the stepwise analytical structure for hypothesis 2.

[23] To enable comparisons between family-owned firms and other types of ownership/management, we provide t-tests of the equality of coefficients at the bottom of the respective regression tables, which allow comparing the statistical difference between them.

Table 3.3: Differences between Family-owned Firms and Firms with Dispersed Ownership

	Family-owned		**Dispersed**		**T-Test**
	Mean	(SD)	Mean	(SD)	
Firm-level variables					
Industry					
Manufacturing	0.336	(0.473)	0.313	(0.464)	
Metal industries, electronics, vehicle construction	0.304	(0.460)	0.342	(0.475)	
Trade, transport, communication	0.181	(0.385)	0.123	(0.329)	***
Business and financial services	0.120	(0.325)	0.132	(0.339)	
Other services	0.059	(0.237)	0.089	(0.286)	**
Establishment size					
Less than 50	0.029	(0.168)	0.011	(0.105)	**
50-99 employees	0.388	(0.487)	0.217	(0.413)	***
100-249 employees	0.346	(0.476)	0.345	(0.476)	
250-499 employees	0.159	(0.366)	0.217	(0.413)	***
500 and more employees	0.078	(0.268)	0.210	(0.408)	***
Previous fiscal year's net profits					
Positive	0.843	(0.364)	0.776	(0.417)	***
Negative	0.079	(0.270)	0.123	(0.329)	***
Balanced	0.078	(0.268)	0.101	(0.301)	
Single-establishment	0.864	(0.343)	0.519	(0.500)	***
Change of management in previous 2 years	0.182	(0.386)	0.414	(0.493)	***
HR management at executive board level	0.478	(0.500)	0.396	(0.490)	***
Share of qualified staff	0.634	(0.254)	0.633	(0.234)	
Share of high-qualified staff	0.078	(0.118)	0.153	(0.179)	***
Share of women	0.302	(0.238)	0.301	(0.234)	
Share of part-time contracts	0.145	(0.200)	0.120	(0.169)	**
Non-permanent work contracts	0.693	(0.462)	0.808	(0.395)	***
Establishment age	22.73	(6.83)	21.19	(7.49)	***
Individual-level variables					
Job satisfaction	7.46	(1.78)	7.33	(1.82)	***
Turnover intention	1.63	(0.943)	1.61	(0.911)	
Leadership quality	3.69	(0.850)	3.63	(0.812)	***
Procedural fairness	3.40	(0.986)	3.36	(0.982)	**
Goal clarity	3.52	(1.21)	3.48	(1.17)	
Professional development	3.42	(1.27)	3.43	(1.24)	

Notes: For definition of variables, see Section 3.3.2. Firm level: Family-owned firms: $n = 1,832$, firms with dispersed ownership: $n = 447$. Individual level: Family-owned firms: between $n = 8,048$ (procedural fairness) and $n = 8,127$ (job satisfaction), firms with dispersed ownership: between $n = 3,019$ (procedural fairness) and $n = 3,042$ (job satisfaction). Last column shows results of a two-sample t-test with equal variances. * p<0.10, ** p<0.05, *** p<0.01.

The particularities of the regression specification for hypothesis 3 will be discussed in Section 3.5.3. In addition to the controls discussed earlier, we include dummies for female, management position, blue-collar worker, fixed-term contract, part-time work, the survey method (CAWI/CATI), as well as the logarithm of monthly net income and age in years.

The descriptive statistics in Table 3.3 show that family-owned firms tend to be smaller than firms with dispersed ownership, are more likely to be single-establishment firms and report positive net profits in a higher number of cases. Furthermore, the proportion of highly-qualified employees is smaller in family-owned firms and they tend to be slightly older. At the individual level, employees in family-owned firms tend to be more satisfied with their job and perceive the quality of leadership and procedural fairness slightly better. As Table 3.4 documents, both in family-owned and firms with dispersed ownership, our measures of perceived job quality are slightly declining over time (and turnover intention increases, respectively). A descriptive overview of all variables used in the statistical analyses is presented in Table B.2 in Appendix B.3. In all regressions, standard errors are clustered at the firm level.

Table 3.4: Individual Outcome Measures Over Time

Individual-level outcomes	2012	2014	2016	2018	2020
Panel A: Family-owned firms					
Job satisfaction	7.62	7.51	7.50	7.21	7.18
Turnover intention	1.61	1.62	1.62	1.63	1.69
Leadership quality	3.79	3.70	3.76	3.53	3.49
Procedural fairness	3.49	3.40	3.47	3.28	3.25
Goal clarity	3.57	3.57	3.61	3.44	3.28
Professional development	3.56	3.49	3.52	3.13	3.12
Panel B: Dispersed ownership					
Job satisfaction	7.52	7.47	7.44	6.93	7.05
Turnover intention	1.56	1.54	1.59	1.64	1.81
Leadership quality	3.66	3.70	3.73	3.42	3.53
Procedural fairness	3.45	3.41	3.39	3.25	3.23
Goal clarity	3.48	3.52	3.62	3.32	3.38
Professional development	3.63	3.52	3.56	3.03	3.18

Notes: For detailed definitions of the variables, see Section 3.3.2.

3.5 Results

3.5.1 Management Practices and the Type of Management in Family-owned Firms

Table 3.5 reports our baseline establishment-level regression results for the three management sub-scores using the unbalanced panel data set including all types of firm ownership, as well as management types for family-owned firms.

As discussed earlier, our stepwise analysis begins with an examination of the combined group of family-/management-owned companies. As shown in columns 1, 4, and 7, these types of firms score, on average, significantly lower on all our three management sub-scores relative to the reference group of firms with dispersed ownership, controlling for our rich set of establishment-level controls such as establishment size and financial performance. The differences range from 0.22 to 0.36 standard deviations, indicating that closely-held firms are negatively associated with the professionalism of management practices of up to one third of a standard deviation. This also constitutes an economically significant relationship.

The results in columns 2, 5 and 8 indicate that splitting the initial group into management- and family-owned firms does not largely change this result. Both categories of ownership perform significantly lower on each management sub-score compared to widely-held firms. Variations span from 0.2 to 0.36 standard deviations, where family-owned enterprises perform, in absolute size, marginally better in *planning and recruitment* and slightly worse in *development and participation* compared to management-owned companies. Nevertheless, the Wald tests in the first row at the bottom of Table 3.5 show that for all sub-scores, the coefficients for management-owned and family-owned firms are not significantly different from each other. Hence, relative to dispersed ownership, it seems that it does not make much of a difference whether the firm is owned by a family or the management.

Columns 3, 6, and 9 in Table 3.5 analyze the findings in more detail, breaking down family-owned firms by the three different types of management. The results indicate that the lower management scores between family-owned companies and widely-held companies are primarily driven by family-owned enterprises that are exclusively run by family managers. These firms have significantly lower scores compared to the reference group, ranging up to one-half a standard deviation. For family-owned companies run exclusively by non-family managers, on the other hand, the differences compared to widely-held firms are smaller in magnitude (up to 20% of a SD) and statistically significant in only two of the three cases (for *performance management* and *development and participation*). The coefficient differences between family-owned enter-

prises with family management and family-owned enterprises with non-family management are statistically significant for all management practice categories (see Wald tests at the bottom of the table), pointing again to the decisive role of family-owned firms' management boards run exclusively by family managers. In addition, our results show that management-owned businesses and family-owned companies with a top management team composed of both family managers and non-family managers score only slightly better (i.e., less negative) than family-owned firms with family managers.

Besides, single-establishments (i.e., single-site firms) also show lower management scores on all dimensions, controlling, among others, for firm size. One explanation might be that establishments that belong to a parent company (holding) or conglomerate (so-called multi-establishment firms) might also be characterized by foreign ownership or are multinational enterprises operating in more complex business environments. Moreover, changes in the CEO position in the previous two years are associated with increased management sub-scores of 0.09 to 0.13 standard deviations.

Overall, the coefficient sizes are comparable to those documented by Bloom and Van Reenen (2007) and Lemos and Scur (2019). For instance, Lemos and Scur (2019) suggest that, when incorporating comprehensive control variables, family-owned enterprises with a family member serving as CEO exhibit management scores that are as much as one-third of a standard deviation lower compared to widely-held firms. Meanwhile, family-owned businesses with a non-family CEO have a management score that is approximately 0.12 standard deviations lower. In addition, by splitting the management score into its main components, they show that this seems to be a general relation and is not driven by a specific sub-area of management. Consequently, besides confirming hypothesis 1, our results contribute to earlier findings by showing that this pattern is driven not only by the CEO, but also by the entire management board and applies beyond operational-focused management practices. The conceptual framework offered by Lemos and Scur (2019) tends to align with our findings. It appears that the utility functions of family and non-family managers vary in important aspects. One such element explicitly used in their framework is the higher cost of disciplining employees (via formal controls) due to reputation concerns experienced by family managers.[24]

[24] As laid out in Section 3.2, differences in the implementation of management practices could be partially driven by productivity differences between the different firm types. In our baseline estimates, we already control for the previous fiscal year's profit situation. As a robustness check, we replicate our baseline results in Table 3.5 by additionally controlling for the logarithm of the value added of a firm as productivity proxy. Value added is defined as sales turnover of a firm minus intermediate inputs. The average log value added is 16.09 (median: 15.90) with a minimum of 11.67 and a maximum of 23.46. Table B.3 in Appendix B.4 shows the respective regression results. Our baseline estimates are broadly confirmed, indicating that the differences in the use of formal management practices between family-owned firms and firms with dispersed ownership are not driven by differences in average productivity levels.

Table 3.5: Determinants of Management Practices: Ownership Structure and Management Board

Dep. var.:	Planning and recruiting (std.)			Performance management (std.)			Development and participation (std.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dispersed ownership	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref
Family-/Management-owned	-0.219***			-0.360***			-0.295***		
	(0.052)			(0.053)			(0.059)		
Family-owned		-0.201***			-0.364***			-0.315***	
		(0.056)			(0.058)			(0.063)	
Management-owned		-0.255***	-0.283***		-0.361***	-0.388***		-0.266***	-0.284***
		(0.060)	(0.061)		(0.060)	(0.060)		(0.068)	(0.068)
Family-owned: non-family management			-0.064			-0.188***			-0.177**
			(0.065)			(0.070)			(0.073)
Family-owned: family management			-0.352***			-0.504***			-0.408***
			(0.070)			(0.068)			(0.074)
Family-owned: mixed management			-0.201***			-0.432***			-0.388***
			(0.074)			(0.074)			(0.079)
Single-establishment	-0.144***	-0.154***	-0.119**	-0.360***	-0.358***	-0.320***	-0.403***	-0.396***	-0.369***
	(0.047)	(0.048)	(0.047)	(0.048)	(0.048)	(0.048)	(0.050)	(0.051)	(0.052)
CEO change within last two years	0.127***	0.124***	0.100**	0.119***	0.121***	0.095**	0.089**	0.085**	0.066
	(0.042)	(0.042)	(0.042)	(0.039)	(0.039)	(0.039)	(0.041)	(0.042)	(0.042)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,944	2,944	2,944	2,944	2,944	2,944	2,944	2,944	2,944
# of firms	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560
R-squared	0.25	0.26	0.22	0.22	0.21	0.22	0.21	0.21	0.22
Wald tests (p-values):									
Family-owned = Management-owned		0.30			0.96			0.39	
Management-owned = Family: non-family man.			0.00			0.00			0.15
Management-owned = Family: family man.			0.27			0.06			0.06
Management-owned = Family: mixed man.			0.15			0.72			0.22
Family: non-family management = family man.			0.00			0.00			0.00
Family: non-family management = mixed man.			0.10			0.00			0.01
Family: family management = mixed man.			0.02			0.19			0.61

Notes: Standard errors clustered at the firm level in parentheses. Additional firm controls include: competitive strategy, hierarchical level of HR department, share of qualified employees, share of highly qualified employees, share of female employees, share of part-time employees, share of fixed-term contracts, annual net profits, works council, collective agreement, founded before year 1990, firm age. For detailed definitions of the variables, see Section 3.3.2. * p<0.10, ** p<0.05, *** p<0.01.

3.5.2 Perceived Job Quality in Family-owned Firms

As shown above, professional management practices play a less prominent role in motivating, developing, and controlling employees in family-owned firms as compared to companies with dispersed ownership. An explanation could be that family-owned businesses substitute formal practices with more informal measures and implicit norms. So far, related literature indeed suggests that employees in family-owned enterprises are compensated for a reduced usage of formal management practices with higher levels of job security (we confirm this finding, see Table B.4 in Appendix B.5). However, it remains unclear whether additional forms of substitution are present in family-owned companies and whether these are ultimately linked to poorer, similar, or higher perceived work quality, compared to firms with dispersed ownership. To shed light on this, we make use of our linked employer-employee data and regress employee outcomes that are associated with perceived job quality on ownership structure and management types, including our wide set of controls. As family-owned enterprises are typically smaller, we would expect a reduced power distance to management and a more personal culture to be present here. This implies that the size of the establishment constitutes an important control variable in the following regressions.

Table 3.6 presents the results of the previously discussed (standardized) behavioral measures of job quality. We start by investigating the relationship between our six constructs and ownership structure. In the first stage of our stepwise analysis we show that employees in closely-held firms rate, ceteris paribus, *job satisfaction* (0.06 SD, column 1), *leadership quality* (0.06 SD, column 7), and *procedural fairness* (0.08 SD, column 10) significantly higher than employees in firms with dispersed ownership. Similarly, employees in family-/management-owned firms show lower levels of *turnover intention* (column 4) and higher perceptions of *goal clarity* and *professional development* (column 13 and 16), but the coefficients are not significant at conventional levels. The division of closely-held companies into family-owned and management-owned categories demonstrates that all coefficients that were previously significant retain their importance and show a slight increase for family-owned companies. On the other hand, the estimates for management-owned businesses turn insignificant, except for *procedural fairness* (significant at the 10% level), and decrease marginally in absolute size. Nonetheless, the Wald tests in the first row at the bottom of Table 3.6 indicate that the coefficients for management-owned and family-owned enterprises do not differ significantly for any of the employee-level outcomes. Thus, whether companies are owned by a family or by management does not significantly alter the outcomes.

Investigating the type of management more closely yields that the higher associations between our behavioral measures of job quality and family-owned firms are significantly driven by

family-owned firms which are exclusively run by non-family managers. In detail, employees in non-family-managed family-owned companies show a 10% of a SD higher *job satisfaction* than their counterparts in widely-held businesses (column 3). They also rate the quality of their managers (12% of a SD, column 9), the fairness of internal procedures (12% of a SD, column 12), as well as perceptions about the clarity of goals (11% of a SD, column 15) and *professional development* (13% of a SD, column 18) higher than employees in firms with dispersed ownership. Again, regarding turnover intention, the coefficients point to a similar interpretation, but statistical significance is lacking (column 6). Investigating the Wald tests in the lower section of the table discloses that family-owned companies with non-family management are also significantly higher rated than family-managed family-owned enterprises in four out of six job quality scores (*leadership quality*, *procedural fairness*, *goal clarity*, *professional development*) and management-owned companies in three out of six cases (*job satisfaction*, *leadership quality*, *professional development*).

To conclude, our regression results partially support hypothesis 2. Despite the lack of formal practices, family-owned firms, compared to firms with dispersed ownership, on average exhibit a higher (or at least comparable) level of perceived direct and indirect job quality. This is particularly true for family-owned firms with non-family management. As these firms are not as much different from dispersed ownership firms in terms of applied formal management practices, but show significantly higher levels of employees' perceived job quality, our results suggest that they likely benefit from a unique and more informal organizational culture. Hence, we highlight an important caveat when assessing management quality in firms with different ownership structures. Results on the impact of ownership structures on the quality of management, proxied via the use of formal management practices, might be misleading, even in case the type of management is accounted for. The reason is the existence of informal practices, such as trust-based relationships and personalized interactions, that are likely to be captured by employees' perceptions of job quality. Our findings suggest that family-owned businesses can compensate for their comparative deficiency in formal management practices by cultivating a more effective informal workplace culture. This could result from a combination of professional management practices and the relational strengths and superior firm culture often found in family-owned businesses, creating a workplace that employees perceive as especially positive. Hence, the results provide indirect evidence of the implementation of informal management practices in family-owned companies.

Table 3.6: Firm Ownership and Perceived Job Quality

Dep. var.:	Job satisfaction (std.)			Turnover intention (std.)			Leadership quality (std.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dispersed ownership	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Family-/Management-owned	0.062**			-0.032			0.059*		
	(0.030)			(0.034)			(0.033)		
Family-owned firm		0.076**			-0.037			0.062*	
		(0.032)			(0.037)			(0.036)	
Management-owned		0.037	0.033		-0.023	-0.019		0.052	0.043
		(0.035)	(0.035)		(0.039)	(0.039)		(0.043)	(0.043)
Family-owned firm: non-family management			0.096**			-0.057			0.121***
			(0.038)			(0.043)			(0.040)
Family-owned firm: family management			0.040			-0.005			-0.003
			(0.043)			(0.047)			(0.054)
Family-owned firm: mixed management			0.077**			-0.037			0.031
			(0.038)			(0.045)			(0.041)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm and individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,251	13,251	13,251	13,246	13,246	13,246	13,191	13,191	13,191
# of firms	1,199	1,199	1,199	1,199	1,199	1,199	1,199	1,199	1,199
R-squared	0.05	0.05	0.05	0.10	0.10	0.10	0.05	0.05	0.05
Wald tests (p-values):									
Family-owned = Management-owned		0.21			0.66			0.80	
Management-owned = Family: non-family man.			0.09			0.34			0.09
Management-owned = Family: family man.			0.85			0.75			0.39
Management-owned = Family: mixed man.			0.24			0.67			0.80
Family: non-family management = family man.			0.18			0.25			0.02
Family: non-family management = mixed man.			0.60			0.65			0.02
Family: family management = mixed man.			0.32			0.47			0.48

Table continued on next page.

Table 3.6 continued

Dep. var.:	Procedural fairness (std.)			Goal clarity (std.)			Professional development (std.)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Dispersed ownership	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Family-/Management-owned	0.075**			0.053			0.048		
	(0.033)			(0.041)			(0.040)		
Family-owned firm		0.076**			0.049			0.053	
		(0.035)			(0.044)			(0.050)	
Management-owned		0.072*	0.063		0.059	0.048		0.038	0.024
		(0.039)	(0.040)		(0.048)	(0.049)		(0.050)	(0.050)
Family-owned firm: non-family management			0.122***			0.110**			0.127***
			(0.042)			(0.049)			(0.050)
Family-owned firm: family management			0.010			-0.028			-0.058
			(0.048)			(0.062)			(0.056)
Family-owned firm: mixed management			0.064			0.025			0.038
			(0.043)			(0.051)			(0.050)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm and individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,150	13,150	13,150	13,216	13,216	13,216	13,223	13,223	13,223
# of firms	1,197	1,197	1,197	1,199	1,199	1,199	1,199	1,199	1,199
R-squared	0.05	0.05	0.05	0.04	0.04	0.04	0.11	0.11	0.11
Wald tests (p-values):									
Family-owned = Management-owned		0.91			0.81			0.75	
Management-owned = Family: non-family man.			0.17			0.19			0.05
Management-owned = Family: family man.			0.24			0.16			0.14
Management-owned = Family: mixed man.			0.97			0.62			0.79
Family: non-family management = family man.			0.01			0.02			0.00
Family: non-family management = mixed man.			0.17			0.07			0.06
Family: family management = mixed man.			0.22			0.33			0.06

Notes: Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Further firm controls include: dummy whether the establishment is a single-establishment firm, hierarchical level of HR department, share of qualified employees, share of highly qualified employees, share of women, share of part-time employees, share of fixed-term contracts, sales, firm age dummy. Individual-level controls include: female, age, management position, blue collar dummy, fixed term contract, part-time, ln net income, survey method (CAWI/CATI).

3.5.3 Worker Heterogeneity in Perceived Job Quality

While the previous subsection highlights that, on average, family ownership and management are correlated with employee-level outcomes, we want to supplement these findings by addressing potential heterogeneity related to key employee characteristics. Specifically, we investigate whether various segments of the workforce, defined by their individual hierarchical position and average skill composition, are differently affected by the formalized management structures in family firms. As argued in hypothesis 3, we expect that the observed non-negative relationship between family ownership and perceived job quality is less pronounced for employees at higher hierarchical levels and for firms with higher average workforce skills, suggesting that less formalized structures may not be sufficient to meet the needs of highly-qualified and managerial employees.

Qualification level is operated by five dummies for quintiles of the distribution of highly-qualified employees in a firm labeled as *k-th quintile highly-qualified*. These dummies take the value 1, if the share of highly-qualified employees within a firm falls into the respective quintile, based on all firm observations in a given year, and 0 else. Note that the share of employees with high qualifications might also point to more complex tasks or a more complex business model. Managerial responsibility is proxied through an indicator variable *supervisory position*, which indicates whether an employee holds a supervisory position (1, if this holds true, and 0 otherwise).

We replicate our baseline regression in Table 3.6 while including interaction terms between our ownership variables and each of the two characteristics, individually in a separate regression. Table 3.7 displays the heterogeneity analysis with respect to the average skill composition of the workforce. The coefficients *Family-owned firm: non-family management* and *Family-owned firm: family management* display the association for the respective non-family- or family-managed family-owned companies within the first quintile of highly-qualified workers. Throughout the different job quality measures, the coefficients on *Family-owned firm: non-family management* remain positive but lose significance, except for *professional development*, which could reflect a power problem. Notably, the estimates for *Family-owned firm: family management* turn into significant positive ones in four of six cases. It appears that employees of family-run family-owned firms ranging in the lowest quintile of the distribution of the highly-qualified workforce perceive job quality as significantly superior compared to employees in widely-held firms within the same quintile. Therefore, the proposed substitution of formal management practices with informal approaches appears to be particularly successful for this group of workers. Further, investigating the interaction terms of *Family-owned firm: family management* with the remaining quintiles vastly disclose a significant negative relationship (or

positive, in case of *turnover intention*). This implies that employees working in family-managed family-owned firms from these upper quintiles rate, ceteris paribus, their perceived job quality significantly lower than those from the lowest quintile. Consequently, the findings indicate that a reduced degree of structured management practices can be detrimental to a workforce with relatively high skill levels, as such individuals typically depend on these practices more heavily. Thus, the proposed substitution via superior firm culture and/or informal management practices appears to be less effective for this part of the workforce.

Results from Table 3.8 show a similar pattern. In this case, the estimates for *Family-owned firm: non-family management* and *Family-owned firm: family management* illustrate the relationship between the different management styles of family-owned firms and the perceived job quality of employees who do not occupy a supervisory role. Be aware that in this analysis the reference group consists of non-supervisory employees employed by widely-held companies. Hence, the positive and significant coefficients for *Family-owned firm: non-family management* denote that, relative to the reference group, individuals occupying a comparable position within family-owned enterprises governed by an exclusively non-family management board report, on average, a superior perception of job quality. Examining managerial employees in non-family-managed family-owned companies, this correlation seems to be largely diminished (negative interaction terms). As a result, even within family-owned firms operated by non-family management, formal management practices tend to play an essential role at higher hierarchical levels.

In summary, even though family-owned companies are generally likely to successfully substitute the lack of structured management practices, this view offers an incomplete account of the situation. Our heterogeneity analyses suggest that family-owned firms should not completely forgo the implementation of formal management structures. This holds true for both types of family-owned firms, with family and non-family management, and specifically for those employees that are highly-skilled and/or hold a leadership position, as positions requiring more expertise (i.e., high-qualified and supervisory roles) will benefit from more formal management strategies. Lastly, the general non-negative associations for family-owned enterprises found in Section 3.5.2 together with the results of the heterogeneity analyses indicate that the quality of the workforce is more compressed in those firms than in widely-held enterprises.

Table 3.7: Heterogeneity Analysis: Share of Highly-qualified Employees (Quintiles)

Dep. var.:	Job satisfaction (std.) (1)	Turnover intention (std.) (2)	Leadership quality (std.) (3)	Procedural fairness (std.) (4)
Family-owned firm: non-family management	0.153	-0.088	0.118	0.195
	(0.140)	(0.142)	(0.143)	(0.135)
Family-owned firm: family management	0.280**	-0.181	0.321***	0.257**
	(0.125)	(0.121)	(0.106)	(0.129)
Family firm: non-family man. × 2nd quintile highly qualified	0.031	0.055	-0.156	-0.144
	(0.176)	(0.181)	(0.193)	(0.183)
Family firm: non-family man. × 3rd quintile highly qualified	-0.076	0.016	-0.044	-0.188
	(0.172)	(0.165)	(0.162)	(0.156)
Family firm: non-family man. × 4th quintile highly qualified	-0.079	0.085	-0.022	-0.033
	(0.149)	(0.154)	(0.153)	(0.151)
Family firm: non-family man. × 5th quintile highly qualified	-0.066	-0.004	0.094	-0.040
	(0.155)	(0.161)	(0.156)	(0.152)
Family firm: family man. × 2nd quintile highly qualified	-0.223	0.244	-0.540***	-0.429**
	(0.156)	(0.167)	(0.152)	(0.170)
Family firm: family man. × 3rd quintile highly qualified	-0.377**	0.247*	-0.441***	-0.348**
	(0.155)	(0.147)	(0.131)	(0.149)
Family firm: family man. × 4th quintile highly qualified	-0.289**	0.348**	-0.435***	-0.302**
	(0.134)	(0.148)	(0.131)	(0.149)
Family firm: family man. × 5th quintile highly qualified	-0.171	0.025	-0.204	-0.171
	(0.136)	(0.138)	(0.140)	(0.145)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Firm and individual controls	Yes	Yes	Yes	Yes
Observations	13,251	13,246	13,191	13,150
# of firms	1,199	1,199	1,199	1,197
R-squared	0.05	0.10	0.06	0.05

Table continued on next page.

Table 3.7 continued

	Goal clarity (std.) (5)	Professional development (std.) (6)
Family-owned firm: non-family management	0.170	0.336**
	(0.114)	(0.155)
Family-owned firm: family management	0.179*	0.220
	(0.103)	(0.143)
Family firm: non-family man. × 2nd quintile highly qualified	-0.294*	-0.313
	(0.172)	(0.195)
Family firm: non-family man. × 3rd quintile highly qualified	-0.174	-0.286
	(0.140)	(0.176)
Family firm: non-family man. × 4th quintile highly qualified	-0.110	-0.232
	(0.134)	(0.166)
Family firm: non-family man. × 5th quintile highly qualified	0.073	-0.148
	(0.140)	(0.178)
Family firm: family man. × 2nd quintile highly qualified	-0.498***	-0.376**
	(0.150)	(0.182)
Family firm: family man. × 3rd quintile highly qualified	-0.387***	-0.420***
	(0.136)	(0.163)
Family firm: family man. × 4th quintile highly qualified	-0.316**	-0.342**
	(0.130)	(0.161)
Family firm: family man. × 5th quintile highly qualified	0.017	-0.186
	(0.133)	(0.168)
Industry, region, and firm size F.E.	Yes	Yes
Year F.E.	Yes	Yes
Firm and individual controls	Yes	Yes
Observations	13,216	13,223
# of firms	1,199	1,199
R-squared	0.05	0.12

Notes: Reference category: firms with dispersed ownership. Qualification level is operated by five dummies for quintiles of the distribution of highly-qualified employees in a firm labeled as *k-th quintile highly-qualified*. These dummies take the value 1, if the share of highly-qualified employees within a firm falls into the respective quintile, based on all firm observations in a given year, and 0 else. Coefficients on "Family-owned firm: mixed management" and quintiles of share of highly-qualified employees suppressed due to space constraints. Further firm and individual controls: see Table 3.6. Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 3.8: Heterogeneity Analysis: Employees in Supervisory Positions

Dep. var.:	Job satisfaction (std.) (1)	Turnover intention (std.) (2)	Leadership quality (std.) (3)	Procedural fairness (std.) (4)
Family firm: non-family management	0.122***	-0.087*	0.166***	0.158***
	(0.042)	(0.047)	(0.043)	(0.047)
Family firm: family management	0.021	0.013	0.016	0.010
	(0.049)	(0.052)	(0.059)	(0.053)
Family firm: non-family man. × supervisory position	-0.083	0.097*	-0.151***	-0.119*
	(0.058)	(0.056)	(0.059)	(0.066)
Family firm: family man. × supervisory position	0.043	-0.060	-0.068	-0.005
	(0.059)	(0.061)	(0.060)	(0.067)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Firm and individual controls	Yes	Yes	Yes	Yes
Observations	13,251	13,246	13,191	13,150
# of firms	1,199	1,199	1,199	1,197
R-squared	0.05	0.10	0.06	0.05

Table continued on next page.

Table 3.8 continued

	Goal clarity (std.) (5)	Professional development (std.) (6)
Family firm: non-family management	0.158***	0.162***
	(0.053)	(0.057)
Family firm: family management	-0.011	-0.047
	(0.065)	(0.061)
Family firm: non-family man. × supervisory position	-0.159***	-0.114*
	(0.055)	(0.062)
Family firm: family man. × supervisory position	-0.061	-0.041
	(0.063)	(0.061)
Industry, region, and firm size F.E.	Yes	Yes
Year F.E.	Yes	Yes
Firm and individual controls	Yes	Yes
Observations	13,216	13,223
# of firms	1,199	1,199
R-squared	0.04	0.11

Notes: Reference category: firms with dispersed ownership. Coefficients on "Family-owned firm: mixed management" and quintiles of share of highly-qualified employees suppressed due to space constraints. Further firm and individual controls: see Table 3.6. Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

3.5.4 Robustness Tests

We supplement our analysis with a series of robustness tests. First, as documented in Section 3.3.2, we observe only few transitions between ownership status and board management. Nevertheless, in the spirit of Bassanini et al. (2013), we estimate a first-differences estimation, in which we also control for time-invariant unobserved heterogeneity across workers. The dependent variable, Δ *Job satisfaction*, is the change in standardized *job satisfaction* of worker i between two adjacent survey waves t and $t-1$. Changes in the ownership structure of family-owned firms are captured through a variable defined as Δ *Family-owned firm*, which indicates the state of family ownership in the survey wave t minus the state of family ownership in wave $t-1$. This variable may thus take the values 0 (no changes in ownership, reference category), $+1$ (family-owned firm in wave t but not in wave $t-1$) and -1 (family-owned firm in $t-1$ but not in t). We further control for changes in individual and establishment controls, i.e., all variables are calculated as differences between survey waves t and $t-1$. Table B.5 in Appendix B.6 reports the corresponding regression estimates that make use of these ownership transitions, exemplary for job satisfaction, our global indicator of work quality. When firms switch from non-family ownership to family ownership (Δ *Family-owned firm* $= 1$), job satisfaction increases by 0.13 standard deviations and vice versa when ownership status changes in the opposite direction. This confirms the positive association shown between family-owned firms and job satisfaction reported in baseline Table 3.6.

Second, in our individual-level baseline regressions, we abstain from controlling for formal management practices, because these management practices are likely to be the outcome of a firm's ownership structure and therefore tend to be endogenous and bad controls (see Section 3.5). To avoid a potential endogeneity problem between management practices and ownership status, we focus on the subsample of family-owned firms to analyze which management practices are correlated with our outcome variables. We include the same set of control variables as in our baseline specification and cluster standard errors at the firm level. The corresponding subsample estimates are shown in Table B.6 in Appendix B.7. Employees working in firms that score higher on management quality from the dimension *development and participation* show significantly higher values in the vast majority of outcome variables, whereas for practices from the dimension *planning and recruiting*, no significant results are observed. Finally, employees from firms scoring higher on *performance management* rate the clarity of organizational goals significantly higher. This emphasizes the strong importance of management practices tailored to career development and employee feedback for the perceived workplace quality of the workforce.

Our third robustness test tries to elaborate more on whether the observed differences in outcome variables between family-owned firms and firms with dispersed ownership are driven

by differences in employees' perceptions or rather reflect practice differences. For this test we make use of the fact that selected questions from the individual questionnaire are also mirrored, although slightly modified, in the employer survey. Regarding goal setting, the employer survey asks for the existence of formal, i.e., written, target agreements in the firm. In a regression with goal clarity (standardized) as the dependent variable, the respective coefficient on the existence of formal target agreements (independent variable) is 0.1013 (significant at the 1%-level). Hence, we observe a strong positive correlation between the existence of formal targets and an employee's assessment of goal clarity. The same is true for the existence of a career development plan, a question from the employer survey, where we also observe a positive correlation between firms employing this management practice and employees' ratings of professional development as outcome variable (effect size 0.1331, significant at the 1%-level). Importantly, additionally controlling for the existence of target agreements and career development plans does not change our baseline results. For instance, replicating column 15 of Table 3.6, while additionally controlling for formal target agreements, leads to a coefficient for *Family-owned firm: non-family management* of 0.1232 (significant at the 5%-level). Similarly, replicating column 18 of Table 3.6, while controlling for the existence of career development plans, again leaves the coefficient for *Family-owned firm: non-family management* almost unchanged (effect size 0.1389, significant at the 1%-level). This result indicates that if we compare two firms with equal (observable) characteristics, and importantly, both firms with formal target agreements (or career development plans) in place (ceteris paribus), then we still observe that employees in family-owned firms with non-family management rate goal clarity (professional development) significantly higher than their counterparts in firms with dispersed ownership. This suggests that these higher ratings of workplace quality we observe are not entirely driven by perceived differences in the use of management practices, but rather seem to point to a comparative advantage of family-owned firms with non-family management when it comes to workplace quality (as compared to firms with dispersed ownership and family-owned firms with family members on the management board).

In a last robustness test we address the ordinary nature of the individual-level dependent variables and replicate our baseline regressions using ordered probit models. As indicated in Table B.7 in Appendix B.8, all our results remain robust to this alternative specification.

3.6 Conclusion

Previous research has provided mixed evidence on the organizational performance of family-owned firms as compared to non-family-owned firms. The studies by Bloom and Van Reenen (2007, 2010), Bloom et al. (2012) and Bloom et al. (2019) offer some evidence that differences

can partly be driven by variations in the implementation of management practices. We contribute to the literature by shedding light on the differences in the use of a large variety of management practices between family and non-family-owned firms. We are not only able to identify family ownership but can also differentiate whether the whole management board is governed by family and/or non-family members. This allows us to develop a detailed understanding of a broader bundle of formalized management practices (planning and recruiting, performance management, and development and participation) and how these practices are used across different types of firms and industries. Moreover, and perhaps more critically, our study also significantly contributes to the literature by exploring the nuanced relationship between family ownership, management styles, and perceived work quality, which has been scarcely addressed in existing research. Unlike previous studies focusing on wages and job security, we delve into potential differences in work quality and corporate culture by analyzing a series of behavioral outcome measures such as job satisfaction, turnover intention, leadership quality, and procedural fairness, filling a gap in understanding how family-owned firms may offer a superior or comparable workplace through informal management practices. To accomplish our research goal, we use five waves of the Linked Personnel Panel (LPP) from 2012 to 2021, to which we merge additional information from the IAB Establishment Panel.

Our analyses show that family-owned firms apply less formal management practices than widely-held businesses. However, also management-owned companies apply significantly fewer practices. This holds for each of our three management sub-scores. In addition, we include detailed management information that reveals that family-owned firms governed exclusively by family members drive the correlations. They not only apply fewer management practices than widely-held businesses, but also report fewer practices than exclusively non-family-managed family-owned businesses. Next, we show that family-owned companies have a higher or at least a similar level of job quality compared to companies with dispersed ownership. Thereby, family-owned firms exclusively run by non-family managers perform particularly well. This gives some evidence for a family firm culture that increases the intrinsic motivation of employees and might partially substitute formal management practices. Lastly, our heterogeneity analyses reveal that these substitution efforts by family-owned firms are mainly effective for lower-skilled employees, as they seem to drive the higher perceived job quality. On the other hand, this approach appears less beneficial for better skilled workforces and those in leadership positions, who tend to benefit more from formal management structures.

The results suggest several important managerial implications for family-owned firms. First, family-owned firms should carefully evaluate the role of family culture as it can enhance intrinsic motivation among employees and serve as a partial substitute for formal management practices. Second, incorporating non-family managers into the governance structure appears to

enhance overall work quality, highlighting the value of a balanced governance approach. Third, while less structured management practices may suffice for workforces with lower skill levels, positions requiring more expertise (i.e., high-qualified and supervisory roles) will benefit from more formal management strategies. As a result, disregarding this significant heterogeneity may result in adverse outcomes and potential inefficiencies. Therefore, while family-owned companies can often substitute structured management practices, they should not completely abandon them. Finally, implementing targeted employee development and participation practices, such as regular feedback and development plans, tends to be most effective when it comes to optimizing perceived job quality.

Finally, our study is not without limitations. First, as outlined above, we are not able to distinguish between the family and the founder(s) as potential owners of a firm. Although previous research suggests that this might be less severe in the German context, future research should try to elicit this information separately. Second, our main variables of interest, the type of ownership and board management, are rather time-constant, which limits the use of estimation procedures relying on within-firm changes to establish causality. Closely related, we have limited ability to track changes in employment. When workers change employers, they typically fall out of our sample, hence within-employee variations of firm ownership and management are also observed rarely, limiting our findings on perceived job quality to cross-sectional evidence. Upcoming studies might be capable of detecting and monitoring such changes (more often), allowing more in-depth analyses. Third, while we know whether family-owned firms in our sample are managed exclusively by family or non-family members, we lack information on who of the family members exactly occupies which position in the board. Future research needs to investigate whether there is heterogeneity among family members in occupying management board seats and, if yes, whether this impacts important outcome variables such as perceived job quality.

Overall, our study confirms previous findings that family-owned firms apply fewer management practices, yet it demonstrates that the whole picture is more complex, especially with regard to employee-level outcomes. Further research is needed to disentangle the cause for differences in management practices and perceived employee-outcomes in family- and non-family-owned firms in more depth. Potential drivers may include more informal management in family-owned firms and/or the underlying firm culture, the latter of which has received particularly limited attention in prior research.

Appendix B

B.1 Transition Probabilities

Table B.1: Transition Probabilities of Ownership Status (in Percent)

Panel A: Changes in ownership status	Family-owned firm	Management-owned firm	Dispersed ownership
Family-owned firm	91.4	6.8	1.8
Management-owned	17.7	71.3	11.1
Dispersed ownership	2.7	18.9	78.4
Panel B: Changes in type of management within family-owned firms	**Non-family mgmt.**	**Family mgmt.**	**Mixed family and non-family mgmt.**
Non-family management	89.0	4.6	6.4
Family management	1.8	94.8	3.4
Mixed family and non-family management	6.9	10.1	83.1

Notes: Panel A shows transition probabilities in ownership status. Panel B shows transition probabilities in the type of board management for family-owned firms. The transition probabilities are calculated as changes in the respective categorical variables over time, i.e., from the year 2012 to the year 2020.

B.2 Classification of Ownership Structure

To validate our ownership information derived from the LPP employer survey, we use the Bureau van Dijk (BvD) historical ownership database 2022 Q4 data set. This dataset includes information on a firm's direct owner and historical linkages to a potential global ultimate owner (GUO). We retrieved the BvD IDs of all GUO 25 and GUO 50 companies and their subsidiary independence indicator for a subset of 217-226 establishments of our LPP firms. GUO 25 is defined as global ultimate owners with a minimum ownership of 25%, i.e., using the entity type of the establishment itself if the independence indicator is equal to "A+", "A", "A-" and "U". GUO 50 is defined as global ultimate owners with a minimum ownership of 50%, i.e., using the entity type "A+", "A", "A-", "B+", "B", "B-", and "U".[25] The final ownership type is then determined via the subsidiary independence indicator of the matched LPP establishments. This validation exercise shows family ownership in 59% of cases for the GUO 25 indicator and in 38% of cases for the GUO 50 indicator. This is broadly in line with our LPP employer survey measure on ownership structure and the numbers reported by previous literature as discussed in Section 3.3.2, which validates the use of our ownership variable from the LPP employer survey.

[25]The indicator "A+/-" is attached to any business in which there are known, registered shareholders, none of whom possess more than 25% of the business directly or overall. "B+/-" applies to any business having one or more shareholders with an ownership percentage above 25% but no known documented shareholders with a direct, total, or calculated ownership percentage above 50%. And "U" relates to any entity that is deemed to be "independent" by definition, such as states, persons, or inactive businesses.

B.3 Descriptive Statistics

Table B.2: Descriptive Statistics for Main Explanatory Variables

Variables	Mean	SD	Min	Max	N
Firm level					
Industry					
Manufacturing	0.32	0.47	0.00	1.00	2,944
Metal industries, electronics, vehicle construction	0.29	0.45	0.00	1.00	2,944
Trade, transport, communication	0.17	0.37	0.00	1.00	2,944
Business and financial services	0.15	0.35	0.00	1.00	2,944
Other services	0.08	0.26	0.00	1.00	2,944
Region					
North	0.19	0.39	0.00	1.00	2,944
East	0.35	0.48	0.00	1.00	2,944
South	0.19	0.39	0.00	1.00	2,944
West	0.26	0.44	0.00	1.00	2,944
Establishment size					
Less than 50	0.03	0.16	0.00	1.00	2,944
50-99 employees	0.37	0.48	0.00	1.00	2,944
100-249 employees	0.34	0.47	0.00	1.00	2,944
250-499 employees	0.16	0.37	0.00	1.00	2,944
500 and more employees	0.10	0.29	0.00	1.00	2,944
Single-establishment	0.78	0.42	0.00	1.00	2,944
Competitive strategy					
Cost leadership	0.06	0.23	0.00	1.00	2,944
Benefit leadership	0.47	0.50	0.00	1.00	2,944
Both strategies	0.47	0.50	0.00	1.00	2,944
Change of management in previous 2 years	0.23	0.42	0.00	1.00	2,944
HR management at executive board level	0.46	0.50	0.00	1.00	2,944
Share of qualified staff	0.63	0.26	0.00	1.00	2,944
Share of high-qualified staff	0.10	0.14	0.00	0.98	2,944
Share of women	0.31	0.24	0.00	1.00	2,944
Share of part-time contracts	0.15	0.20	0.00	0.99	2,944
Non-permanent work contracts	0.72	0.45	0.00	1.00	2,944
Collective agreement	0.54	0.50	0.00	1.00	2,944

Table B.2 continued

Variables	Mean	SD	Min	Max	N
Works council	0.57	0.49	0.00	1.00	2,944
Establishment age	22.17	7.02	0.00	31.00	2,944
Previous fiscal year's net profits					
Positive	0.83	0.34	0.00	1.00	2,944
Negative	0.09	0.28	0.00	1.00	2,944
Balanced	0.09	0.28	0.00	1.00	2,944
Founded before year 1990	0.53	0.50	0.00	1.00	2,944
Year					
2012	0.27	0.44	0.00	1.00	2,944
2014	0.19	0.39	0.00	1.00	2,944
2016	0.20	0.40	0.00	1.00	2,944
2018	0.18	0.38	0.00	1.00	2,944
2020	0.16	0.37	0.00	1.00	2,944
Employee level					
Female	0.26	0.44	0.00	1.00	12,820
Leadership position	0.30	0.46	0.00	1.00	12,820
Blue collar worker	0.40	0.49	0.00	1.00	12,820
Fixed term contract	0.96	0.21	0.00	1.00	12,820
Full-time	0.88	0.32	0.00	1.00	12,820
CATI	0.83	0.37	0.00	1.00	12,820
Logarithm of monthly net income	7.64	0.51	0.00	12.04	12,820
Age	47.88	10.61	19.00	77.00	12,820
Job satisfaction	7.37	1.82	0.00	10.00	12,820
Turnover intention	1.66	0.96	1.00	5.00	12,815
Leadership quality	3.64	0.85	1.00	5.00	12,760
Procedural fairness	3.37	0.99	1.00	5.00	12,720
Goal clarity	3.49	1.21	1.00	5.00	12,784
Professional development	3.38	1.27	1.00	5.00	12,792

B.4 Management Practices, Ownership & Board: Add. control *log value added*

Table B.3: Management Practices, Ownership Structure, and Management Board: Add. control *log value added*

	Planning and recruiting (std.)		Performance management (std.)		Development and participation (std.)	
	(1)	(2)	(3)	(4)	(5)	(6)
Dispersed ownership	ref.	ref.	ref.	ref.	ref.	ref.
Family-owned	-0.235***		-0.388***		-0.316***	
	(0.072)		(0.077)		(0.083)	
Management-owned	-0.302***	-0.327***	-0.447***	-0.472***	-0.351***	-0.364***
	(0.080)	(0.080)	(0.080)	(0.080)	(0.091)	(0.092)
Family-owned: non-family management		-0.121		-0.251***		-0.230**
		(0.084)		(0.090)		(0.096)
Family-owned: family management		-0.332***		-0.479***		-0.359***
		(0.087)		(0.088)		(0.095)
Family-owned: mixed management		-0.265***		-0.458***		-0.379***
		(0.096)		(0.099)		(0.105)
Log value added	0.118***	0.111**	0.082**	0.076**	0.065*	0.062*
	(0.036)	(0.036)	(0.033)	(0.033)	(0.034)	(0.034)
Single-establishment	-0.147**	-0.123**	-0.315***	-0.288***	-0.398***	-0.382***
	(0.062)	(0.062)	(0.068)	(0.068)	(0.074)	(0.075)
CEO change within last two years	0.145***	0.123**	0.142***	0.119**	0.088	0.075
	(0.053)	(0.053)	(0.051)	(0.051)	(0.055)	(0.055)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Additional firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,763	1,763	1,763	1,763	1,763	1,763
# of firms	958	958	958	958	958	958
R-squared	0.25	0.25	0.22	0.23	0.19	0.19
Wald tests (p-values):						
Family firm = Management-owned	0.30		0.36		0.64	
Family: non-family man. = Family: family man.		0.02		0.01		0.15

Notes: Standard errors clustered at the firm level in parentheses. Additional firm controls include: competitive strategy, hierarchical level of HR department, share of qualified employees, share of highly qualified employees, share of female employees, share of part-time employees, share of fixed-term contracts, annual net profits, works council, collective agreement, founded before year 1990, firm age. * p<0.10, ** p<0.05, *** p<0.01.

B.5 Ownership and Perceived Job Security

Table B.4: Firm Ownership and Perceived Job Security

	Job security (std.)		
	(1)	(2)	(3)
Dispersed ownership	ref.	ref.	ref.
Family-/Management-owned	0.142***		
	(0.043)		
Family-owned firm		0.182***	
		(0.044)	
Management-owned		0.066	0.065
		(0.053)	(0.053)
Family-owned firm: non-family management			0.173***
			(0.050)
Family-owned firm: family management			0.168***
			(0.053)
Family-owned firm: mixed management			0.209***
			(0.051)
Industry, region, and firm size F.E.	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes
Further firm controls	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Observations	12,813	12,813	12,813
# of firms	1,189	1,189	1,189
R-squared	0.07	0.07	0.05
Wald tests (p-values):			
Family firm = Management-owned		0.00	
Management-owned = Family: non-family man.			0.03
Management-owned = Family: family man.			0.04
Management-owned = Family: mixed man.			0.00
Family: non-family management = Family: family man.			0.92
Family: non-family management = Family: mixed man.			0.43
Family: family management = Family: mixed man.			0.38

Notes: For description of variables, see Table 3.6. Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Further firm controls include: dummy whether the establishment is a single-establishment firm, hierarchical level of HR department, share of qualified employees, share of highly qualified employees, share of women, share of part-time employees, share of fixed-term contracts, sales, firm age dummy. Individual-level controls include: female, age, management position, blue collar dummy, fixed term contract, part-time, income, survey method (CAWI/CATI).

B.6 First-Differences Estimation: Changes in Ownership and Job Satisfaction

Table B.5: First-Differences Estimation: Changes in Ownership and Job Satisfaction

	Δ Job satisfaction (std.)
No ownership change	ref.
Δ Family-owned firm	0.130***
	(0.043)
Changes in individual controls	Yes
Changes in firm controls	Yes
Observations	3,508
# of firms	548
R-squared	0.02

Notes: The table reports results of a first-differences estimation. The dependent variable, Δ *job satisfaction*, is the change in standardized *job satisfaction* of worker i between two adjacent survey waves t and $t-1$. Changes in the ownership structure of family-owned firms are captured through a variable defined as Δ *Family-owned firm*, which indicates the state of family ownership in the survey wave t minus the state of family ownership in wave $t-1$. This variable may thus take the values 0 (no changes in ownership, reference category), +1 (family-owned firm in wave t but not in wave $t-1$) and −1 (family-owned firm in $t-1$ but not in t). We further control for changes in individual and establishment controls, i.e., all variables are calculated as differences between survey waves t and $t-1$. For a description of the remaining firm- and individual controls, see Table 3.6. Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

B.7 Management Practices and Employee-level Outcomes: Subsample of Family-Owned Firms

Table B.6: Management Practices and Employee-level Outcomes: Subsample of Family-Owned Firms

Dep. var.:	Job satisfaction (std.) (1)	Turnover intention (std.) (2)	Leadership quality (std.) (3)	Procedural fairness (std.) (4)
Planning and recruiting (std.)	-0.016	0.028	-0.028	-0.016
	(0.016)	(0.019)	(0.019)	(0.019)
Performance management (std.)	-0.001	-0.029	0.032	0.014
	(0.017)	(0.023)	(0.021)	(0.022)
Development and participation (std.)	0.061***	-0.028	0.080***	0.061***
	(0.016)	(0.022)	(0.022)	(0.020)
Industry, region, and firm size F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Further firm controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
Observations	7,568	7,565	7,531	7,502
# of firms	757	757	757	754
R-squared	0.05	0.10	0.06	0.06

Table continued on next page.

Table B.6 continued

	Goal clarity (std.) (5)	Professional development (std.) (6)
Planning and recruiting (std.)	0.000	0.008
	(0.020)	(0.019)
Performance management (std.)	0.064***	0.017
	(0.023)	(0.022)
Development and participation (std.)	0.085***	0.109***
	(0.024)	(0.022)
Industry, region, and firm size F.E.	Yes	Yes
Year F.E.	Yes	Yes
Further firm controls	Yes	Yes
Individual controls	Yes	Yes
Observations	7,543	7,554
# of firms	757	757
R-squared	0.06	0.13

Notes: Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Further firm controls include: dummy whether the establishment is a single-establishment firm, hierarchical level of HR department, share of qualified employees, share of highly qualified employees, share of women, share of part-time employees, share of fixed-term contracts, sales, firm age dummy. Individual-level controls include: female, age, management position, blue collar dummy, fixed term contract, part-time, ln net income, survey method (CAWI/CATI).

B.8 Ordered Probit regressions

Table B.7: Ordered Probit: Firm Ownership and Employee-level Outcomes

	Job satisfaction		Turnover intention		Leadership quality		Procedural fairness	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dispersed ownership	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Family-owned	0.084**		-0.058		0.079**		0.078**	
	(0.035)		(0.047)		(0.037)		(0.038)	
Management-owned	0.045	0.041	-0.031	-0.026	0.055	0.047	0.073*	0.064
	(0.038)	(0.038)	(0.049)	(0.049)	(0.042)	(0.043)	(0.042)	(0.042)
Family-owned: non-family management		0.099**		-0.082		0.126***		0.122***
		(0.042)		(0.057)		(0.043)		(0.043)
Family-owned: family management		0.051		-0.025		0.024		0.006
		(0.047)		(0.058)		(0.056)		(0.051)
Family-owned: mixed management		0.090**		-0.052		0.056		0.075*
		(0.043)		(0.058)		(0.042)		(0.045)
Observations	12,818	12,818	12,813	12,813	12,759	12,759	12,719	12,719
# of firms	1,189	1,189	1,189	1,189	1,189	1,189	1,187	1,187
Pseudo R-squared	0.01	0.01	0.05	0.05	0.01	0.01	0.02	0.02

Table continued on next page.

Table B.7 continued

	Goal clarity		Professional development	
	(9)	(10)	(11)	(12)
Dispersed ownership	ref.	ref.	ref.	ref.
Family-owned	0.053		0.055	
	(0.044)		(0.047)	
Management-owned	0.062	0.052	0.040	0.024
	(0.048)	(0.048)	(0.054)	(0.054)
Family-owned: non-family management		0.113**		0.130**
		(0.049)		(0.057)
Family-owned: family management		-0.023		-0.065
		(0.064)		(0.060)
Family-owned: mixed management		0.031		0.050
		(0.052)		(0.056)
Observations	12,783	12,783	12,791	12,791
# of firms	1,189	1,189	1,189	1,189
Pseudo R-squared	0.01	0.01	0.03	0.03

Notes: Standard errors clustered at the firm level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Further controls include: industry, region dummy, size dummies, dummy whether the establishment is a single-establishment firm, hierarchical level of HR department, share of qualified employees, share of highly qualified employees, share of women, share of part-time employees, share of fixed-term contracts, sales, firm age dummy, year. Individual level controls include: female, age, management position, blue collar dummy, fixed term contract, part-time, income, survey method (CAWI/CATI).

Chapter 4

ESG Incentives in Executive Compensation[‡]

We study whether the growing use of ESG performance metrics in executive compensation creates meaningful incentives to improve ESG outcomes. Using granular data on compensation design and realized performance at large European firms, we show that ESG-linked pay often violates necessary conditions for strengthening ESG incentives. Specifically, we find that ESG metrics either carry immaterial weight in pay calculations or are largely insensitive to ESG performance. ESG metrics are also not tailored to executives' specific responsibilities but concentrate in highly visible positions. Overall, ESG-linked pay appears largely symbolic.

Key Words: Executive Compensation, ESG, ESG metrics, ESG Contracting, CSR Contracting, Sustainability, Incentive Contracting, Optimal Contracts

JEL classification: G30, G35, J33, M12, M52

[‡]This chapter is based on joint work with Matthias Efing, Patrick Kampkötter, and Raphael Moritz. The current working paper version is available at `http://dx.doi.org/10.2139/ssrn.4974204` (see Efing et al. (2026)).

4.1 Introduction

"*ESG metrics are now one of the most prevalent metrics in executive incentive plans.*"
Willis Towers Watson (2023)

In recent years, there has been a growing trend of incorporating Environmental, Social, and Governance (ESG) metrics into executives' performance-linked compensation plans, as observed by many industry experts and scholars.[26] Typical ESG metrics include, for example, emissions, working conditions, workforce diversity, or employee satisfaction. In this paper, we ask whether the inclusion of such ESG metrics in compensation plans creates genuine incentives for executives to improve ESG performance.

The main challenge in answering this question is endogeneity. For instance, if a firm introduces ESG metrics into executive pay and subsequently exhibits improved ESG performance, this need not reflect a causal effect of stronger ESG incentives. The adoption of ESG-linked pay could coincide with a range of ESG-relevant developments in regulation, accounting standards, technological innovation, and investor and consumer preferences, among others. In this paper, we sidestep this endogeneity problem, as we do not attempt to estimate the effect of ESG-linked pay on ESG performance. Instead, we show that the design of ESG-linked pay often violates conditions that are necessary for such contracts to strengthen incentives. Our analysis suggests that the inclusion of ESG metrics in the compensation plans of large listed European firms remains largely symbolic.

As an example, consider an executive who receives only a cash bonus that is initially linked to a target for EBITDA growth. The firm pays one million for on-target performance; but will scale this target bonus by a factor f if EBITDA growth deviates from the target:

$$Bonus = 1,000,000 \times f(EBITDA)$$

The scaling factor f is called the target achievement rate. It is monotonic and normalized to one (100%) for on-target performance. Otherwise, f can take any functional form. Now suppose that the firm revises the bonus plan and introduces an additional performance metric: pay now depends both on a target for EBITDA growth and on a target for reductions in CO_2 emissions:

$$Bonus = 1,000,000 \times (w_1 \times f_1(EBITDA) + w_2 \times f_2(CO_2)),$$

where w_1 and w_2 are non-negative weights, summing to one. The bonus now depends on a

[26] See for example, PricewaterhouseCoopers (2021); Willis Towers Watson (2023); Cohen et al. (2023); Rajan et al. (2023); Michaely et al. (2024); Agarwal et al. (2025); Hazarika et al. (2025); Gantchev et al. (2025).

weighted average of target achievement rates f_1 and f_2 for EBITDA growth and emissions.

Whether this revised bonus plan increases the executive's incentives to reduce CO_2 emissions depends on many factors. For now, we focus on one necessary condition: the sensitivity of the bonus to emissions must not be zero. Assuming that EBITDA does not depend on emissions in this example, the bonus sensitivity to emissions equals:

$$\frac{\partial Bonus}{\partial \mathrm{CO}_2} = 1,000,000 \times w_2 \times \frac{\partial f_2}{\partial \mathrm{CO}_2} \neq 0$$

Clearly, this inequality is only satisfied if neither weight w_2 nor $\frac{\partial f_2}{\partial \mathrm{CO}_2}$ are zero.

Our paper evaluates this condition for ESG performance metrics in executive pay in general. Specifically, we analyze the weights of ESG relative to conventional criteria in executives' short-term incentive (STI) and long-term incentive (LTI) plans. Granular data on achievement rates f allow us to study performance sensitivities separately for different ESG and non-ESG components of pay. In a nutshell, our findings show that for the average executive, ESG performance criteria either carry immaterial weight or are insensitive to ESG performance changes. This does not imply that ESG considerations do not matter to executives at all. For instance, if we allow emissions to affect EBITDA in our example, the executive would still care about them.[27] Similarly, she could be intrinsically motivated to reduce emissions. However, explicit ESG targets with immaterial weight or low sensitivity are unlikely to generate meaningful *incremental* incentives.

We use a novel dataset with comprehensive information both on the ex-ante design of executives' compensation plans (such as numbers and weights of different performance metrics and target bonuses) and on realized performance (e.g., target achievement rates) and payout (e.g., salary, bonuses, etc.).[28] We collect these data for 73 constituent firms in Europe's leading stock indices, the EURO STOXX 50 and the STOXX Europe 50, between 2013 and 2020. Previous work has shown that the recent rise in the adoption of ESG-linked pay is more pronounced in Western Europe (including the UK) than in most other developed economies (Barontini and Hill (2024); Hazarika et al. (2025); Willis Towers Watson (2023)). Hence, our analyses likely reveal an upper bound on the relevance of ESG incentives in compensation plans worldwide. As compared to prior studies, we do not limit the analysis to CEOs but include all executive

[27] In this case, the bonus sensitivity to emissions becomes: $\frac{\partial Bonus}{\partial \mathrm{CO}_2} = 1,000,000 \left(w_1 \frac{\partial f_1}{\partial EBITDA} \frac{\partial EBITDA}{\partial \mathrm{CO}_2} + w_2 \frac{\partial f_2}{\partial \mathrm{CO}_2} \right)$. The incremental incentives of the explicit emissions target are still given by the term $w_2 \frac{\partial f_2}{\partial \mathrm{CO}_2}$.

[28] We also collect some data about the functional relationship between achievement rates f and ESG performance, such as performance hurdles and caps. However, ultimately, we need to estimate performance sensitivities empirically.

positions, considering 674 executives in total. Performance metrics are classified into ESG and non-ESG criteria according to the terminology of LSEG Data & Analytics.

Our analyses yield several important insights into the relative importance of ESG and non-ESG performance in executives' compensation plans. First, we find that firms do not always commit ex ante to the weights with which these metrics enter pay calculations. Instead, they often retain discretion to determine the relative importance of performance criteria only ex post—that is, after performance has been realized. Conceptually, the weight of such a discretionary metric is a random variable that introduces additional uncertainty, potentially harming incentives.

The following example illustrates this distinction between *binding* metrics with explicit weights and *discretionary* metrics without. Suppose a firm states that targets for EBITDA growth (non-ESG) and emissions reductions (ESG) enter the calculation of STI with weights of 60% and 15%, respectively. The remaining 25% is attributed *jointly* to two additional performance metrics—"progress on diversity" (ESG) and "setup of IT infrastructure" (non-ESG). In this example, the firm commits to the relative importance of ESG and non-ESG performance only for the first two metrics. Hence, we call them *binding*. By contrast, the diversity and IT criteria are only jointly evaluated, so the firm retains *discretion* over their relative importance. For instance, the firm could disregard the diversity target (ESG) and assign the full 25% entirely to IT performance (non-ESG).

We find that ESG metrics are more often included as discretionary metrics, whereas firms commit explicit weights typically to conventional performance metrics. For the average executive, all *binding non-ESG metrics* together enter with an explicit weight of 72% in the calculation of STI and 89% in the calculation of LTI pay.[29] By contrast, all *binding ESG metrics* together enter with explicit weights of, on average, less than 3% in both STI and LTI plans.

It seems unlikely that *binding ESG metrics* will generate meaningful incremental ESG incentives if their weights are, on average, 24 to 30 times smaller than those assigned to conventional metrics. Of course, if these ESG targets lack ambition and are easy to achieve, requiring little to no effort from executives, they might still somewhat strengthen ESG incentives. However, unambitious ESG targets constitute a limitation to the effectiveness of ESG contracting themselves, as the resulting improvements to ESG outcomes are likely marginal.

[29]We do not require firms to specify weights at the level of each individual performance metric. For our purposes, it is sufficient that firms commit explicitly to a clear aggregate weight on ESG performance relative to non-ESG performance. Hence, for each compensation plan, we calculate the combined explicit weight assigned to ESG metrics relative to non-ESG metrics, as well as the residual weight placed on discretionary pay that pools ESG and non-ESG criteria without a clear distinction regarding their relative importance.

If we are willing to assume that the weights of *binding ESG metrics* are too small to strengthen ESG incentives, should we then conclude that ESG-linked pay is ineffective in general? Not necessarily. ESG considerations could still matter through the *discretionary* pay components that pool conventional and ESG criteria without specifying their relative importance. In STI plans, these discretionary metrics receive a substantial *combined* weight for ESG and non-ESG together of, on average, 25% (in LTI plans, the corresponding weight is roughly three times smaller). Hence, to assess the overall incentive relevance of ESG-linked pay, we must analyze *discretionary* STI compensation in more detail and, in particular, whether it responds to ESG performance.

In our second analysis, we therefore estimate the sensitivities of discretionary STI and of other STI components to changes in ESG performance. In panel regressions, we use the achievement rates f for binding ESG, binding non-ESG, and discretionary metrics as dependent variables. This granular information allows us to decompose performance sensitivities across STI components and to test whether discretionary STI, as the only ESG-linked pay component with substantial weight, has potential to enhance ESG incentives.[30]

We regress the component-level achievement rates on different performance variables. Importantly, executive fixed effects control for cross-sectional variation in levels, which is uninformative about sensitivities, and restrict identification to within-executive *changes* in target achievement and performance. We begin by considering companies' stock returns—not as a measure of ESG performance, but to assess whether STI achievement responds to shareholder value creation.

Unsurprisingly, we find that overall target achievement (calculated as the weighted average of component-level achievement rates), as well as achievement for *binding non-ESG* metrics, is positively sensitive to stock returns. Interestingly, achievement rates for *binding ESG* metrics exhibit a negative sensitivity to stock returns, which could point to a potential trade-off between financial and ESG performance. Importantly, achievement for *discretionary* STI is again positively sensitive. Thus, with respect to stock returns, discretionary STI behaves more like STI tied to conventional performance criteria than like STI linked to binding ESG metrics.

To measure ESG performance, we first consider changes in firms' total CO_2 emissions equivalents and employee turnover, because we observe that most environmental metrics are related to

[30]Importantly, in data sets such as ISS Incentive Lab and Executive Compensation Analytics, target achievement rates and weights for different STI components are often missing. Furthermore, a systematic comparison of target values and actual outcomes (target achievement rates) across firms is difficult to implement, as this information, if available, is often disclosed in unstructured form.

emissions and most social metrics are related to companies' workforce.[31] Again, we find important differences across STI components. Binding ESG metrics exhibit negative and significant sensitivities to emissions and employee turnover, suggesting that executives receive lower STI payouts on these metrics when firms emit more CO_2 equivalents or employee turnover increases. By contrast, the sensitivities of STI linked to binding non-ESG metrics and of discretionary STI are insignificant.

It is perhaps not surprising that discretionary STI pay is not sensitive to emissions or employee turnover. These outcomes can be measured and observed relatively well, and we find that such verifiable metrics are indeed more common among binding ESG criteria. By contrast, discretionary STI tends to pool more qualitative ESG and non-ESG objectives that are difficult to verify.[32] Estimating sensitivities to hard-to-verify performance changes is challenging, particularly for small changes in performance. However, very large shocks should still have measurable effects on pay. For example, although an ESG objective such as "progress on diversity and inclusion" may be difficult to assess precisely, a major scandal—such as the loss of a large discrimination case in court—should be observable and plausibly affect target achievement.

To capture such large shocks to ESG performance, we estimate STI sensitivities to high-reach ESG news with negative press coverage in global media outlets. We find that overall target achievement (across all STI components combined) is negatively sensitive to these adverse ESG events. However, as before, it is primarily the binding ESG metrics that respond to negative ESG news. Contrary to our prediction, discretionary STI is largely insensitive, except in the case of governance news, where its sensitivity remains roughly three times smaller than that of binding ESG metrics.

Taken together, our estimates reveal a mismatch between the weights and performance sensitivities of different ESG metrics. *Binding* ESG metrics are the only pay component that responds systematically to changes in ESG performance, yet they carry only negligible weight in the average executive compensation plan. By contrast, *discretionary* STI represents the only ESG-linked pay component with substantial weight, but it appears largely insensitive even to major ESG controversies. Overall, ESG-linked pay is unlikely to contribute materially to executives' pay risk. To formalize this intuition, we calculate the within-executive variance of STI as a proxy for pay risk, and decompose it into variance shares attributable to different STI components.

[31] Arguably, within-company changes in employee turnover capture within-company changes in employee satisfaction in a way that can be verified and measured consistently.

[32] This observation aligns with those in Michaely et al. (2024).

As expected, the variance share of binding ESG metrics, as well as their covariance terms with other components, is negligible given the immaterial weight these metrics receive. By contrast, binding *non-ESG* metrics account for the largest share of total STI variance, nearly twice as much as discretionary metrics. By definition, we cannot decompose the discretionary STI component further into an ESG and a non-ESG part. However, we observe that the variance of discretionary STI is lower when a greater fraction of discretionary metrics is classified as ESG. As predicted, variation in STI is mostly driven by conventional (non-ESG) performance.

When firms design incentive pay, they should take into account executives' cost of effort for improving ESG performance. Specifically, firms should only use performance criteria that are not prohibitively costly to control and informative about executives' actions. Tying pay to outcomes that executives cannot meaningfully influence would increase compensation risk without motivating effort. We therefore expect firms to tailor ESG metrics to executives' tasks and responsibilities—for example, *employee satisfaction targets* for chief human resources officers (CHROs) and *clean-technology or emissions targets* for chief technology officers (CTOs). Surprisingly, we find no evidence of such tailoring. Fixed-effects regressions show that CHROs and CTOs are no more likely to face employee satisfaction or emissions metrics than the CEO of the same firm in the same year. More broadly, ESG metrics appear most frequently in the STI plans of *generalist* positions with a broad spectrum of tasks. As these generalists are more visible, firms might link their pay to ESG less for incentive reasons but to signal ESG engagement to external audiences.

Our findings raise the question of why firms design ESG-linked pay with such weak incentive power. One possible explanation is that firms are still learning how to structure ESG compensation, given that ESG-linked pay is a relatively recent phenomenon. To test this hypothesis, we analyze whether firms update and redesign STI packages in response to past ESG performance. Our results show that firms reallocate weight from discretionary to binding metrics following adverse ESG press coverage. However, it is less clear whether this adjustment reflects an attempt to strengthen ESG incentives or rather a broader effort to reduce discretion. Indeed, some regression specifications suggest that binding *non-ESG* criteria gain the most weight in these revisions.

Overall, our analyses suggest that ESG-linked pay in large European firms is unlikely to generate meaningful incremental ESG incentives relative to compensation plans based solely on conventional performance metrics. At the same time, we document important industry differences. In particular, firms in the financial sector emphasize discretionary ESG metrics, whereas energy- and emissions-intensive sectors appear more inclined to commit to binding ESG criteria with explicit weights. Possibly, ESG-linked pay serves a more symbolic function in the financial

sector, while it is more incentive-relevant in firms that also have a financial interest in improved ESG performance—for example, due to their direct exposure to energy and carbon costs.

4.2 Literature Review

A growing literature documents the rise of ESG (or corporate social responsibility (CSR)) contracting, specifically through the integration of non-financial sustainability and CSR metrics in executive pay contracts (e.g., Kolk and Perego (2014)). These terms describe "a firm's voluntary actions to manage its environmental and social impact and increase its positive contribution to society" (Khan et al. (2016), p. 1697).[33] Recent contributions find that ESG contracting is associated with firm-level outcomes such as long-term orientation, financial and social performance outcomes, and improvements in environmental and social initiatives (Hong et al. (2016); Flammer et al. (2019); Al-Shaer et al. (2023); Ikram et al. (2023)).

Most prior empirical studies measure the adoption of ESG metrics in executive pay with a binary variable that equals one if a firm adopts *at least one ESG metric* or a broader ESG-related compensation policy (Hong et al. (2016); Tsang et al. (2021); Aresu et al. (2023); Cohen et al. (2023); Hazarika et al. (2025); Al-Shaer et al. (2023); Ikram et al. (2023)). This extensive margin approach can document whether ESG pay exists, but cannot assess whether it is also relevant for the provision of meaningful incentives. We contribute to the literature by studying the intensive margin of ESG contracting. In particular, our hand-collected information on the ex-ante structure of compensation contracts enables us to distinguish between binding performance metrics - whose weights are fixed by the firm ex ante - and discretionary metrics, and to compare the relative weights that firms assign to ESG versus non-ESG criteria.[34] By combining these contractual weights with realized target achievement rates, we can assess whether ESG metrics create meaningful incremental incentives relative to conventional performance plans based on financial metrics.

Several papers test whether the adoption of ESG pay in CEO contracts is associated with

[33]ESG and CSR contracting have been used interchangeably so far. The term ESG is now used to group all sustainability-related issues (Khan et al. (2016)). More recently, Edmans (2024) has added the term "rational sustainability" to this debate.

[34]While Cohen et al. (2023) document median weights of ESG metrics in STI and LTI bonus plans, they do not report further analyses based on these weights. Bebchuk and Tallarita (2022) show in a cross-section of 2021 S&P 100 firms that only less than one-third of firms report ex-ante weights of ESG metrics. Similarly, Michaely et al. (2024) report summary statistics on ES weights and pay magnitudes for explicit ES incentive schemes, but primarily base their analysis on dummy variables distinguishing explicit and implicit ES pay. Specifically, they classify a scheme as explicit if it includes at least one quantitative target directly linked to executive compensation via a fixed formula, and as implicit if it relies solely on qualitative targets or discretionary links.

companies' ESG performance (e.g., Hong et al. (2016); Flammer et al. (2019); Al-Shaer et al. (2023); Cohen et al. (2023); Hazarika et al. (2025); Ikram et al. (2023); Barontini and Hill (2024)). These studies provide valuable insights, but need to overcome endogeneity problems. In particular, firms' adoption of ESG pay is unlikely to be exogenous to contemporaneous changes in regulation, accounting standards, technological innovation, investor and consumer preferences, and other factors that also directly affect ESG outcomes.

Instead of attempting to overcome this endogeneity problem directly, we take a fundamentally different approach. Specifically, we show that in most large European firms, ESG metrics either (i) receive immaterial weight in pay calculations, or (ii) are largely insensitive to ESG performance. Hence, two necessary conditions for generating meaningful ESG incentives are violated. We conclude that ESG-linked pay remains largely symbolic, possibly serving to signal ESG engagement to external stakeholders. Consistent with this interpretation, Gantchev et al. (2025) show that the inclusion of ESG metrics in executive compensation helps firms secure say-on-pay approval and reduce shareholder dissent, rather than materially altering managerial incentives.

Most large-scale studies use global databases such as ISS Incentive Lab or Executive Compensation Analytics. These databases are useful for documenting ESG adoption worldwide and analyzing broad cross-country patterns (e.g., Carter et al. (2023); Cohen et al. (2023); Hazarika et al. (2025); Barontini and Hill (2024)). However, as noted by Gantchev et al. (2025) and Michaely et al. (2024), these databases have important limitations: restricted company coverage, sparse categorization of metrics (particularly prior to 2018), widely missing data on weights, target values, and target achievement rates as well as a lack of coverage for executive positions other than the CEO. As a result, Cohen et al. (2023) explicitly call for *"[...] further access to the exact compensation vehicles, the relative weights attached to different performance metrics, and the use of discretionary bonus rules* (Cohen et al. (2023), p. 810)." We contribute detailed compensation data that addresses these limitations. Instead of relying on commercial databases, we hand-collect comprehensive information on the ex-ante design of compensation plans for 674 executives at the 73 largest listed European firms. Critically, our data include: (i) contractually assigned weights for binding ESG and non-ESG metrics, (ii) joint weights for discretionary components that pool ESG and non-ESG criteria, (iii) target achievement rates at the component level (binding ESG, binding non-ESG, and discretionary), and (iv) granular executive-level information across all executive positions in the sampled firms. This allows us to quantify the relative contribution of ESG and non-ESG metrics to overall pay risk and to uncover novel insights into heterogeneity in contract design across industries and executive positions.

While our geographic focus is narrower than global studies, our goal is to provide deeper insight into how ESG metrics are actually implemented in executive compensation contracts. Additionally, although previous work has shown that the adoption of ESG-linked pay is mostly prevalent in Europe during recent years (Barontini and Hill (2024); Hazarika et al. (2025); Willis Towers Watson (2023)), the vast majority of existing studies use US data (e.g., Flammer et al. (2019); Ikram et al. (2023); Grabner et al. (2024); Badawi and Bartlett (2024)). Hence, the evidence for European firms is still scarce. Our focus on large European firms addresses this gap and likely provides an upper bound on the incentive character of ESG-linked pay, given that ESG adoption is more prevalent in Europe than in most other developed economies. The fact that in this favorable setting, even ESG metrics contribute little to incentive provision suggests that the limitations we document may be even more pronounced outside of Europe.

Our findings also relate to several theoretical and empirical mechanisms in the literature. Michaely et al. (2024) show that explicit, quantitative ES-linked pay schemes with specific targets are most effective when ES performance is precisely measurable, while qualitative or discretionary schemes work better for hard-to-measure outcomes. Our distinction between binding and discretionary metrics provides empirical support for this mechanism: we find that binding ESG metrics are indeed more sensitive to verifiable ESG outcomes (emissions, employee turnover), whereas discretionary metrics, which often include softer ESG criteria, show little sensitivity even to major ESG controversies.

Recent theoretical work by Bonham and Riggs-Cragun (2024) analyzes principal-agent models where boards contract with managers to maximize a combination of financial and ESG outcomes, showing that ESG-contingent stock awards can align incentives when shareholders treat financial and ESG outcomes as complements. Geelen et al. (2025) model diverging pro-social preferences of owners and managers and show that variable pay dependent on both financial and social metrics can be inefficient, helping explain why ESG compensation constitutes a small fraction of total pay. Although not studying contract design directly, unlike us, Maeckle (2025) examines how managerial incentives—arising from financial performance pressures—affect firms' environmental performance.

Agarwal et al. (2025) demonstrate empirically that the pay-performance sensitivity for financial metrics decreases after ESG metric adoption, indicating a substitutionary relationship. Our variance decomposition extends these insights by showing that even when ESG metrics are present, they contribute minimally to pay risk, suggesting that the substitution effect documented by Agarwal et al. (2025) may be more symbolic than substantive in European firms. Finally, our analysis of contract revisions following negative ESG news contributes to understanding how firms update compensation design in response to performance shocks, comple-

menting work by Carter et al. (2023) on the role of say-on-pay voting in driving ESG adoption.

The literature has also emphasized the role of institutional characteristics in ESG contracting (Tsang et al. (2021)). Aresu et al. (2023) find that firms in countries with greater social and environmental regulatory pressures are more likely to adopt ESG contracting, moderated by internal corporate governance structures. Hazarika et al. (2025) show that a country's culture predicts ESG pay adoption and that mandatory ESG disclosure requirements strengthen the link between ESG contracting and outcomes. Recent papers also emphasize the pivotal role of sustainability committees in this context and the complementary relationship between the public disclosure of sustainability information and CSR contracting (Al-Shaer and Zaman (2019); Grabner et al. (2024)). We contribute to this strand by documenting substantial heterogeneity across industries and executive positions within a relatively homogeneous European regulatory environment. Our finding that energy-intensive sectors rely more on binding ESG metrics while financial firms favor discretionary approaches suggests that industry-specific exposure to ESG costs and benefits shapes contract design in ways that studies covering heterogeneous regulatory frameworks may not fully capture.

4.3 Data

We first describe the composition of our firm and executive samples. Then we describe the different data sources and information we will study in our empirical analysis and also provide a brief comparison of our dataset to commercial databases on executive pay. A detailed description of the notation and definitions of our main variables can be found in Section 4.4 and Appendix C.1.

4.3.1 Firm Sample

We consider the constituent firms of Europe's two leading stock market indices, i.e., the STOXX Europe 50, covering the largest firms in terms of free-float market capitalization regardless of currency, and the EURO STOXX 50, tracking the 50 largest firms in the eurozone. To be included in our analysis, we require firms to be in at least one of both indices for at least ten days between December 31, 2014 and December 31, 2020.[35] For all firms, we collect eight

[35]Hence, we discard South32 (8 days, 2015), Uniper (1 day, 2016), Alcon, M&G (both 1 day, 2019), and Siemens Energy (1 day, 2020). These firms entered the indices for short transitional periods, e.g., due to corporate restructuring and (de)mergers. More generally, when firms enter the stock indices as a result of a merger or acquisition, we do not include any predecessor companies that were not already index constituents before.

years of data (2013 to 2020) even if a firm enters or exits the stock indices during this sample period. The resulting firm sample comprises 73 distinct firms or 584 firm-year observations.[36]

In Panel A of Table 4.1, we report the composition of this firm sample by industry and headquarter country.[37] Most firm-year observations come from the financial services sector, followed by consumers and industrials & materials. Unsurprisingly, most firms are headquartered in Germany, the UK, and France. We provide summary statistics about firms' B/S and P&L statements, as well as their governance and ownership structure, in Table IC.1 of the Internet Appendix C.5.

4.3.2 Executive Sample

We consider all executives (across the entire C-suite) employed by the firms in our sample. However, we discard 762 executive-year observations for executives with an annual work period of less than 360 days to make sure that compensation remains comparable and is not distorted by intra-annual turnover resulting in few months of employment.[38] We end up with 2,560 executive-year observations or 674 distinct executives. This corresponds to a mean of 4 and a median of 3 executives per firm.

Table 4.1, Panel B reports our executive sample in more detail. The average executive is 54 years old and predominantly male, with average tenure in the firm of 6 years. CEOs account for 576 executive-year observations, followed by CFOs with 399 observations. CHROs and COOs are also relatively frequent (148 and 141 observations each). All other expert C-suite positions (e.g., chief marketing officer, chief sales officer, chief legal officer, etc.) are grouped as *other specialists* (416 observations). Executives heading the management of individual divisions of firms, i.e., executives responsible for certain product lines or geographical markets, account for 825 observations.

[36] To maintain a balanced firm panel, we discard 22 firms without available data in all sample years. Throughout our analysis, the reporting year refers to the calendar year in which the company's fiscal year ends.

[37] A complete list of our sample firms is provided in Table C.1 of Appendix C.2.

[38] In those cases, special contractual clauses and severance pay apply.

Table 4.1: Sample Composition

Panel A: Firm sample

Industry sector	Firm-year obs.
	(1)
Consumers (discretionary & staples)	120
Energy & Utilities	48
Financials	176
Industrials & Materials	112
Health	64
Information & Communic. Technology (ICT)	64
Total	584

Headquarter country	Firm-year obs.
	(2)
Australia	16
Denmark	8
Finland	16
France	104
Germany	136
Ireland	8
Italy	24
Netherlands	24
Spain	40
Switzerland	72
United Kingdom	136
Total	584

Panel B: Executive sample

	Exec.-year obs.	Mean	S.D.	P5	P25	P50	P75	P95
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age (years)	2,560	54	5	46	51	54	58	63
Tenure (years)	2,216	6	5	1	2	4	8	15
Female (1/0)	2,560	0.106	0.308	0	0	0	0	1
CEO	576							
CFO	399							
CHRO	148							
COO	141							
Others	1,264							
Chairman	23							
Divisional Head (Region/Segment)	825							
Other specialists	416							
Total	2,528							

Notes: This table describes the composition of our firm and executive samples. Panel A shows a breakdown of firm-year observations by industry sector (column 1) and headquarter country (column 2). Panel B reports summary statistics on executives' demographic characteristics and a breakdown of executive-year observations by position. Industry sectors are classified using the primary Global Industry Classification Standard (GICS) sector code from LSEG Data & Analytics. The original 11 GICS sectors are aggregated into six major sectors: Consumers, Energy & Utilities, Financials, Industrials & Materials, Health, and Information & Communication Technology (ICT). We hand-collect information on executives' age, gender, tenure, and position on the executive or management board. Executive positions include chief executive officer (CEO), chief financial officer (CFO), chief human resources officer (CHRO), and chief operating officer (COO). All other expert C-suite positions (e.g., chief marketing officer, chief sales officer, chief legal officer) are grouped as "Other specialists." Executives chairing a division, region, or product segment are classified as "Regional/Segment." Age and tenure are winsorized at the 1% level in each tail of the distribution.

4.3.3 Data on Realized Remuneration

Information on realized remuneration is provided by the international consulting and board advisory firm Mercer hkp/// group (now part of Mercer/Marsh McLennan). These data comprise executives' fixed salaries, annual short-term incentive (STI) plans, multi-year performance elements including deferred compensation, and long-term (typically equity-based) incentive (LTI) plans, as well as the demographic information in Table 4.1, Panel B. As a market leader for executive remuneration benchmarking and advisory, Mercer hkp/// group ensures high accuracy and comparability across countries and companies by adhering to the consultancies' International Compensation Disclosure Standard (ICDS) methodology (Mercer (2016)).[39]

4.3.4 Data on Compensation Design and Performance Achievement

We hand-collect information about the ex-ante design of executives' compensation plans from companies' annual reports, corporate governance reports, and compensation reports. First, we identify the metrics that firms use to evaluate performance and to calculate variable remuneration. We carefully distinguish between ESG and non-ESG performance metrics, based on the ESG framework provided by LSEG Data & Analytics (formerly Thomson Reuters Refinitiv), "one of the largest ESG content collection operations in the world" (LSEG (2024)).[40] According to this taxonomy, the "E", "S", and "G" categories are divided into the following 10 subcategories: Emissions, Innovation, Resource use (E), Community, Product responsibility, Human rights, Workforce (S), and Corporate Social Responsibility strategy, Management, Shareholders (G). These classifications can then be subdivided into 375 more detailed metrics. Table C.2 in Appendix C.3 provides an overview of exemplary metrics per category. We use this detailed taxonomy to classify non-financial metrics in executive compensation contracts into non-financial-ESG-related and non-financial-non-ESG-related metrics and to group them accordingly.

Second, we record the weights with which performance along the different metrics enters the calculation of incentive pay, according to the firms' public reportings. To the extent that firms report joint weights for several metrics together, we also record these joint weights. For example, a given STI plan might report a weight of 50% for a performance metric called EBITDA growth, a weight of 20% for Scope-1 emissions reduction, and a joint weight of 30% for two metrics called "progress on diversity, equity, and inclusion initiatives" and "leadership effectiveness

[39] Whenever the data of the pay consultancy is incomplete, we hand-collect the missing information from company reports following their methodology.

[40] This data has been used in more than 1,200 scientific articles, mainly in finance journals (Berg et al. (2020)).

or strategic execution quality."[41] In Section 4.4, we will provide notation along with a detailed discussion of how we define and aggregate these weights at the level of ESG versus non-ESG performance metrics.

Third, we record performance targets and target pay. For example, a given STI plan might report that the firm targets to pay out a bonus amounting to 150% of fixed base salary if the executive neither overachieves nor underachieves but exactly meets the targets set for the various performance metrics ex ante. Fourth, whenever they are reported, we also record the actual, realized target achievement rates (also called target fulfillment rates), again distinguishing between the achievement rates for ESG and non-ESG performance targets. Again, we refer notation and further discussion of realized achievement rates to Section 4.4.[42]

Finally, we record information on bonus caps and hurdles, whether directors / compensation committees reserve the right to scale STI at their discretion (e.g., via ex post multipliers), and whether this right was also exercised in a given year. Whenever possible, we collect this information for STI as well as LTI plans.

All compensation data is converted into Euro using year-specific conversion rates provided by the European Central Bank. Compensation elements such as pension schemes and fringe benefits (company cars, insurances) are not included in our analyses. For notation and definitions of our main variables, see Section 4.4 as well as Appendix C.1.

4.3.5 Comparison with ISS Incentive Lab and ECA

Alternative datasets, such as ISS Incentive Lab (IL) and Executive Compensation Analytics (ECA), also provide information about executive compensation. As noted by Gantchev et al. (2025), IL is limited by both its restricted company coverage and its sparse categorization of metrics, particularly prior to 2018. Although ECA provides comparatively more detail on metrics than IL, Michaely et al. (2024) document substantial data gaps: in their sample, about half of the weights and three-quarters of the target values for ES metrics are missing. Furthermore, a systematic comparison of information on target values and actual outcomes across firms is

[41] We record this information for STI and for LTI plans. As LTI plans are granted over multiple years, information on the number and weights of metrics in LTI plans always refer to the first year of granting such an LTI plan.

[42] Note that for several reasons, we do not report target values and realized achievement rates for LTI plans. First, LTI target values are reported in a minority of cases, with missing information being present particularly in earlier years. Additionally, LTI details are sometimes not specified at the individual level, but only for the entire executive team. Lastly, it often occurs that multiple LTI plans in the same firm overlap over time, as each of these plans have different vesting periods, which makes it difficult to determine LTI target achievement across time.

difficult to implement, as this information, if available, is often disclosed in unstructured form. For instance, goals might be described verbally in qualitative terms (e.g., "rating: good", "exceeded or achieved targets"), which makes a clear, quantitative linkage between ex-ante goals and ex-post achievement hard to establish. Additionally, after matching the ECA data to our sample firms, only about 6% of observations contain non-missing information on both the performance goal and the actual outcome for at least one metric (ESG or non-ESG) in STI plans at the company–year level. Finally, ECA reports compensation plans only at the company level, whereas our hand-collected dataset provides executive-level information, which is essential for analyzing executive-specific heterogeneity and the tailoring of incentives to positions. Still, information on actual compensation amounts between ECA and our dataset is highly consistent, with correlation coefficients > 0.9 for base salary and STI bonus and > 0.7 for total compensation. The remaining differences are mainly due to the timing of the payout structure of LTI plans.

4.3.6 Firm-level Data

Information from firms' balance sheets and income statements, as well as stock price information for the majority of firms in our sample, comes from LSEG Data & Analytics. Corporate governance variables, similar to the ones used in Cohen et al. (2023) and Hazarika et al. (2025), comprise, for example, the fraction of independent and female directors, and ownership composition such as the number of institutional investors and blockholders per firm-year.[43] Moreover, we include ESG-relevant, verifiable metrics such as total CO_2 emissions and employee turnover.

We further retrieve information from RepRisk, which provides firm-level ESG risk data based on negative news coverage and has been used in prior research (Gantchev et al. (2022); Derrien et al. (2025); Duchin et al. (2025); Fritsch et al. (2025); Dube et al. (2026)). RepRisk identifies and classifies material ESG-related incidents by screening daily over 100,000 public media sources and stakeholders (e.g., print and online media, social media, government bodies, regulators) and characterizes them along several dimensions, for instance, their reach and their sharpness. As the data capture observable negative ESG outcomes, they provide timely signals of firms' ESG practices and related risks. For our analysis, we focus on news events that reach a large audience through global media outlets and that explicitly mention the name of the firm and precisely specify the underlying ESG incident. To match our annual compensation data, we aggregate the daily RepRisk ESG incidents to the firm-year level. Similar to Gantchev et al. (2022), we study the impact of all ESG-related news jointly, each single ESG category, or a composed Environmental and Social measure.

[43] We define blockholders as large shareholders owning more than 10% of outstanding equity.

We winsorize all balance sheet items, realized compensation data (STI and LTI), target bonuses, and executive characteristics (age, tenure) at the 1%-level in each tail of the distribution.

4.4 Compensation Design in Practice

In this section, we describe how firms design executives' compensation plans in practice. We introduce notation and show through which contract parameters ESG performance can enter the calculation of variable incentive pay. General information on compensation design can be found in Murphy (1999) and Edmans et al. (2017), as well as Edmans et al. (2023) who provide a survey on the objectives, constraints, and determinants of CEO pay from both a director's and investor's perspective.

An executive's compensation contract is typically composed of base salary, annual variable, and multi-year variable compensation elements. In the following, we introduce calculations and notation for the example of annual variable compensation, denoted as short-term incentive (STI) pay, which constitutes firms' preferred contracting margin to tie remuneration to ESG criteria. Note that we adopt the same notation for long-term incentive (LTI) pay.

Executives' STI includes all variable compensation elements with a performance period of one year. Its calculation is generally formula-based.[44] At the beginning of the year, firms set a *target* bonus or *target* STI amount, which we denote as $TSTI_{i,t}$ for executive i in year t. The *realized* bonus or *realized* STI amount $RSTI_{i,t}$ at the end of the year depends on the extent to which the executive has achieved the target, i.e., on the achievement rate (also called fulfillment rate) $f_{i,t}$:[45]

$$RSTI_{i,t} = TSTI_{i,t} \times f_{i,t} \tag{4.1}$$

Usually, the calculation of $RSTI_{i,t}$ in Equation (4.1) is qualified by so-called bonus hurdles and bonus caps. Specifically, firms impose constraints on the achievement rate $f_{i,t}$. If actual performance falls below a lower threshold, realized $RSTI_{i,t}$ is set to a lower bound (typically zero). Similarly, $RSTI_{i,t}$ is sometimes capped at an upper threshold. In the so-called incentive zone between the hurdle and the cap, $RSTI_{i,t}$ increases (typically linearly) in target achieve-

[44] In some firms, directors reserve the right to make adjustments to the formula-based STI payout ex post. We discuss these ex-post adjustments further below.

[45] Payout of realized STI is typically made in cash at or shortly after the end of the performance period.

ment $f_{i,t}$ as in Equation (4.1).

In practice, most contracts include several performance metrics. It then becomes important how much each of these different metrics contributes to overall target achievement $f_{i,t}$. Most bonus plans (STI plans) in our sample are additive:

$$f_{i,t} = \underbrace{w_{i,t}^{B1} \times f_{i,t}^{B1} + w_{i,t}^{B2} \times f_{i,t}^{B2} + ... + w_{i,t}^{Bn} \times f_{i,t}^{Bn}}_{\text{binding metrics}} \quad + \quad \underbrace{w_{i,t}^{D} \times f_{i,t}^{D}}_{\text{discretionary metrics}} \tag{4.2}$$

In Equation (4.2), the firm has committed to individual weights $w_{i,t}^{B1}$, $w_{i,t}^{B2}$, ..., $w_{i,t}^{Bn}$ for n performance metrics at the beginning of the fiscal year. For these n metrics, the executive knows in advance the individual weights with which they will be included in the calculation of the overall target achievement rate $f_{i,t}$. For example, achievement rate $f_{i,t}^{B1}$ of metric $B1$ will enter with weight $w_{i,t}^{B1}$. As the firm commits to these weights ex ante, we call them the *binding metrics*. In practice, these binding metrics are often measured with respect to *hard* and verifiable (quantitative) key performance indicators (KPIs). In our analyses, it will be convenient to distinguish between binding *ESG* and binding *non-ESG* performance metrics.

For this purpose, we rewrite Equation (4.2):

$$f_{i,t} = \underbrace{w_{i,t}^{B,ESG} \times f_{i,t}^{B,ESG}}_{\text{binding ESG metrics}} \quad + \quad \underbrace{w_{i,t}^{B,nESG} \times f_{i,t}^{B,nESG}}_{\text{binding non-ESG metrics}} \quad + \quad \underbrace{w_{i,t}^{D} \times f_{i,t}^{D}}_{\text{discretionary metrics}} \tag{4.3}$$

where $w_{i,t}^{B,ESG}$ and $f_{i,t}^{B,ESG}$ are the total weight and the joint achievement rate of all binding *ESG* metrics together, and $w_{i,t}^{B,nESG}$ and $f_{i,t}^{B,nESG}$ are defined correspondingly for binding *non-ESG* metrics. In other words, Equation (4.3) shows the relative importance of (binding) ESG and non-ESG performance metrics in the calculation of total target achievement $f_{i,t}$.

Besides binding metrics, many firms also include *discretionary* (often called *soft*) metrics. We define these discretionary metrics for instances in which the firm does not use sufficiently granular weights that would allow us to identify the importance of the ESG criteria relative to non-ESG metrics.[46] An example would be an STI plan that considers the two metrics "Continue

[46]For our purpose, this distinction between separate weights for ESG and non-ESG criteria is enough to identify the relative importance of ESG metrics. More generally, the defining characteristic of discretionary metrics is that their individual weights are not known to the executive at the beginning of the fiscal year (Maas (2018)). Ikram et al. (2023) differentiate between objective or formulaic contracts and subjective contracts. They define contracts as objective if a proxy statement clearly specifies the weights (dollar amount of compensation or the percentage) attached to CSR-related metrics.

to drive strategic initiatives for market access (non-ESG)" and "Improve employee satisfaction (ESG)", but without specifying their relative importance (their individual weights). At the beginning of the year, the executive only knows the joint weight of both discretionary metrics *together*, i.e., $w_{i,t}^{D} = 1 - \sum_{j}^{n} w_{i,t}^{Bj}$. In Equations (4.2) and (4.3), the joint contribution of all discretionary metrics together to overall target achievement $f_{i,t}$ enters as the product between their joint weight $w_{i,t}^{D}$ and achievement rate $f_{i,t}^{D}$. Economically, these discretionary metrics expose the executive to additional risk as the (supervisory) board / the compensation committee will pick their relative weights only ex post. The executive cannot be certain how much strong performance on one of these discretionary metrics will actually be rewarded ex post.[47]

4.4.1 Example of an STI Plan (Schneider Electric, 2016)

Figure 4.1 shows an example of an STI plan for the CEO of Schneider Electric in 2016.[48] The plan lists various performance metrics that the company labels as economic and non-economic criteria, and a category called "individual assessment by the board". Additional information in the annual report allows us to identify the economic and non-economic criteria as non-ESG and ESG metrics, respectively. As the STI plan allows us to calculate separate weights of $w_{i,t}^{B,ESG} = 20\%$ and $w_{i,t}^{B,nESG} = 60\%$ for these ESG and non-ESG metrics, we call them *binding* metrics. The category "individual assessment by the board" comprises three additional performance metrics (described in the annual report), of which only one qualifies as ESG. As Schneider Electric does not commit to a specific weight for this ESG metric, we speak of one *discretionary ESG* and two *discretionary non-ESG* metrics with only one joint weight of $w_{i,t}^{D} = 20\%$.[49]

The STI plan in Figure 4.1 further shows that the company targeted to pay the CEO $RSTI_{i,t} = 130\% \times$ *Base Salary* (EUR 1,235,000) for a target achievement of, on (weighted) average, 100%. In practice, the CEO achieved an overall achievement rate of $f_{i,t} = 129.4\%$, resulting in actual payout of 168.22% of base salary or EUR 1,598,090. Based on the information in Figure 4.1, we also observe the break down of target achievement into $f_{i,t}^{D} = 100\%$, $f_{i,t}^{B,ESG} = 132\%$, and $f_{i,t}^{B,nESG} = 138\%$.

[47] In our example, the firm is free to concentrate on "Continue to drive strategic initiatives" at the expense of employee satisfaction (and vice-versa) when calculating their joint achievement rate $f_{i,t}^{D}$ at the end of the year.

[48] Appendix C.4 shows additional examples.

[49] If the three individual goals had all been classified as ESG (hypothetically speaking), then we would have an example in which $w_{i,t}^{D} = 0$, $w_{i,t}^{B,ESG} = 40\%$, and $w_{i,t}^{B,nESG} = 60\%$. Put differently, for our classification into binding and discretionary ESG metrics, we only require that ESG metrics are not pooled with non-ESG criteria. Within the ESG category, no individual weights are necessary for our purpose of identifying the importance of ESG relative to non-ESG criteria.

The case of Schneider Electric is an example of high reporting granularity. While the majority of STI plans in our sample fit the framework of Equations 4.1 to 4.3, not all firms report, for example, separate weights $w_{i,t}^{B1}$, ..., $w_{i,t}^{Bn}$ and achievement rates $f_{i,t}^{B1}$, ..., $f_{i,t}^{Bn}$ for each performance metric individually. However, we typically observe aggregate weights $w_{i,t}^{B,ESG}$, $w_{i,t}^{B,nESG}$, and $w_{i,t}^{D,ESG}$, allowing us to infer the joint importance of binding ESG metrics relative to non-ESG metrics.

Finally, a minority of STI plans do not connect metrics purely additively but multiplicatively. Typically, these are cases in which the firm calculates STI as described above, but the board retains the right to scale the resulting payout by an additional multiplier to adjust STI, for example, for financial constraints (e.g., during a pandemic) or non-anticipated dimensions of (individual) performance. In those cases, we record the presence of such a scaling multiplier, whether the board exercised its right to apply it, and whether it is linked to ESG or non-ESG performance. This allows us to exclude such STI plans in robustness analyses.

Figure 4.1: Schematic CEO STI Contract at Schneider Electric (2016)

		Binding Targets	Target relative to base salary: On- Target payment	Achievement rate-base 100	Actual payout: In % of base salary	Actual payout: Amount (in €)
60%		Economic Criteria		By-target achiev.		
40%		**Group Financial indicators**				
	13.33%	Organic Growth of the Group	17.33%	155%	26.90%	255,233
	13.33%	Adjusted EBITA	17.33%	200%	34.70%	329,333
	13.33%	Group cash conversion rate	17.33%	200%	34.70%	329,333
20%		**Company program priorities**				
	6.67%	Transcactional Sales growth	8.67%	135%	11.70%	111,150
	6.67%	Field Services (without Process Automation) Sales growth	8.67%	0%	0%	0
	6.67%	Systems Gross Margin (Projects & Equipments)	8.67%	0%	0%	0
20%		Non-economic Criteria				
	10%	Customer satisfaction	13%	64%	8.30%	79,040
	10%	Planet & Society Barometer	13%	200%	26%	247,000
20%		Individual assessment by the board				
	20%	Individual goals determined by the board based on the company's strategic plan	26%	100%	26%	247,000
100%	TOTAL	Discretionary Targets	130%	129.40%	168.22%	1,598,090

Notes: The figure shows the schematic STI contract of the CEO of Schneider Electric (2016), including binding non-ESG targets, binding ESG targets, discretionary targets, as well as target achievement rates and payout amounts. Extracted from the Financial and Sustainable Development Annual Report, Registration Document, Schneider Electric (2016), p. 165.

4.4.2 Summary Statistics

Table 4.2 reports summary statistics for STI plans, that is, target payout $TSTI_{i,t}$, the number of binding and discretionary metrics for ESG and non-ESG performance, as well as their (joint) weights $w_{i,t}^{B,ESG}$, $w_{i,t}^{B,nESG}$, and $w_{i,t}^{D,ESG}$. Further reported are variables describing the presence of hurdles, caps, and board discretion, for example, in the form of ex-post multipliers. Table 4.3 shows the same variables for LTI plans. Target achievement rates (only for STI plans) are reported alongside with realized pay amounts in euros in Table 4.4. Finally, we break down the different ESG metrics into their environmental, social, and governance components in Table IC.2 of the Internet Appendix C.5. As we provide a detailed analysis of these variables in Section 4.5, we refrain from discussing these summary statistics here.

4.5 Empirical Analysis

To what extent do firms link executive compensation to ESG performance metrics? Have these links the potential to enhance ESG incentives relative to compensation plans without ESG metrics?[50] To answer these questions, we analyze the number and weights of ESG metrics, pay-to-performance sensitivities, and the resulting pay risk borne by executives in Section 4.5.1. In Section 4.5.2, we analyze whether firms redesign compensation plans and revise the relative importance of ESG and non-ESG metrics following bad ESG news. Heterogeneity across industry sectors and executive positions is studied in Section 4.5.3.

4.5.1 Relevance of ESG Pay

In Equation (4.3), ESG performance metrics enter the calculation of executive pay through two parameters: their weight relative to other metrics and their target achievement rates. Therefore, ESG-linked pay can strengthen ESG incentives only if two necessary conditions are met: ESG metrics must receive material weight, and ESG target achievement must be sensitive to changes in ESG performance. If either condition is violated, executives will not work harder to improve ESG performance than if these metrics were not included in their compensation plans to begin with.

[50] Even compensation plans without ESG metrics could generate *some* ESG incentives. If ESG performance enhances shareholder value and pay is tied to stock prices, executives could still value ESG improvements. Moreover, even when ESG performance reduces financial performance, executives with intrinsic ESG preferences could seek to maintain a minimum level of ESG performance.

Table 4.2: Summary Statistics for Short-term Incentive (STI) Pay

		Obs (1)	Mean (2)	S.D. (3)	P5 (4)	P10 (5)	P25 (6)	P50 (7)	P75 (8)	P90 (9)	P95 (10)
Target STI amount:											
$TSTI_{i,t}$	(€)	1,915	988,677	577,729	366,600	420,555	623,583	813,934	1,240,250	1,670,918	2,200,000
$TSTI_{i,t}/Base\ salary_{i,t}$	(%)	1,915	100.614	39.665	50	60	80	100	105	150	173
Number of binding STI metrics:											
$\#_{i,t}^{B,ESG}$	(#)	2,500	0.426	1.597	0	0	0	0	0	1	3
$\#_{i,t}^{B,nESG}$	(#)	2,490	4.217	4.051	0	0	1	3	5	9	12
$\#_{i,t}^{B,ESG} + \#_{i,t}^{B,nESG}$	(#)	2,490	4.645	4.408	0	0	2	4	6	10	13
Weight of binding STI metrics:											
$w_{i,t}^{B,ESG}$	(%)	2,490	2.714	7.315	0	0	0	0	0	15	20
$w_{i,t}^{B,nESG}$	(%)	2,490	72.067	34.869	0	0	60	80	100	100	100
$w_{i,t}^{B,ESG} + w_{i,t}^{B,nESG}$	(%)	2,490	74.781	35.831	0	0	65	100	100	100	100
Number of discretionary STI metrics:											
$\#_{i,t}^{D,ESG}$	(#)	2,500	1.433	2.408	0	0	0	0	2	5	5
$\#_{i,t}^{D,nESG}$	(#)	2,490	3.513	5.866	0	0	0	0	5	12	17
$\#_{i,t}^{D,ESG} + \#_{i,t}^{D,nESG}$	(#)	2,490	4.951	7.564	0	0	0	0	8	16	21
Weight of discretionary STI metrics:											
$w_{i,t}^{D}$	(%)	2,490	25.214	35.830	0	0	0	0	35	100	100
STI hurdle $Y/N_{i,t}$	(1/0)	2,560	0.652	0.476	0	0	0	1	1	1	1
STI cap $Y/N_{i,t}$	(1/0)	2,560	0.960	0.197	1	1	1	1	1	1	1
$STI\ cap_{i,t}/Base\ salary_{i,t}$	(%)	2,135	181.173	66.948	90	113	140	180	200	250	281
STI board discretion allowed $Y/N_{i,t}$	(1/0)	2,557	0.759	0.428	0	0	1	1	1	1	1
STI ex post multiplier (ESG) $Y/N_{i,t}$	(1/0)	2,560	0.041	0.199	0	0	0	0	0	0	0
STI ex post multiplier (nESG) $Y/N_{i,t}$	(1/0)	2,560	0.143	0.350	0	0	0	0	0	1	1

Notes: Reported are summary statistics for the main remuneration elements in executives' short-term incentive contracts (STI) at the executive-year level. The superscripts B and D designate *binding* and *discretionary* performance metrics, respectively. The superscripts ESG and $nESG$ designate *ESG* and *non-ESG* metrics. For more information about the design of binding and discretionary performance metrics, see Section 4.4. For variable definitions, see Appendix C.1.

Table 4.3: Summary Statistics for Long-term Incentive (LTI) Pay

		Obs (1)	Mean (2)	S.D. (3)	P5 (4)	P10 (5)	P25 (6)	P50 (7)	P75 (8)	P90 (9)	P95 (10)
Number of binding LTI metrics:											
$\#_{i,t}^{B,ESG}$	(#)	2,390	0.210	0.730	0	0	0	0	0	0	2
$\#_{i,t}^{B,nESG}$	(#)	2,334	2.589	1.912	0	1	2	2	3	5	5
$\#_{i,t}^{B,ESG} + \#_{i,t}^{B,nESG}$	(#)	2,334	2.804	2.148	0	1	2	2	3	5	6
Weights of binding LTI metrics:											
$w_{i,t}^{B,ESG}$	(%)	2,334	2.591	9.304	0	0	0	0	0	5	20
$w_{i,t}^{B,nESG}$	(%)	2,334	88.668	27.642	0	50	100	100	100	100	100
$w_{i,t}^{B,ESG} + w_{i,t}^{B,nESG}$	(%)	2,334	91.259	26.874	0	67	100	100	100	100	100
Number of discretionary LTI metrics:											
$\#_{i,t}^{D,ESG}$	(#)	2,390	0.197	0.686	0	0	0	0	0	1	1
$\#_{i,t}^{D,nESG}$	(#)	2,334	0.633	2.084	0	0	0	0	0	3	5
$\#_{i,t}^{D,ESG} + \#_{i,t}^{D,nESG}$	(#)	2,334	0.834	2.655	0	0	0	0	0	4	7
Weight of discretionary LTI metrics:											
$w_{i,t}^{D}$	(%)	2,334	8.741	26.873	0	0	0	0	0	33	100
LTI hurdle $Y/N_{i,t}$	(1/0)	2,560	0.787	0.409	0	0	1	1	1	1	1
LTI cap $Y/N_{i,t}$	(1/0)	2,560	0.429	0.495	0	0	0	0	1	1	1
LTI $cap_{i,t}/Base\ salary_{i,t}$	(%)	1,097	292.521	119.171	133	140	200	275	380	438	500
LTI board discretion allowed $Y/N_{i,t}$	(1/0)	2,258	0.753	0.431	0	0	1	1	1	1	1
LTI ex post multiplier (ESG) $Y/N_{i,t}$	(1/0)	2,560	0.000	0.000	0	0	0	0	0	0	0
LTI ex post multiplier (nESG) $Y/N_{i,t}$	(1/0)	2,560	0.016	0.127	0	0	0	0	0	0	0

Notes: Reported are summary statistics for the main remuneration elements in executives' long-term incentive contracts (LTI) at the executive-year level. Information on the numbers and weights of metrics in LTI plans always refer to the first year of granting. The superscripts B and D designate *binding* and *discretionary* performance metrics, respectively. The superscripts ESG and $nESG$ designate *ESG* and *non-ESG* metrics. For more information about the design of binding and discretionary performance metrics, see Section 4.4. For variable definitions, see Appendix C.1.

Table 4.4: STI Target Achievement and Realized Pay

		Obs (1)	Mean (2)	S.D. (3)	P5 (4)	P10 (5)	P25 (6)	P50 (7)	P75 (8)	P90 (9)	P95 (10)
Panel A: Target achievement rates in STI											
Overall target achievement rate:											
$f_{i,t}$	(%)	1,859	104	40	36	59	87	102	123	145	166
Achievement rate of binding targets:											
$f_{i,t}^{B,ESG}$	(%)	227	103	32	44	69	92	100	121	148	150
$f_{i,t}^{B,nESG}$	(%)	1,324	100	40	34	51	82	100	119	139	159
$f_{i,t}^{B}$	(%)	1,584	101	39	37	54	82	100	120	143	159
Achievement rate of discretionary targets:											
$f_{i,t}^{D}$	(%)	709	108	41	27	63	90	104	124	162	180
Panel B: Realized pay (base, STI, LTI)											
$RSTI_{i,t}/Base\ salary_{i,t}$	(%)	2,388	84	54	0	19	47	79	115	149	170
$RSTI_{i,t}$	(€)	2,388	821,716	612,654	0	174,867	424,817	707,572	1,053,200	1,608,147	2,027,285
$Base\ salary_{i,t}$	(€)	2,464	1,093,465	576,840	513,338	600,000	720,000	882,466	1,300,000	1,923,100	2,339,052
$RSTI_{i,t} + Base\ salary_{i,t}$	(€)	2,560	1,819,335	985,311	511,539	841,699	1,207,902	1,622,384	2,349,000	3,053,200	3,621,199
$RLTI_{i,t}$	(€)	2,489	1,736,526	1,334,585	35,692	317,600	870,526	1,429,362	2,241,950	3,627,090	4,450,626
$RSTI_{i,t} + Base\ salary_{i,t} + RLTI_{i,t}$	(€)	2,556	3,530,735	2,024,399	961,180	1,503,386	2,183,970	3,013,434	4,582,079	6,397,592	7,576,694
$Board\ exercised\ discretion\ Y/N_{i,t}$	(1/0)	2,500	0.136	0.342	0	0	0	0	0	1	1
$STI\ partly\ deferred\ Y/N_{i,t}$	(1/0)	2,560	0.534	0.499	0	0	0	1	1	1	1
$STI\ deferral_{i,t}/RSTI_{i,t}$	(%)	945	53	17	33	33	50	50	60	80	80
$STI\ deferral\ period_{i,t}$	(years)	1,032	3.25	0.82	2.00	2.00	3.00	3.00	4.00	5.00	5.00

Notes: Panel A shows summary statistics for target achievement rates (target fulfillment rates) in short-term incentive pay (STI). Panel B shows summary statistics for realized compensation (base salary, STI, and LTI), board discretion, and deferred pay. For more information about the design of STI and LTI, see Section 4.4. For variable definitions, see Appendix C.1.

4.5.1.1 Relative Numbers and Weights of ESG and Non-ESG Metrics

The majority of existing studies has not studied the relative weights of ESG and non-ESG metrics, but focuses solely on whether a compensation plan is tied to at least one ESG criterion (e.g., Carter et al. (2023); Cohen et al. (2023); Hazarika et al. (2025); Ikram et al. (2023)). In a first step, we replicate this "extensive-margin" approach before leveraging our more granular data on weights. Figure 4.2 considers 674 executives from 73 firms listed on the EURO STOXX 50 or the STOXX Europe 50. Panel A shows that in 2013, only 40% of these executives have STI pay linked to at least one ESG metric (dashed black line). This share increases to around 60% by 2020. Panel A further distinguishes between the three ESG dimensions. In 2020, 56% of executives were subject to at least one social metric, just over 30% to one or more governance metrics, and only 20% to environmental criteria. In contrast, all executives are evaluated on at least one non-ESG performance metrics throughout the entire sample period (solid line at 100%).

We further break down each of the three main ESG categories into their subcategories (see Table C.2 in Appendix C.3 for definitions and examples of these subcategories). Panel B of Figure 4.2 shows that emissions reduction metrics are the most common environmental criteria, increasing from 3% to 15% of executives by 2020, with the most significant rise from 2018 to 2020, i.e., precisely at a time when the price of EU carbon permits quadruples (see Figure IC.1 of the Internet Appendix C.5). Metrics for resource use, such as water and energy efficiency, increase from 0% in 2013 to an executive share of 8.4% in 2020. Innovation and green R&D metrics remain uncommon and are found for only 3.1% of executives by 2020.

For social metrics, Panel C reveals that workforce-related and product quality metrics are the most common, reaching executive shares of 46% and 31%, respectively, by 2020. Social criteria that consider possible externalities on local communities are much rarer, with a share of about 10%, and not a single executive has any metrics explicitly related to human rights. Panel D shows that governance metrics related to CSR strategy and management are the most prevalent, covering 21% of executives in 2020, where the strong increase over our sample period is likely driven by an anticipation of new regulations towards improved corporate governance, such as board independence and board diversity.

Overall, Figure 4.2 suggests that while a substantial share of executives have at least one ESG performance metric, firms tend to prefer criteria that may improve financial performance, such as emissions and workforce-related criteria, over metrics addressing broader externalities, such as human rights.

While the extensive margin approach in Figure 4.2 tells us about the *presence* of ESG in STI, it remains silent about its *importance* relative to non-ESG criteria. In a first attempt to overcome this limitation, Figure 4.3 plots the *number* of ESG versus non-ESG metrics per executive. Panel A shows that, on average, executives have four times more non-ESG than ESG metrics, and Panel B reveals that the average executive has three times more social metrics than governance-linked metrics and very few environmental criteria.[51] Together, Panels A and B suggest that ESG metrics, particularly social outcomes, carry only about one-fourth the importance of non-ESG criteria. Although this is a step forward compared to Figure 4.2, there remain two limitations to this analysis.

First, it remains unclear if ESG metrics are binding and how heavily they factor into year-end performance assessments. As explained in Section 4.4, *binding* metrics have a pre-set weight known to the executive, whereas *discretionary* metrics lack an explicit ex-ante weight and may even be disregarded by the firm ex post. Interestingly, Panel C of Figure 4.3 shows that especially the number of discretionary ESG metrics increases during our sample period, whereas binding ESG metrics only increase slightly after 2018.

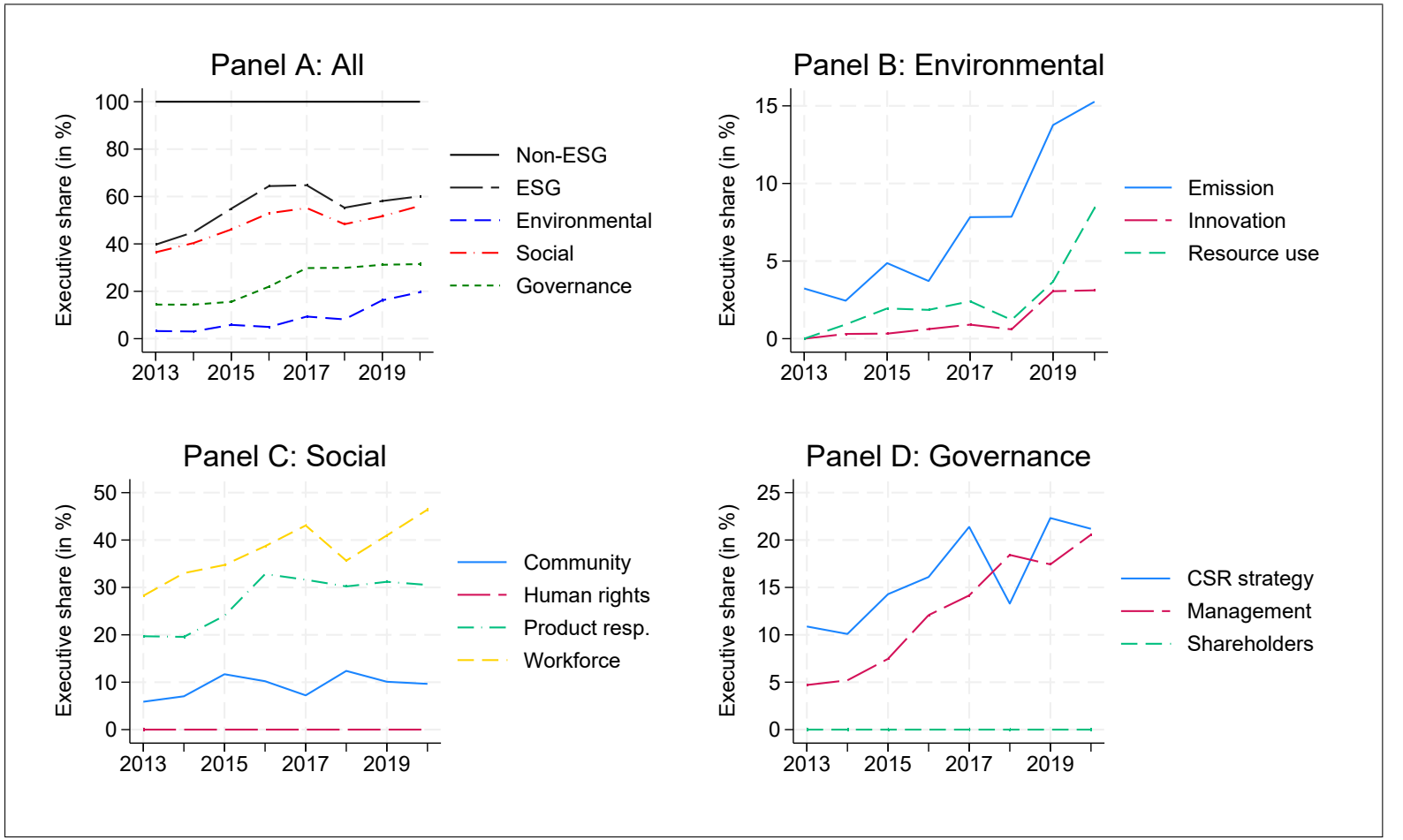
A second limitation is that firms can inflate the perceived importance of ESG-linked pay by splitting a given ESG outcome into several criteria, even if they lack substantial weight or are non-binding.[52] To overcome these limitations, Figure 4.4 plots the joint weight that firms give to binding ESG metrics together. Initially, the total weight of all binding ESG metrics combined is less than 2% for the average executive. Even after a strong increase, this ESG weight reaches only 5%. Averaged across the entire sample period, binding ESG metrics enter STI with an explicit weight of just 3%, compared to 72% for binding non-ESG metrics (see Table 4.2). In other words, firms commit a roughly 24 times larger weight to non-ESG than to ESG performance metrics.

Importantly, the immaterial weight of ESG metrics is not confined to STI pay only. Table 4.3 shows that binding ESG metrics receive an explicit weight of only 3% in LTI plans as well, compared to 89% for conventional metrics. Overall, firms' commitment to incorporating ESG performance in either STI or LTI payout calculations is close to zero.

[51] We plot average numbers of ESG metrics per executive because plotting median numbers would not reveal any new information. Figure 4.2 already shows that the median executive has zero metrics in most ESG categories.

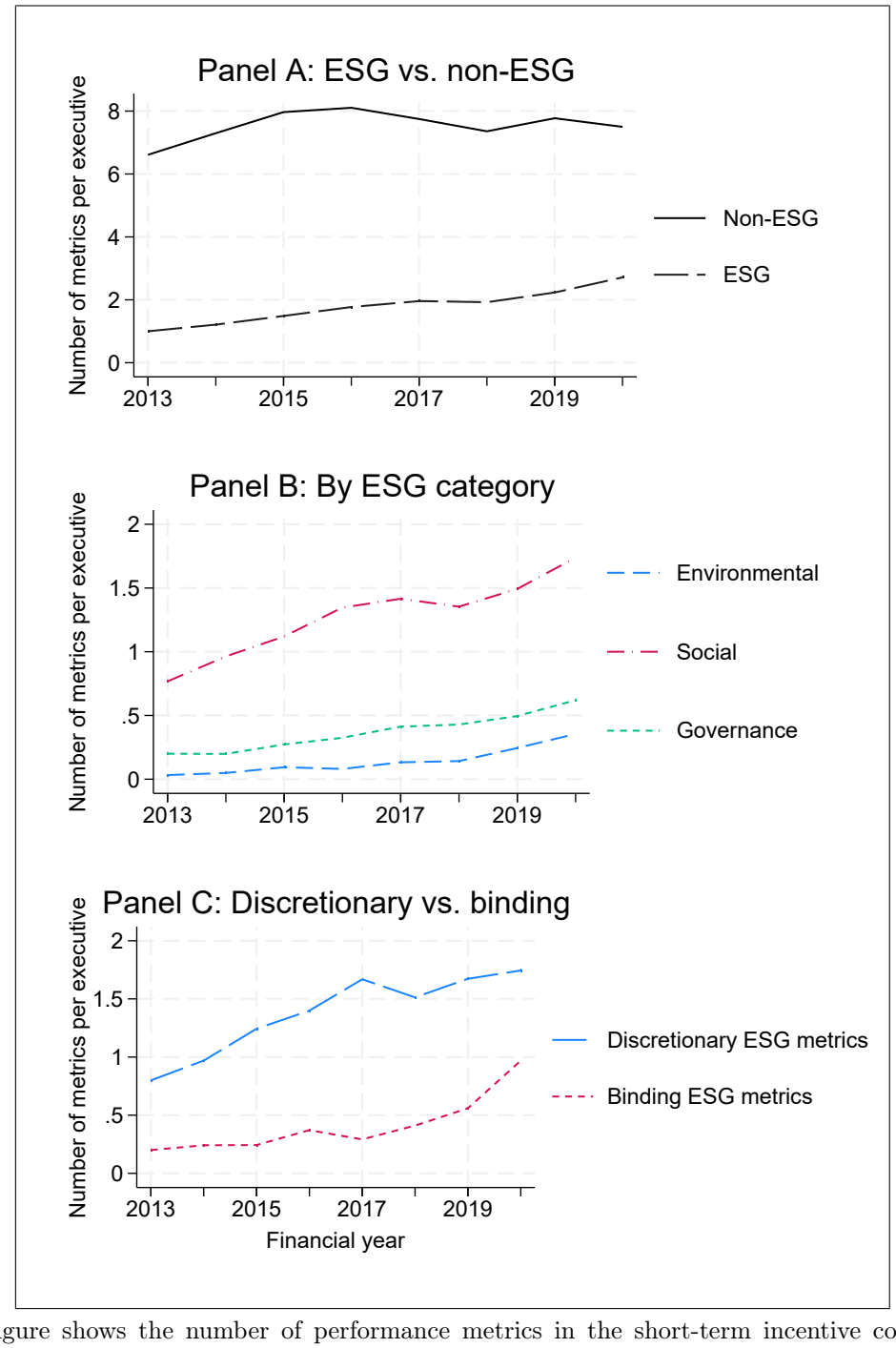
[52] Whereas this is a conceptually valid concern, we find no direct evidence for an increase in reporting granularity. We examine the variable *ESG Reporting Scope*, which captures the percentage of activities covered in Environmental and Social reporting. It remains stable over our sample period, ranging between 86% and 91% with no clear time trend. Furthermore, we find no significant relationship between the number of ESG goals and reporting granularity. Results are reported in Table IC.3 of the Internet Appendix C.5.

Figure 4.2: Share of Executives with at least one ESG or Non-ESG-related Performance Metric (STI)



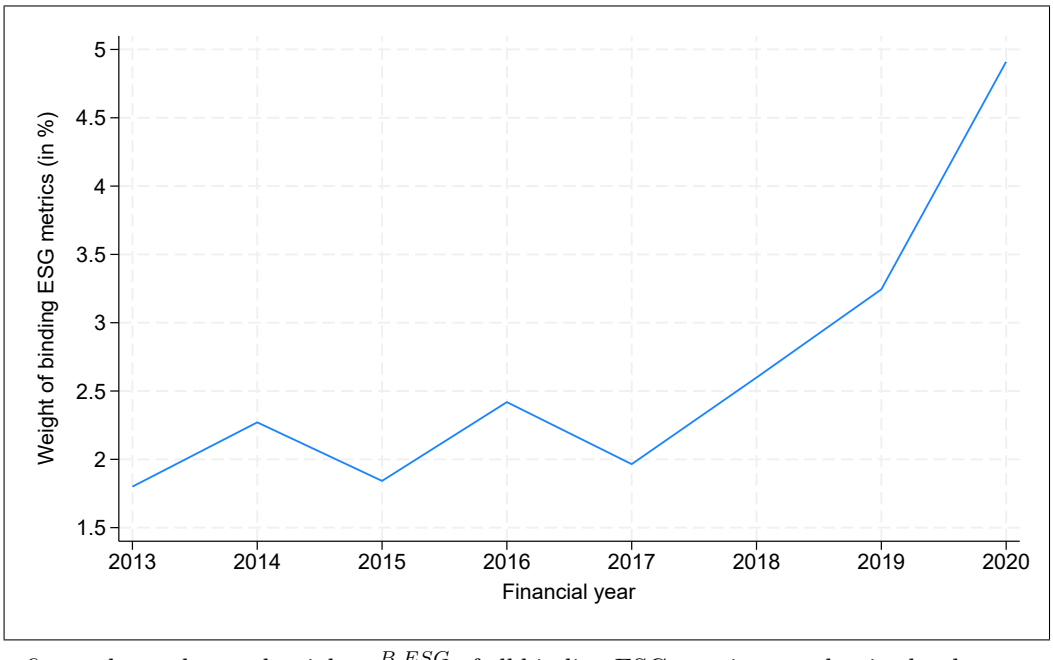
Notes: The figure shows the share of executives with at least one ESG- or non-ESG performance metric in their short-term incentive contracts (STI). Panel A differentiates between ESG and non-ESG metrics, while also splitting the ESG part into its three main categories Environmental, Social, and Governance. Panels B to D show the corresponding executive shares for each of the three subcategories in each main ESG category, i.e., Environmental (Panel B), Social (Panel C), and Governance (Panel D). Analyses are based on a sample of 674 executives from 73 firms that have been constituents of the EURO STOXX 50 and the STOXX Europe 50 for at least ten days between 2013 and 2020.

Figure 4.3: Number of ESG vs. Non-ESG Performance Metrics per Executive (STI)



Notes: The figure shows the number of performance metrics in the short-term incentive contract (STI) of the average executive in our sample. Panel A differentiates between ESG and non-ESG metrics and Panel B splits ESG up into its three main categories Environmental, Social, and Governance. Panel C differentiates between discretionary and binding ESG metrics. We define binding performance metrics as metrics where the firm commits already at the beginning of the fiscal year with what weight it will consider each of them in the calculation of realized STI at year-end. For discretionary metrics, however, the respective weights are not known to the executive ex ante. For more information about the design of discretionary and binding performance metrics, see Section 4.4.

Figure 4.4: Weight of Binding ESG Performance Metrics in STI



Notes: The figure shows the total weight $w_{i,t}^{B,ESG}$ of all binding ESG metrics together in the short-term incentive contract (STI) of the average executive in our sample. For more information about the design of binding performance metrics and their weights, see Section 4.4.

4.5.1.2 Pay-to-Performance Sensitivities

Do the immaterial weights of *binding* ESG metrics in STI and LTI calculations imply that ESG-linked pay provides no stronger ESG incentives than conventional compensation plans? Not necessarily. In STI plans, ESG criteria also enter compensation through *discretionary* metrics (Figure 4.3). Although these discretionary components pool ESG and non-ESG criteria without specifying their relative importance, they carry a substantial combined weight of, on average, 25% (see $w_{i,t}^{D}$ in Table 4.2). Interestingly, this weight of discretionary pay in STI is also three times larger than in LTI plans (Table 4.3).

If discretionary STI metrics are sufficiently sensitive to ESG performance, they could still generate meaningful ESG incentives. Hence, to assess the incentive relevance of ESG-linked pay, we next examine the performance sensitivities of different STI components. Equation (4.4) shows the regression specification we use to estimate performance sensitivities:

$$f_{i,t} = \beta \; Performance \; metric_{i,t} + Executive \; FE + Industry \times Year \; FE + \epsilon_{i,t} \qquad (4.4)$$

where $f_{i,t}$ is the overall achievement rate of executive i for performance in year t. $f_{i,t}$ is measured in percentage points and defined as the weighted average of target-level achievement

rates (Equation (4.2)). A strength of our data is that we also observe achievement rates at a more granular level, at least for a subset of executives. Hence, we also estimate Equation (4.4) for the target achievement rates of binding ESG ($f_{i,t}^{B,ESG}$) and non-ESG metrics ($f_{i,t}^{B,nESG}$), as well as for discretionary pay ($f_{i,t}^{D}$), which pools ESG and non-ESG metrics. Hence, we can check explicitly which "type" of metrics is more sensitive to a certain type of performance changes.

The main regressor in Equation (4.4) is *Performance metric*$_{i,t}$. The executive fixed effects ensure that we estimate its coefficient β only based on *within-executive time variation*. Hence, β identifies the *sensitivity* of executive i's achievement rate to changes in performance. Its estimate tells us by how many percentage points the achievement rate changes when the performance metric increases by one unit. Finally, we also include interacted industry $\times$ year fixed effects to control for industry-wide trends in performance, as firms could assess executives relative to peer companies.

Before we study ESG performance metrics, we first study sensitivities to stock returns, as a measure of overall financial performance. The reason why we consider the stock return is *not* because STI plans would necessarily include the stock return itself as a performance metric. But a positive sensitivity to the stock return would imply that STI pay is indeed sensitive to performance metrics that capture shareholder value creation. Table 4.5, Panel A shows the corresponding estimates. Unsurprisingly, executives' overall achievement rate $f_{i,t}$ is positively sensitive to stock returns (column 1). For a one-standard deviation increase in the annual stock return (22.41 percentage points), the overall achievement rate and, therefore, realized pay $RSTI$ increase by 12% (see Equation (4.1)).

In columns 2 to 4, we report the estimated sensitivities at a more granular level. We have fewer observations than in column 1 because not all STI plans include each "class" of metrics, and ESG metrics, and especially *binding* ESG metrics, are less common than non-ESG metrics (see Section 4.5.1.1 and Figure 4.3). Interestingly, the sensitivity for binding ESG metrics is statistically significant despite small sample size (Panel A, column 2). Importantly, this sensitivity of $f_{i,t}^{B,ESG}$ to the stock price is negative and rather large in absolute size. Possibly, for the average executive, there exists a trade-off between shareholder value creation and ESG performance, as measured by binding ESG metrics. By contrast, the achievement rates for binding non-ESG metrics and for discretionary metrics are all positively sensitive to stock returns.

Next, we consider sensitivities to changes in ESG performance. Ideally, for each executive, we would estimate Equation (4.4) for the specific performance metrics included in his or her STI plan. This is unfeasible given the variety of metrics across the 674 executives of our 73 sample firms. Moreover, for many metrics, only the firms, but not the econometrician, observe

performance changes. Instead, we focus on a few verifiable performance metrics that are common and likely to be positively correlated with many similar metrics.

Panel B shows sensitivities to total CO_2 emissions equivalents (normalized by sales). We choose this metric as a proxy for environmental performance because emissions-related metrics are by far the most common in the E category (Figure 4.2, Panel B). Column 2 shows that target achievement for binding ESG metrics has a large negative sensitivity to emissions. An increase by one standard deviation reduces $f_{i,t}^{B,ESG}$ by 113%.[53] One should not overinterpret this exact value, but binding ESG metrics seem to be an important mechanism to link STI pay to emissions. By contrast, target achievement of discretionary metrics (column 4) shows no significant sensitivity. Interestingly, the sensitivities of overall achievement rate $f_{i,t}$ in column 1 and of $f_{i,t}^{B,nESG}$ in column 3 are positive and significant. Possibly, high emissions tend to occur when non-ESG performance is high, which could again point to a trade-off.

Finally, Panel C shows sensitivities to employee turnover, which we consider as a proxy for workforce-related performance. Indeed, Figure 4.2, Panel C shows that workforce-related metrics are the most common in the S category. Moreover, employee turnover is verifiable and, conditionally on firm or executive fixed effects, a potentially good measure for changes in employee satisfaction. Again, we only find a significant and negative sensitivity for binding ESG metrics.

Overall, Table 4.5 suggests that *binding* metrics are a suitable way to make STI pay sensitive to verifiable environmental and social performance. Based on our findings, we cannot say the same for *discretionary* ESG and non-ESG metrics, which firms evaluate jointly without specifying their relative importance (see Section 4.4). One possible explanation relates to the use of *verifiable* performance in Table 4.5.

Indeed, when performance is verifiable, it makes sense for firms to use binding metrics and to commit to a weight, as this reduces uncertainty for the executive and should have stronger incentive effects. However, for performance that is hard to measure and verify anyway, such a binding commitment might be less important (Michaely et al. (2024)). Possibly, discretionary metrics are mostly linked to such hard-to-measure dimensions of performance.

[53] The standard deviation of emissions equals 0.422 in column 2. In other words, an executive with on-target achievement for binding ESG metrics ($f_{i,t}^{B,ESG} = 100\%$) sees target achievement drop essentially to zero ($f_{i,t}^{B,ESG} = 0$) when emissions increase by one standard deviation.

Table 4.5: Sensitivities of Target Achievement Rates to Verifiable Performance Metrics

	(1)	(2)	(3)	(4)
Dependent variable:	*Target achievement rate*			
	overall	binding metrics		discretionary
	$f_{i,t}$	$f_{i,t}^{B,ESG}$	$f_{i,t}^{B,nESG}$	$f_{i,t}^{D}$
Panel A: Stock return				
Stock return	0.526***	−0.727**	0.431***	0.399**
	(0.105)	(0.320)	(0.116)	(0.158)
Observations	1,710	179	1,185	584
Adjusted R^2	0.49	0.43	0.50	0.74
Panel B: Total CO_2 emission equivalents (normalized by sales)				
Total CO_2 emission equivalents	7.310***	−267.819**	5.149*	21.829
	(1.965)	(104.099)	(2.949)	(23.888)
Observations	1,684	179	1,159	584
Adjusted R^2	0.45	0.48	0.47	0.72
Panel C: Employee turnover				
Employee turnover	1.043	−3.828**	0.059	−0.494
	(1.131)	(1.668)	(1.364)	(0.914)
Observations	1,463	136	992	514
Adjusted R^2	0.41	0.59	0.42	0.74
Executive F.E.	Yes	Yes	Yes	Yes
Industry × Year F.E.	Yes	Yes	Yes	Yes

Notes: This table reports panel regressions at the executive-year level. The dependent variables are the overall target achievement rate $f_{i,t}$ in column 1, the achievement rate for binding ESG metrics $f_{i,t}^{B,ESG}$ in column 2, for binding non-ESG metrics $f_{i,t}^{B,nESG}$ in column 3, and for discretionary metrics $f_{i,t}^{D}$ (without further distinguishing between ESG and non-ESG) in column 4. The main regressors are *Stock return* in Panel A, *Total CO_2 emission equivalents normalized by sales* in Panel B, and *Employee turnover* in Panel C. Executive fixed effects and industry-by-year fixed effects are included throughout. Robust standard errors are clustered by firm and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 4.6: Sensitivities of Target Achievement Rates to ESG Scandals

	(1)	(2)	(3)	(4)
Dependent variable:	*Target achievement rate*			
	overall	binding metrics		discretionary
	$f_{i,t}$	$f_{i,t}^{B,ESG}$	$f_{i,t}^{B,nESG}$	$f_{i,t}^{D}$
Panel A: All ESG news				
High-reach bad ESG news	−8.023***	−7.474**	-6.091	-3.937
	(2.865)	(3.516)	(3.785)	(5.060)
Observations	1,710	179	1,185	584
Adjusted R^2	0.46	0.38	0.48	0.73
Panel B: Environmental news				
High-reach bad E news	−16.671**	−15.389**	−10.297	−2.957
	(7.325)	(6.424)	(6.672)	(4.256)
Observations	1,710	179	1,185	584
Adjusted R^2	0.46	0.37	0.48	0.72
Panel C: Social news				
High-reach bad S news	−8.156	6.577	−0.156	0.186
	(5.503)	(5.061)	(5.509)	(5.367)
Observations	1,710	179	1,185	584
Adjusted R^2	0.45	0.37	0.47	0.72
Panel D: Governance news				
High-reach bad G news	−7.600**	−22.251**	−4.241	−8.243**
	(3.409)	(9.888)	(3.257)	(4.046)
Observations	1,710	179	1,185	584
Adjusted R^2	0.46	0.48	0.48	0.73
Executive F.E.	Yes	Yes	Yes	Yes
Industry × Year F.E.	Yes	Yes	Yes	Yes

Notes: This table reports panel regressions at the executive-year level. The dependent variables are the overall target achievement rate $f_{i,t}$ in column 1, the achievement rate for binding ESG metrics $f_{i,t}^{B,ESG}$ in column 2, for binding non-ESG metrics $f_{i,t}^{B,nESG}$ in column 3, and for discretionary metrics $f_{i,t}^{D}$ (without further distinguishing between ESG and non-ESG) in column 4. The main regressor is the logarithm of one plus the number of ESG incidents with negative press coverage. We only consider incidents that 1) explicitly mention the name of the firm and precisely specify the underlying ESG incident and 2) reach a large audience by being published in global media outlets (called "high-reach" and "sharp" incidents in RepRisk). We consider news with respect to all three ESG categories in Panel A, but focus only on environmental news in Panel B, social news in Panel C, and on governance news in Panel D. See C.1 for details. Executive fixed effects and industry-by-year fixed effects are included throughout. Robust standard errors are clustered by firm and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Estimating sensitivities to non-verifiable performance is, by definition, difficult, especially for small performance changes. Whether an executive deserves a target achievement rate of 1 or of 2% might be debatable for a soft target like "progress on diversity, equity, and inclusion". However, for very large performance changes, the corresponding change in target achievement should be more visible. For example, a firm that loses a major discrimination case in court might find it difficult to justify why it reduces target achievement for "progress on diversity, equity, and inclusion" by only 1 or 2%. To capture these large performance changes, our next analysis focuses on events that are sufficiently important to generate significant press coverage and media attention.

Table 4.6 reports our analysis of high-reach press coverage of bad ESG performance. Only bad news that explicitly identify both the firm and the underlying ESG issue and reach a large audience through coverage in global media outlets are considered. Column 1 shows that such scandals reduce executives' overall achievement rate $f_{i,t}$ when we consider all ESG categories jointly (Panel A) and when we focus on environmental and governance news (Panels B and D). Social news have a similar point estimate in column 1 as environmental and governance news, but it is statistically insignificant.

More importantly, discretionary metrics only exhibit a significant sensitivity to bad governance news (-8.243), which also falls short of the corresponding sensitivity of binding ESG metrics (-22.251). For total ESG news and for environmental news, only binding ESG metrics show a significant (and negative) sensitivity.

Overall, we find little evidence that discretionary metrics are sensitive to changes in ESG performance, not even to high-reach press coverage of bad ESG events. This result is consistent with firms emphasizing the non-ESG targets at the expense of the ESG targets among the discretionary metrics. By contrast, binding ESG metrics are highly sensitive to ESG performance changes.

4.5.1.3 Contribution of ESG metrics to Pay Risk

We stated above that ESG-linked pay is ineffective unless ESG metrics are given material weight in STI calculations and ESG target achievement is responsive to changes in ESG performance. The evidence so far suggests that these necessary conditions do not hold. The *binding* ESG metrics are sensitive to performance, but rare and enter pay calculations with small weight in the average STI plan. The *discretionary* metrics have larger weight (for ESG and non-ESG combined), but seem insensitive to ESG performance. Given these results, we predict that ESG

contracting does not expose executives to additional remuneration risk.

In this section, we estimate the contribution of ESG performance metrics to the pay risk posed by the STI contract of executives. To this end, we decompose total STI variance into its different components. First, we substitute Equation (4.3) into Equation (4.1) to rewrite total realized STI:

$$\begin{aligned} RSTI_{i,t} &= TSTI_{i,t} \times w_{i,t}^{B,ESG} \times f_{i,t}^{B,ESG} + TSTI_{i,t} \times w_{i,t}^{B,nESG} \times f_{i,t}^{B,nESG} + TSTI_{i,t} \times w_{i,t}^{D} \times f_{i,t}^{D} \\ &= RSTI_{i,t}^{B,ESG} + RSTI_{i,t}^{B,nESG} + RSTI_{i,t}^{D}, \end{aligned} \tag{4.5}$$

where $RSTI_{i,t}^{B,ESG}$ and $RSTI_{i,t}^{B,nESG}$ are the parts of total realized STI that the executive receives for the achievement of binding ESG and non-ESG metrics. As explained in Section 4.4, discretionary ESG and non-ESG metrics lack individual weights and are only assessed jointly by definition. Hence, it is not possible to split $RSTI_{i,t}^{D}$ into an ESG and a non-ESG part.

In a second step, we eliminate cross-sectional variation in STI pay between executives. We only want to consider within-executive variation in pay, that is, by how much the STI of a given executive varies over the years she works at a given firm. To eliminate all between-executive variation in STI, we regress $RSTI_{i,t}^{B,ESG}$, $RSTI_{i,t}^{B,nESG}$, and $RSTI_{i,t}^{D}$ in Equation (4.5) on interacted executive $\times$ firm fixed effects $\theta_{i,f}$. In a third and final step, we calculate the variance decomposition for the regression residuals, which are free of between-executive STI variation, as follows:

$$\begin{aligned} Var\left(RSTI_{i,t}\right) \; &= Var\left(RSTI_{i,t}^{B,ESG}\right) + Var\left(RSTI_{i,t}^{B,nESG}\right) + Var\left(RSTI_{i,t}^{D}\right) \\ &\quad + 2Cov\left(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{B,nESG}\right) + 2Cov\left(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{D}\right) \\ &\quad + 2Cov\left(RSTI_{i,t}^{B,nESG}, RSTI_{i,t}^{D}\right) \end{aligned} \tag{4.6}$$

This variance decomposition is shown in Table 4.7. The variance shares are calculated by dividing the terms in Equation (4.6) with the total (within-executive) variance $Var\left(RSTI_{i,t}\right)$, which itself is reported in column 2.[54]

[54]To remain parsimonious, we do not introduce additional notation for the regression residuals free of between-executive variation. But all terms in Equation (4.6) and in Table 4.7 are based on these residuals.

Table 4.7: Decomposition of Within-executive STI Variance

			Binding metrics				Discretionary metrics		Covariance terms (2×cov)		
			$RSTI_{i,t}^{B,ESG}$		$RSTI_{i,t}^{B,nESG}$		$RSTI_{i,t}^{D}$		$(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{B,nESG})$	$(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{D})$	$(RSTI_{i,t}^{B,nESG}, RSTI_{i,t}^{D})$
	obs	$Var(RSTI_{i,t})$ $\times 10^{-6}$	STI weight	Var share	STI weight	Var share	STI weight	Var share	Var share	Var share	Var share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
All observations	1,251	106,586	0.027	0.015	0.694	1.284	0.279	0.733	0.015	−0.026	−1.021
No discretion	1,091	96,681	0.023	0.016	0.702	1.329	0.275	0.738	0.011	−0.031	−1.063
Min. 1 ESG metric	663	86,054	0.050	0.034	0.423	1.154	0.527	1.078	0.041	−0.065	−1.242

Notes: This table shows the decomposition of the variance of executives' realized short-term incentive compensation $RSTI_{i,t}$ into its different components. The analysis focuses on within-executive variation in pay. To this end, we first compute the residuals of STI parts $RSTI_{i,t}^{B,ESG}$, $RSTI_{i,t}^{B,nESG}$, and $RSTI_{i,t}^{D}$ that are not explained by executive fixed effects. We sum up these residuals, free of cross-sectional variation, to total $RSTI_{i,t}$. Column 2 reports the variance $Var\left(RSTI_{i,t}\right)$ of the remaining within-executive variation in total STI. The variance shares reported in columns 4, 6, 8, and 9 to 11 correspond to the variance and covariance terms in Equation (4.6) divided by total STI variance. For example, the variance share of binding ESG metrics in column 4 is defined as $Var(RSTI_{i,t}^{B,ESG})/Var(RSTI_{i,t})$. Columns 3, 5, and 7 report the ex ante weights that firms assign to the different STI performance metrics, and are calculated as the sample averages of $w_{i,t}^{B,ESG}$, $w_{i,t}^{B,nESG}$, and $w_{i,t}^{D}$. As in previous tables, superscripts B and D designate binding and discretionary performance metrics, and superscripts ESG and $nESG$ designate ESG and non-ESG metrics. We estimate the variance decomposition in the largest possible sample (*All observations*), in the subsample of observations without board discretion (*No discretion*), and in the subsample of observations with at least one ESG metric (*Min. 1 ESG metric*). For more information on the variance decomposition approach, see Section 4.5.1.3.

Table 4.8: Discretionary ESG Metrics and RSTI Variance

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Discretionary STI Variance $Var\left(RSTI^D\right) \times 10^{-6}$				
ESG share of discretionary metrics (%)	−2080.86**	−2044.38*	−2760.96**	−2670.49**	−3093.27**
	(1040.10)	(1070.54)	(1187.54)	(1239.61)	(1219.48)
Log(Base Salary)					119127.90**
					(59060.99)
Tenure					−892.17
					(3969.61)
Age					3013.89
					(3853.91)
Female					49231.00
					(39523.44)
Industry F.E.	No	Yes	No	Yes	Yes
Executive position F.E.	No	No	Yes	Yes	Yes
Observations	243	243	238	238	227
Adjusted R^2	0.01	0.03	0.06	0.08	0.09

Notes: This table shows cross-sectional regressions at the executive level. The dependent variable is the variance of $RSTI^D$ ($\times 10^{-6}$), defined as the variance of the part of STI that is paid for performance along discretionary metrics. We winsorize $Var\left(RSTI^D\right)$ at the 1% level in the cross-section of executives. To calculate the main regressor *ESG share of discretionary metrics (%)*, we compute the ratio (in percent) of discretionary ESG metrics over total discretionary metrics (ESG and non-ESG) and take the average of this ratio over all sample years for each executive. Controls for executives' fixed base salaries and demographic characteristics (averaged over sample years), as well as industry and position fixed effects are included as indicated. If an executive switches positions during the sample period, we take the fixed effect with respect to the position held for the majority of years (i.e., the mode). Robust standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

In our largest sample with 1,251 observations, binding ESG metrics account for only 1.5% of total STI variance (column 4). Binding non-ESG performance metrics have by far the largest variance share of 128.4% (column 6), followed by discretionary ESG and non-ESG metrics with a share of 73.3% (column 8). The contribution of the covariance terms with binding ESG metrics are negligible (columns 9 and 10). However, the covariance term between binding non-ESG metrics and discretionary metrics is −1.021 (column 11), which is largely mechanical as weights must sum to 100%.[55]

As discussed in Section 4.4, some firms (i.e., their supervisory boards and/or compensation committees) reserve the right to make a final adjustment to total STI at their discretion. Such adjustments can change total pay variance, and it is unclear how these changes in overall pay variance should be attributed to the individual variance shares. However, Table 4.7 shows that the variance shares change only marginally if we drop these firms with supervisory board discretion and focus on the remaining subset of 1,091 observations.

As shown in Figure 4.2, 40% to 60% of the executives in our sample (black, dashed line in Panel A) do not have a single ESG criterion in their STI. For these executives, the contribution of ESG metrics to total STI variance is (mechanically) zero. After excluding the corresponding observations and retaining only executives with at least one ESG metric, the weight of binding ESG metrics increases from, on average, 2.7% to 5.0% (column 3), but the variance share of these metrics remains low at 3.4% (column 4). Hence, binding ESG metrics contribute little to total STI variance, even in a sample of executives whose contracts are all related to at least one ESG metric.

The first take-away from the variance decomposition is that *binding* ESG performance metrics contribute very little to STI pay risk, which is not surprising given their small weight in STI calculations. However, ESG could still matter because *discretionary* metrics are, in part, linked to ESG targets and account for a large variance share of 73.3% overall. As explained, we cannot decompose this 73.3% further into another ESG and another non-ESG variance share. Instead, we analyze whether this variance of discretionary STI pay, $Var(RSTI^D)$, is different for executives with a larger number of *ESG* targets among their discretionary metrics.

Table 4.8 shows cross-sectional regressions that use executives' $Var(RSTI^D)$ as the dependent variable. The main regressor is the share of discretionary metrics that are linked to ESG targets. The regressions show that $RSTI^D_{i,t}$ has a smaller variance for executives with a relatively larger number of ESG targets among the discretionary metrics, and that this result

[55] As $w^{B,ESG}_{i,t}$ and, therefore, changes $\Delta w^{B,ESG}_{i,t}$ are close to zero, we have $\Delta w^{B,nESG}_{i,t} \approx -\Delta w^{D}_{i,t}$, implying a negative correlation between binding non-ESG and discretionary metrics.

is robust to controlling for different fixed effects and executive characteristics. A one standard deviation increase in the ESG share decreases the variance of $RSTI^D_{i,t}$ by about 30% relative to the mean in column 5.[56] Hence, more ESG targets are associated with lower pay risk arising from discretionary pay. This result is consistent with previous analyses showing that discretionary metrics are insensitive to ESG performance (see Tables 4.5 and 4.6).

Overall, our variance decomposition indicates that both binding and discretionary ESG metrics contribute little to the pay risk associated with STI compensation plans. Executives may, of course, still be concerned about ESG performance—for example, because their compensation remains sensitive to financial performance, which may itself depend on ESG outcomes. However, the inclusion of ESG metrics in STI plans does not introduce additional pay risk. It therefore seems unlikely that ESG metrics materially increase executives' concern for ESG performance.

4.5.2 Contract Revisions Following Negative ESG News

Why do firms include ESG metrics with limited incentive power? One possible explanation is that firms are still learning how to design ESG-linked compensation plans efficiently. After all, ESG-linked pay has gained momentum only in recent years. In this section, we analyze how firms update and redesign STI compensation packages in response to bad ESG performance in the past. In particular, we test how firms change the weights of binding and discretionary ESG metrics.

Table 4.9 shows panel regressions with the weights $w^{B,ESG}_{i,t}$, $w^{B,nESG}_{i,t}$, and $w^D_{i,t}$ as dependent variables. In Panel A, the main regressor counts ESG events with negative press coverage that explicitly mention the name of the firm of executive i and precisely specify the underlying ESG incident. We lag this regressor by one year. As we control for executive fixed effects, identification is only based on within-executive time variation. For a given executive, the coefficient of *Lagged bad ESG news* tells us how much firms reallocate weight between binding and discretionary metrics following years with bad ESG news.

Panel A, columns 1 to 3 show the estimates from linear fixed-effects regressions. Only the coefficient in column 3 is significant (at the 10% level). Its negative value of -5.528 implies that firms reduce the weight of discretionary metrics by about 7.4 percentage points when *Lagged bad ESG news* increases by one standard deviation. As weights must sum to 100%, this

[56] The ESG share has a standard deviation of 14.86% in the regression sample, and the dependent variable a sample mean of $147,859$.

decrease in $w_{i,t}^{D}$ implies that the total weight of binding metrics, $w_{i,t}^{B,ESG} + w_{i,t}^{B,nESG}$, increases by the same amount. Indeed, the coefficients in Panel A, columns 1 to 3 sum up to zero. The (insignificant) coefficients in columns 1 and 2 suggest that relatively more weight is allocated to binding *non-ESG* metrics.

Distinguishing between the ESG categories, we find that contract changes are insignificant for governmental news (Panel C), but significant for (lagged) environmental and social news (Panel B). In Panel B, columns 1 to 3, the linear fixed effects regressions suggest that a one-standard-deviation increase of *Lagged bad ES news* (1.269 in column 3) reduces the weight of discretionary metrics by 6.70 percentage points. Again, binding *non-ESG* metrics benefit more and see their weight increase by 5.91 percentage points for the one-standard-deviation increase of *Lagged bad ES news*.

A limitation of linear regressions is their potential inefficiency when dependent variables are skewed and feature a mass of values at zero, as is the case for $w_{i,t}^{B,ESG}$, $w_{i,t}^{B,nESG}$, and $w_{i,t}^{D}$. Cohn et al. (2022) show that fixed-effects Poisson models produce consistent and more efficient estimates in those cases.[57] We implement fixed-effects Poisson regressions in columns 4 to 6. Again, we find that firms allocate weight from discretionary metrics towards binding metrics. As in the linear model, this reallocation is largest when we focus on environmental and social news (Panel B, columns 4 to 6). The coefficients in these Poisson regressions can be interpreted as semi-elasticities of 36% (Panel B, column 4), 6.93% (column 5), and −18.8% (column 6), which are all three statistically significant.[58]

Overall, Table 4.9 suggests that firms redesign STI plans after negative ESG press coverage, and reduce the weight of discretionary metrics. As shown in Section 4.5.1.2, the sensitivity to ESG performance of these discretionary metrics is notably smaller than for binding ESG metrics. Reallocating weight from discretionary to binding metrics could, therefore, strengthen ESG incentives. However, whether firms indeed do so is less clear from Table 4.9. In fact, some regressions suggest that especially the *non-ESG* criteria gain weight among the binding metrics. Overall, it seems that firms are tightening contracts and reducing discretion, but do so in a way that emphasizes financial performance rather than ESG performance per se.

[57] This is also true for continuous dependent variables.

[58] As coefficients do not measure weight changes in percentage points, they need not sum to zero in columns 4 to 6.

Table 4.9: Contract Revisions following Negative ESG News

Panel A: All ESG news

Linear fixed-effects regressions				Fixed-effects Poisson regressions			
	(1)	(2)	(3)		(4)	(5)	(6)
Dependent variable:	$w_{i,t}^{B,ESG}$	$w_{i,t}^{B,nESG}$	$w_{i,t}^{D}$	Dependent variable:	$w_{i,t}^{B,ESG}$	$w_{i,t}^{B,nESG}$	$w_{i,t}^{D}$
Lagged bad ESG news	0.404	5.119	−5.528*	*Lagged bad ESG news*	0.107	0.079*	−0.157*
	(0.439)	(3.108)	(3.039)		(0.186)	(0.046)	(0.083)
Observations	2,443	2,443	2,443	Observations	529	2,283	1,612
Adjusted R^2	0.71	0.61	0.60	Pseudo R^2	0.52	0.31	0.50

Panel B: Environmental and social news

Linear fixed-effects regressions				Fixed-effects Poisson regressions			
	(1)	(2)	(3)		(4)	(5)	(6)
Dependent variable:	$w_{i,t}^{B,ESG}$	$w_{i,t}^{B,nESG}$	$w_{i,t}^{D}$	Dependent variable:	$w_{i,t}^{B,ESG}$	$w_{i,t}^{B,nESG}$	$w_{i,t}^{D}$
Lagged bad ES news	0.613	4.661*	−5.279**	*Lagged bad ES news*	0.307**	0.067*	−0.208**
	(0.423)	(2.532)	(2.635)		(0.155)	(0.035)	(0.088)
Observations	2,443	2,443	2,443	Observations	529	2,283	1,612
Adjusted R^2	0.71	0.61	0.60	Pseudo R^2	0.52	0.31	0.50

Panel C: Governance news

Linear fixed-effects regressions				Fixed-effects Poisson regressions			
	(1)	(2)	(3)		(4)	(5)	(6)
Dependent variable:	$w_{i,t}^{B,ESG}$	$w_{i,t}^{B,nESG}$	$w_{i,t}^{D}$	Dependent variable:	$w_{i,t}^{B,ESG}$	$w_{i,t}^{B,nESG}$	$w_{i,t}^{D}$
Lagged bad G news	0.234	3.507	−3.745	*Lagged bad G news*	−0.099	0.059	−0.077
	(0.483)	(3.378)	(3.376)		(0.164)	(0.047)	(0.113)
Observations	2,443	2,443	2,443	Observations	529	2,283	1,612
Adjusted R^2	0.71	0.61	0.60	Pseudo R^2	0.52	0.31	0.50
Executive F.E.	Yes	Yes	Yes	Executive F.E.	Yes	Yes	Yes
Industry × Year F.E.	Yes	Yes	Yes	Industry × Year F.E.	Yes	Yes	Yes

Notes: This table reports panel regressions at the executive-year level that test whether the design of STI pay in year t depends on the firms' past press coverage in year $t-1$. Columns 1 to 3 (columns 4 to 6) show coefficient estimates from linear fixed-effects regressions (fixed-effects Poisson regressions). The dependent variable is the weight $w_{i,t}^{B,ESG}$ of binding ESG performance metrics in columns 1 and 4, the weight $w_{i,t}^{B,nESG}$ of binding non-ESG performance metrics in columns 2 and 5, and the weight $w_{i,t}^{D}$ of all discretionary performance metrics combined (without further distinguishing between ESG and non-ESG) in columns 3 and 6. The main regressor is the logarithm of one plus the one-year lagged number of ESG incidents with negative press coverage. We only consider incidents that explicitly mention the name of the firm and precisely specify the underlying ESG incident (called "sharp" incidents in RepRisk). We consider news about all three ESG categories in Panel A, but focus on environmental and social news in Panel B, and on governance news in Panel C. See C.1 for details. Executive fixed effects and industry-by-year fixed effects are included throughout. Robust standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

4.5.3 Heterogeneity

Our previous analyses allow conclusions for the average executive in our sample. In this section, we study potential differences across industry sectors and executive positions.

4.5.3.1 Industry Differences

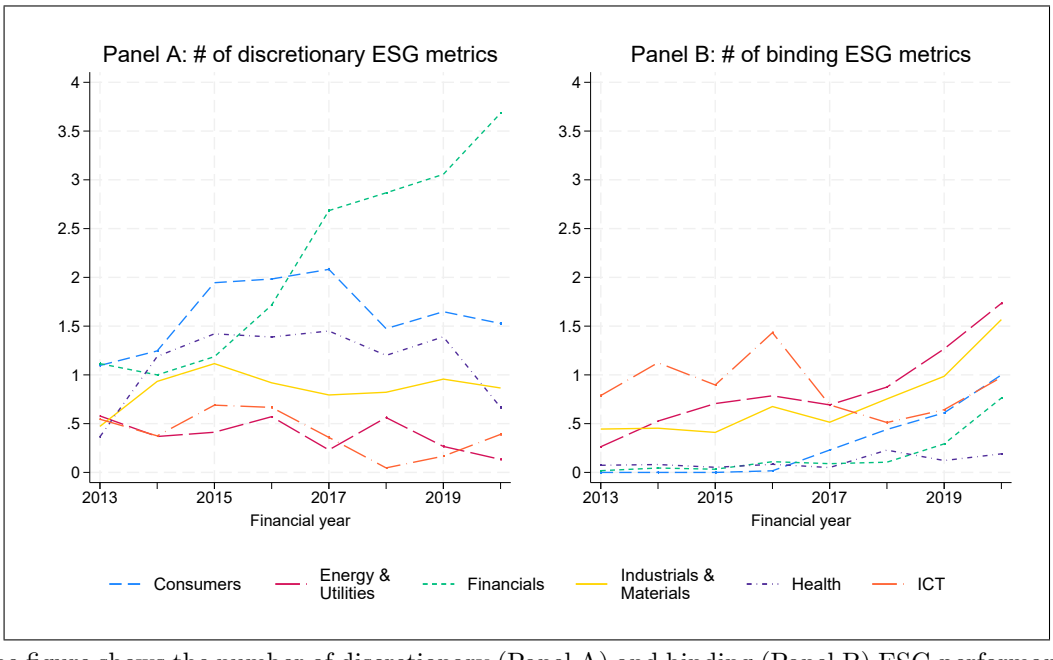
Objectives to link executive pay to ESG performance likely vary across industries. For example, in sectors that also benefit financially from strong ESG performance (doing well by doing good), firms' interest in designing effective ESG incentives could be particularly strong, especially if ESG outcomes add relevant information about executives' work that is not contained in non-ESG performance metrics.[59]

In Figure 4.5, we begin by plotting the numbers of discretionary and binding ESG performance metrics per executive by industry. We observe a striking difference in the prevalence of both types of performance metrics, that is, binding and discretionary ESG metrics are common in different industries. In particular, *discretionary* metrics are most popular in the financial sector, and least common in the energy & utilities sector, whereas the opposite is true for *binding* ESG metrics. Financial firms rapidly increase the number of discretionary ESG metrics from 1 to approximately 4 between 2013 and 2020. These same financial firms have no binding ESG metrics until 2018, and only about 0.75 binding metrics per executive in 2020. By contrast, in the energy & utilities sector, companies start with only 0.5 discretionary ESG metrics in 2013 and further reduce this number to about 0.25 metrics in 2020. For binding ESG metrics, energy & utilities take the opposite trend and increase the number of binding metrics per executive from about 0.25 to 1.75 in 2020.

Unsurprisingly, the industries that favor binding ESG performance metrics over discretionary ones also have the highest weights for these binding metrics in the year-end calculation of total STI. Figure 4.6 shows that industrial & materials, ICT, and energy & utility firms commit to weights that are between 2 and 3 times larger (depending on the year) than in the financial, consumer, and health care industries. These higher ESG weights also translate into a larger contribution of binding ESG metrics to the overall pay risk that STI poses to executives. Binding ESG metrics contribute 4.8% to within-executive STI variance in energy & utilities, but only 0.7% in the financial services sector (see Table IC.4 of the Internet Appendix C.5).

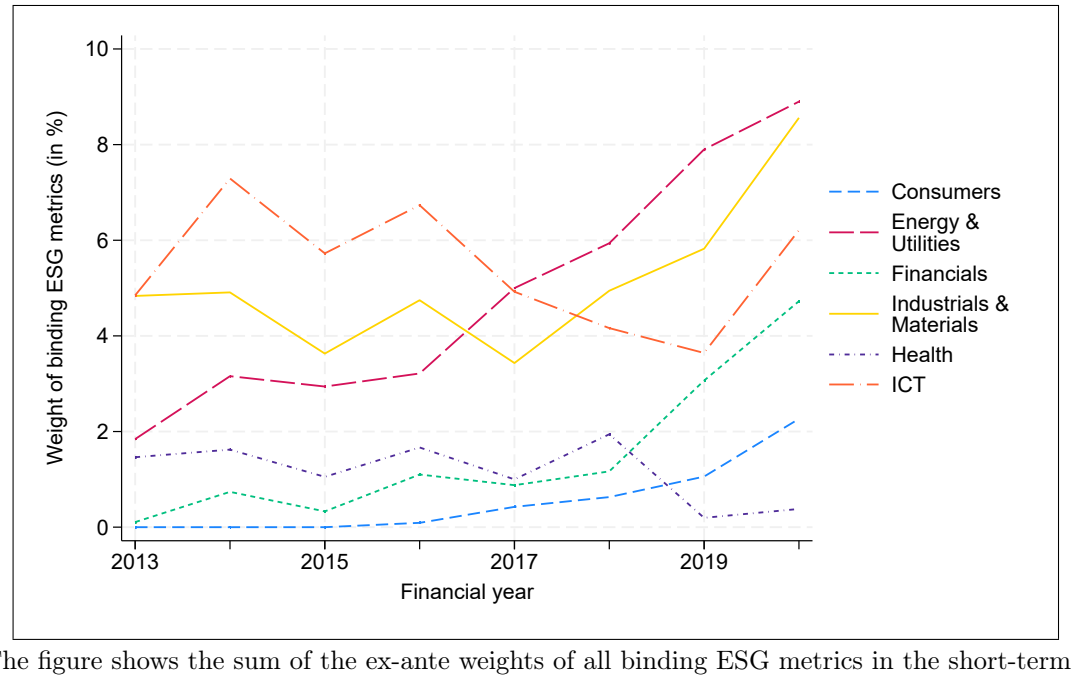
[59] See the informativeness principle by Holmström (1979). Equity-based pay should already provide executives with some ESG incentives in firms that can do well by doing good. However, stock prices might not be fully informative about executives' ESG actions, for example, if markets do not fully understand the effect of ESG on long-term value creation (e.g., Cohen et al. (2023); Edmans et al. (2017); Garel et al. (2024)). In those cases, adding ESG performance metrics to compensation plans could be consistent with optimal contracting.

Figure 4.5: Number of Binding and Discretionary ESG Performance Metrics in STI by Industry



Notes: The figure shows the number of discretionary (Panel A) and binding (Panel B) ESG performance metrics in the short-term incentive contract (STI) of the average executive in a given industry. For more information about the design of discretionary and binding performance metrics, see Section 4.4. Industry classification is based on the primary Global Industry Classification Standard (GICS) sector code, which we aggregate to the six major industry sectors Consumers, Energy & Utilities, Financials, Industrials & Materials, Health, and Information & Communication Technology (ICT).

Figure 4.6: Weight of Binding ESG Performance Metrics in STI by Industry



Notes: The figure shows the sum of the ex-ante weights of all binding ESG metrics in the short-term incentive contract (STI) of the average executive in a given industry. For more information about the design of binding ESG metrics and their weights, see Section 4.4. Industry sectors are defined as in Figure 4.5.

Overall, binding ESG metrics, which are shown to be more sensitive to ESG performance (Tables 4.5 and 4.6), are more prevalent in sectors that are likely to benefit financially from improved ESG performance, for example by reducing their exposure to rising energy prices and emission allowance costs. By contrast, other sectors—and financial institutions in particular—tend to avoid binding ESG metrics. This lack of commitment is unlikely to stem solely from a lack of hard or verifiable ESG measures. Indeed, a growing literature demonstrates that sustainable investments by banks and mutual funds can be measured with considerable precision and is likely contractible. Instead, the financial sector shows a strong preference for discretionary ESG metrics, which firms remain free to disregard in actual pay calculations.

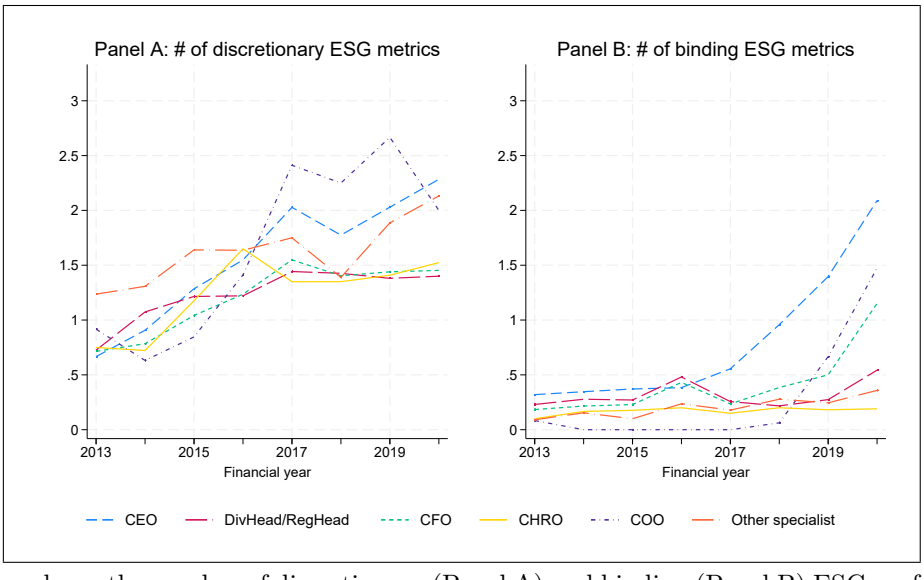
4.5.3.2 Differences across Executive Positions

It is reasonable to assume that firms tailor the design of incentive pay to the tasks and responsibilities of different executive positions. For example, a metric tied to employee satisfaction might be an informative signal of the effort exerted by the CHRO, whereas a metric tied to the adoption of clean technologies and emissions is likely more informative about the work of a CTO. In general, the more specialized a given executive role, the more tailoring we would expect. By contrast, measures of *firm-wide performance* (e.g., stock price or earnings growth) would be noisy signals of specialists' effort, as these metrics depend on many factors outside the control of these specialists. However, firm-wide performance could still be an appropriate metric in the compensation plans of executives that have control over all aspects of corporate policy. An example for such a generalist would be the CEO, whose job is to set the overarching, long-term strategy and vision for the *entire* firm.[60] In the following, we analyze how ESG-linked pay varies across executives and, in particular, between *generalists* and *specialists*.

In Figure 4.7, we begin by studying the numbers of ESG performance metrics by executive position. Between 2013 and 2016, metric numbers are roughly the same across all different positions. However, after 2016, we observe an increase in discretionary ESG metrics (Panel A) for CEOs, COOs, and for our catch-all category "Other specialist", which groups all executive positions with very few observations. After 2016, CEOs and COOs also experience a strong increase in the number of binding ESG metrics (Panel B).

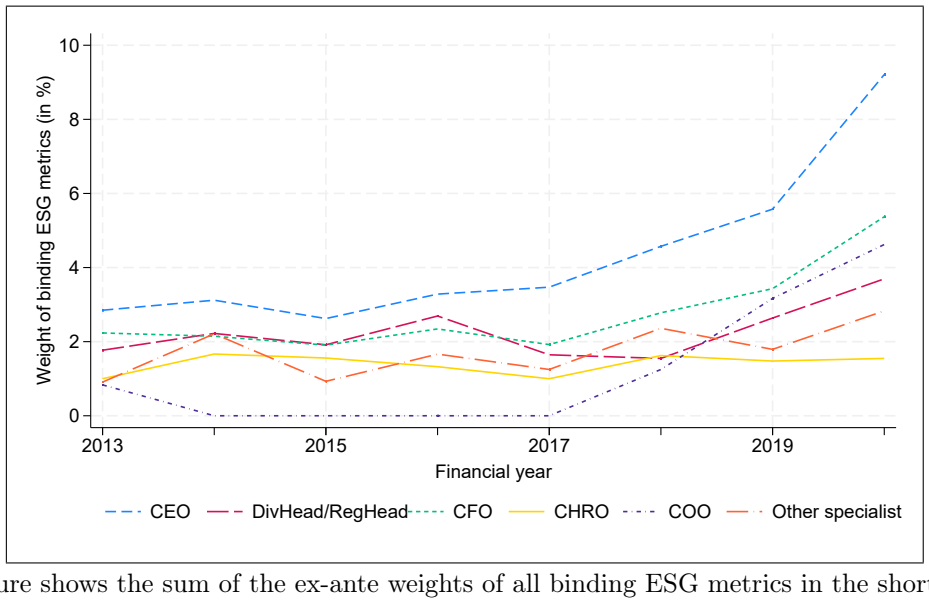
[60] For example, if the objective of a firm is pure shareholder value maximization and the CEO controls all aspects of corporate policy, it might be optimal to tie her compensation simply to the stock price and to let her decide how she best allocates her time and effort to different corporate policies.

Figure 4.7: Number of Binding and Discretionary ESG Performance Metrics in STI by Board Position



Notes: The figure shows the number of discretionary (Panel A) and binding (Panel B) ESG performance metrics in the short-term incentive contract (STI) of the average executive in a given board position. We differentiate between the chief executive officer (CEO), chief financial officer (CFO), chief human resources officer (CHRO), and chief operating officer (COO). "DivHead/RegHead" denotes executives who are chairing a division, region, or are responsible for a product segment. All other expert C-suite positions (e.g., chief marketing officer, chief sales officer, chief legal officer, etc.) are grouped as "Other specialist".

Figure 4.8: Weight of Binding ESG Performance Metrics in STI by Board Position



Notes: The figure shows the sum of the ex-ante weights of all binding ESG metrics in the short-term incentive contract (STI) of the average executive in a given board position. For more information about the design of binding ESG metrics and their weights, see Section 4.4. For a description of the different executive C-suite positions, see Figure 4.7.

Similarly, Figure 4.8 shows that after 2016, firms commit to increasingly large ESG weights in the STI of CEOs, COOs, and also CFOs.[61] As these three positions tend to have a broader spectrum of tasks, we consider them generalist positions.[62] Hence, we cannot confirm our initial prediction that especially the pay of specialists should be tied to (ESG) metrics other than financial performance.

Ideally, to test how much firms tailor ESG-linked pay, we would compare the tasks and responsibilities of each individual to the ESG performance metrics used in her contract. However, executive positions indicate only broad corporate policies, and our data do not clarify how these should correspond to specific ESG performance metrics. Nonetheless, in two cases, this alignment appears relatively straightforward. For example, reducing a firm's environmental footprint often involves developing new technologies and replacing stranded assets. Therefore, we would expect chief technology officers (CTOs) to have more environmental ESG metrics, particularly those linked to emissions. Similarly, a firm aiming to improve working conditions and employee wellbeing will likely assign this task to its CHRO. As a result, CHROs are expected to have social ESG metrics, especially workforce-related metrics, more often than other executives. To test these predictions, we estimate the following linear probability model:

$$y_{i,t} = \alpha + \beta_1 CTO_{i,t} + \beta_2 CHRO_{i,t} + Firm \times Year\ FE + \epsilon_{i,t} \tag{4.7}$$

where the dependent variable $y_{i,t}$ equals one if the STI of executive i is linked to a given ESG metric in year t, and zero otherwise. The regressors $CTO_{i,t}$ and $CHRO_{i,t}$ are also binary and equal one if executive i is a CTO or, respectively, a CHRO, and zero otherwise. We condition the estimation on firm fixed effects interacted with year fixed effects. The regression sample is restricted to CTOs and CHROs (*specialists*) and CEOs (*generalists*). Hence, the coefficient β_1 tells us whether a given CTO is more or less likely to have metric $y_{i,t}$ than the CEO of the same firm in the same year. Similarly, β_2 compares the propensity of a given metric between the CHRO and the CEO in the same firm and year. In keeping with our prediction of "tailoring" metrics to specialist positions, we expect β_1 to be positive for environmental metrics, and β_2 to be positive for social metrics.

[61] In our variance decomposition, the variance share of binding ESG metrics is marginally larger for CEOs, CFOs, and COOs than for other positions, but remains small in absolute terms (see Table IC.5 of the Internet Appendix C.5)

[62] We consider CEOs, CFOs, and COOs as generalists. CEOs are responsible for overall firm performance, for integrating the different corporate policies into a unified strategy, and for shaping the firm's strategic vision. COOs act as the CEO's right hand, overseeing all daily, internal operations. CFOs also require, besides expert knowledge in finance and financial reporting, a high-level understanding of firm-wide strategy and broader management responsibilities, as their financial decisions affect all other corporate policies (Brickley et al. (2017)).

Table 4.10: Tailoring of ESG Metrics to Executive Positions

	(1)	(2)	(3)	(4)
Panel A: Environmental metrics				
	Environm. Y/N		Emissions Y/N	
CTO dummy	-0.068***	-0.066**	-0.053***	-0.052**
	(0.023)	(0.026)	(0.020)	(0.024)
CHRO dummy	-0.037*	-0.040**	-0.034*	-0.034**
	(0.020)	(0.017)	(0.018)	(0.016)
Firm F.E. and Year F.E.	Yes	No	Yes	No
Firm-year F.E.	No	Yes	No	Yes
Observations (CEO)	571	209	571	209
Observations (CTO)	92	92	92	92
Observations (CHRO)	158	155	158	155
Adjusted R^2	0.51	0.73	0.53	0.75
Panel B: Social metrics				
	Social Y/N		Workforce Y/N	
CTO dummy	-0.039	-0.065**	-0.050	-0.066**
	(0.039)	(0.028)	(0.039)	(0.027)
CHRO dummy	-0.032	-0.025	-0.004	-0.012
	(0.028)	(0.017)	(0.029)	(0.018)
Firm F.E. and Year F.E.	Yes	No	Yes	No
Firm-year F.E.	No	Yes	No	Yes
Observations (CEO)	571	209	571	209
Observations (CTO)	92	92	92	92
Observations (CHRO)	158	155	158	155
Adjusted R^2	0.61	0.91	0.59	0.89

Notes: Reported are panel regressions for the adoption of different ESG performance metrics in the STI contracts of CTOs, CHROs, and CEOs. The dependent variables are binary indicators that equal one if the STI of executive i in year t is linked to a given type of ESG metric: environmental metrics in general in columns 1 and 2 of Panel A; metrics measuring emissions in columns 3 and 4 of Panel A; social metrics in general in columns 1 and 2 of Panel B; workforce-related metrics in columns 3 and 4 of Panel B. The independent variables *CTO dummy* and *CHRO dummy* are binary indicators that equal one if executive i is a CTO or a CHRO, respectively (and zero otherwise). The omitted base category consists of CEOs. Firm, year, and interacted firm-year fixed effects are included as indicated. Robust standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

In Table 4.10, we report the coefficient estimates for Equation (4.7). In columns 1 and 2 of Panel A, the dependent variable equals one if executive i has an environmental metric in her compensation plan. The coefficients are all negative and significant, indicating that CTOs are less and not more likely than CEOs to have environmental metrics. Moreover, CTOs are not even more likely to have environmental metrics than CHROs. This result remains unchanged when we look specifically at metrics to reduce emissions (columns 3 and 4 of Panel A). Finally, Panel B shows no evidence that social metrics in general (columns 1 and 2) and workforce-related metrics in particular (columns 3 and 4) are more common among CHROs than among CEOs. If anything, the opposite seems to be true.

Overall, our findings run counter to our prediction that firms tailor ESG metrics to the tasks of different specialists. Instead, ESG-linked pay is more common among generalist positions like the CEO. Coincidentally, these are also positions that tend to be more visible. Indeed, public attention for compensation is generally higher for CEOs than for other specialist positions further down the hierarchy. Possibly, firms do not include ESG metrics for incentive reasons but for signaling to external audiences.

4.6 Conclusion

Overall, our evidence suggests that the rise of "ESG-linked" short-term incentive (STI) pay among large European firms translates into only limited incremental ESG incentives. In particular, three patterns jointly dampen the incentive power of ESG pay. First, executives face substantially more non-ESG than ESG metrics, and in most cases the few ESG metrics are purely discretionary with firms refraining from specifying in advance any explicit weight these metrics will carry in the STI calculation. Second, where firms do commit to binding ESG metrics, achievement rates are meaningfully sensitive to verifiable ESG outcomes (such as emissions intensity and employee turnover) and to salient negative ESG news, implying that binding ESG metrics can, in principle, generate incentives. Yet, these binding ESG metrics are rare and command only modest weight in STI calculations. Third, the more common "discretionary" component of STI metrics—which pools ESG and non-ESG considerations without any quantified commitment to either—exhibits little sensitivity to either verifiable ESG performance or high-reach negative ESG press events. Consistent with this, a variance decomposition shows that ESG performance metrics explain only a small share of within-executive pay variance, suggesting again that executives will mostly be concerned about the effects of conventional performance on pay.

Firms do appear to update compensation design following negative ESG news, but not primarily by strengthening explicit ESG contracting. After such events, companies shift weight away from discretionary assessment toward binding metrics, indicating a general tightening of STI design and reduced ex post discretion. However, the reallocated weight accrues mainly to binding non-ESG criteria rather than to binding ESG metrics, implying that "disciplining" after ESG setbacks is implemented through more conventional, financially-oriented performance contracting.

Finally, we document substantial heterogeneity. Industries with greater exposure to environmental costs and regulatory constraints rely relatively more on binding ESG metrics with higher weights, whereas financial firms disproportionately favor discretionary ESG criteria. Across positions, ESG-linked pay is more prevalent for visible generalists like CEOs than for functional specialists, and we find little evidence that firms systematically tailor environmental and social metrics to executives whose roles would make such targets especially informative (e.g., CTOs and CHROs).

Taken together, these findings suggest that ESG metrics in STI contracts currently play a secondary and limited role in shaping executive incentives rather than serving as a central driver of compensation outcomes. Binding ESG metrics can "work" in the narrow sense of being sensitive to changes in ESG performance, but their limited prevalence and small weights mean that ESG contracting rarely increases executives' exposure to ESG-related pay risk relative to conventional plans. This pattern is consistent with ESG-linked pay primarily fulfilling symbolic, communicative, or transitional functions, while firms continue to rely mainly on financially oriented performance measures when designing high-powered incentive contracts.

Appendix C

C.1 Variable Definitions

Compensation variables:

- *Base salary*$_{i,t}$: Includes all contractually agreed on fixed compensation components, paid regardless of business, company-wide and/or individual performance. (source: hkp, own data collection)
- $RLTI_{i,t}$: Realized long-term incentive compensation defined as the sum of all variable compensation elements (in Euros) based on a performance period of more than one year. The following plan types are considered: cash and equity deferrals, share matching plans, multi-year bonuses, restricted stock, performance shares, and stock options. In case a long-term incentive (LTI) plan is not granted annually, but upfront for a multi-year period, the grant value is distributed over this period and the pro-rated value is shown for each year. (source: hkp, own data collection)
- $RSTI_{i,t}$: Realized short-term incentive compensation defined as the sum of all variable compensation elements (in Euros) with a performance period of one year paid to executive i in fiscal year t. Payment is made in cash at or shortly after the end of the performance period. (source: hkp, own data collection)
- $TSTI_{i,t}$: Target short-term incentive compensation defined as the amount (in Euros) that the firm targets to pay to executive i in fiscal year t if the executive fully meets her targets, i.e., neither over- nor underperforms. In practice, firms typically report $TSTI_{i,t}$ as a proportion of base salary or as a proportion of a bonus cap/STI cap. Occasionally, a target range is reported instead, in which case we take the midpoint. (source: hkp, own data collection)
- $w_{i,t}^{B,ESG}$: Total weight of all *binding* ESG performance metrics of executive i in fiscal year t. In the case of binding metrics, the firm commits to consider target achievement with a weight that is communicated at the beginning of fiscal year t. (source: own data collection)

- $w_{i,t}^{B,nESG}$: Total weight of all *binding* non-ESG performance metrics of executive i in fiscal year t. In the case of binding metrics, the firm commits to consider target achievement with a weight that is communicated at the beginning of fiscal year t. (source: own data collection)

- $w_{i,t}^{D}$: Total weight of all *discretionary* performance metrics of executive i in fiscal year t, calculated as $w_{i,t}^{D} = 1 - w_{i,t}^{B,ESG} - w_{i,t}^{B,nESG}$. Ex ante, executive i knows that the firm will consider the achievement of all discretionary metrics jointly with weight $w_{i,t}^{D}$ in the calculation of compensation at year-end. Specific weights for individual discretionary metrics are not known. (source: own data collection)

- $f_{i,t}$: Overall target achievement rate for all performance metrics together, defined as the weighted average of specific target achievement rates for individual binding metrics and all discretionary metrics combined (Equation (4.2)). A value of $f_{i,t}$ =100% means that executive i has fully met her targets, i.e., neither over- nor underperformed in fiscal year t. Values reported in the tables are values as reported in firms' annual statements, i.e., after discretionary adjustments by the supervisory board or compensation committee. (source: own data collection)

- *STI hurdle* $Y/N_{i,t}$: Dummy variable indicating whether a firm reports using a lower threshold for realized STI. If actual performance falls below this lower threshold, realized $RSTI_{i,t}$ is set to a minimum value, typically zero. (source: own data collection)

- *LTI hurdle* $Y/N_{i,t}$: See *STI hurdle* $Y/N_{i,t}$, but for LTI. (source: own data collection)

- *STI cap* $Y/N_{i,t}$: Dummy variable indicating whether STI amount is capped at an upper threshold. (source: own data collection)

- *LTI cap* $Y/N_{i,t}$: See *STI cap* $Y/N_{i,t}$ but for LTI. (source: own data collection)

- *STI board discretion allowed* $Y/N_{i,t}$: Dummy variable indicating whether the supervisory board and/or compensation committee retains the right to subjectively adjust the STI amount ex post (e.g., to account for unpredicted exogenous events). (source: own data collection)

- *LTI board discretion allowed* $Y/N_{i,t}$: See *STI board discretion allowed* $Y/N_{i,t}$, but for LTI. (source: own data collection)

- *STI ex post multiplier (ESG/nESG)* $Y/N_{i,t}$: Dummy variable indicating whether STI has been adjusted ex post according to a multiplier determined by the supervisory board / compensation committee. (source: own data collection)

- *LTI ex post multiplier (ESG/nESG)* $Y/N_{i,t}$: See *STI ex post multiplier (ESG/nESG)* $Y/N_{i,t}$, but for LTI. (source: own data collection)
- *STI partly deferred* $Y/N_{i,t}$: Dummy variable indicating whether parts of the STI amount were deferred, i.e., paid out in later periods. (source: own data collection)
- *STI deferral period*$_{i,t}$: In case of deferral, this variable denotes the period in years over which the STI is to be deferred. (source: own data collection)
- *Board exercised discretion* $Y/N_{i,t}$: Dummy variable indicating whether supervisory board and/or compensation committee has ultimately exercised its right to subjectively adjust the STI amount ex post. (source: own data collection)

Executive characteristics:

- $Age_{i,t}$: Age of the executive in years. (source: hkp, own data collection)
- $Tenure_{i,t}$: Tenure of the executive in years. (source: hkp, own data collection)
- *Female* $Y/N_{i,t}$: Gender of the executive. (source: hkp, own data collection)
- C-suite position: (source: own data collection)
  - *CEO*: Chief Executive Officer
  - *CFO*: Chief Financial Officer
  - *CHRO*: Chief Human Resources Officer
  - *COO*: Chief Operating Officer
  - *Chairman*: Chairman (e.g., Executive Chairman, Chairman of Executive Committee)
  - *Divisional/regional head*: Executives who are chairing a division, geographic region, or are responsible for a product segment
  - *Other specialists*: All other expert C-suite positions such as chief marketing officer, chief sales officer, chief legal officer, etc.

Company-level variables (all accessed via LSEG Data & Analytics):

- *Market capitalization*$_{f,t}$: Market price year-end multiplied with common shares outstanding (WC08001) (source: Worldscope)
- *Total assets*$_{f,t}$: Total assets (WC02999) (source: Worldscope)

- *Net sales*$_{f,t}$: Net sales or revenues represents gross sales and other operating revenue less discounts, returns, and allowances (WC01001) (source: Worldscope)
- *Book-to-market ratio*$_{f,t}$: Book value of common equity (WC03501) divided by market value of equity (WC08001) (source: Worldscope)
- *Book leverage*$_{f,t}$: Total debt (sum of long- and short-term debt) (WC03255) divided by total assets (WC02999) (source: Worldscope)
- *Net PPE/total assets*$_{f,t}$: Gross property, plant, and equipment (PPE) less accumulated reserves for depreciation, depletion, and amortization (WC02501) divided by total assets (WC02999) (source: Worldscope)
- *Return on assets*$_{f,t}$: ((Net income – bottom line + (interest expense on debt-interest capitalized) * (1 - tax rate))) / Average of last year's and current year's total assets (WC08326) (source: Worldscope)
- *DPS*$_{f,t}$: Dividends per share (WC05101) divided by earnings per share (WC05201) (source: Worldscope)
- *Stock return*$_{f,t}$: Stock (total investment) return of the company divided by last year's market value (WC08801) (source: Worldscope)
- *Historical stock-to-accounting volatility*$_{f}$: Standard deviation of total investment return (WC08801) divided by standard deviation of return on assets (WC08326) between the years 2008 and 2012 (source: Worldscope)
- *Board independence*$_{f,t}$: Share of independent board members (TR.AnalyticIndepBoard) (source: Refinitiv Eikon)
- *Female board membership*$_{f,t}$: Share of female board members (TR.AnalyticBoardFemale) (source: Refinitiv Eikon)
- *Institutional ownership*$_{f,t}$: Share of company's equity owned by institutional investors. Institutional investors include the following categories of financial institutions: investment advisors, hedge funds, pension funds, banks and trusts, insurance companies, sovereign wealth funds, venture capital, and private equity (Shareholders History Report) (source: Refinitiv Eikon)
- *Block ownership*$_{f,t}$: Share of owners that own more than 10% of company's equity (Shareholders History Report) (source: Refinitiv Eikon)
- *Emission policy Y/N*$_{f,t}$: Company policy to improve emission reduction (TR.PolicyEmissions) (source: Refinitiv Eikon)

- *Industry*$_f$: Industry classification is based on the primary Global Industry Classification Standard (GICS) sector code, which we aggregate to six major industry sectors comprising Consumers, Energy & Utilities, Financials, Industrials & Materials, Health, and Information & Communication Technology (ICT) (source: Refinitiv Eikon)

- *Total CO_2 emission equivalents*$_{f,t}$: Total CO_2 and CO_2 equivalents emission in tonnes Scope 1 and Scope 2 (TR.CO2EmissionTotal) divided by net sales or revenues which represents gross sales and other operating revenue less discounts, returns, and allowances (WC01001). The following gases are relevant: carbon dioxide, methane, nitrous oxide, hydrofluorocarbons, perfluorinated compound, sulfur hexafluoride, nitrogen trifluoride. Emission classifications based on greenhouse gas protocol (GHG) (source: Refinitiv Eikon/ Worldscope)

- *Historical Avg. Total CO_2 emission equivalents*$_f$: Average total CO_2 and CO_2 equivalents emission$_{f,t}$ in tonnes Scope 1 and Scope 2 (TR.CO2EmissionTotal) between 2008 and 2012. The following gases are relevant: carbon dioxide, methane, nitrous oxide, hydrofluorocarbons, perfluorinated compound, sulfur hexafluoride, nitrogen trifluoride. Emission classifications based on greenhouse gas protocol (GHG) (source: Refinitiv Eikon)

- *Employee Turnover*$_{f,t}$: Percentage of employee turnover, including employees who left the company for any reason (voluntary or involuntary) such as resignations, retirement, natural departure/death, medical incapacitation, redundancy, layoffs, restructuring, dismissal, retrenchment or end of a fixed-term contract (TR.TurnoverEmployeesScore) (source: Refinitiv Eikon)

- *ESG rating*$_{f,t}$: Yearly mean RepRisk Rating (RRR). The RRR facilitates corporate benchmarking against the peer group and sector of a company, as well as integration of ESG and business conduct risks into business processes. The Rating provides decision support in risk management, compliance, investment management, and supplier risk assessment. The original ESG risk scale, ranging from AAA (very low ESG risk exposure) to D (very high ESG risk exposure), was converted into a numerical rating (source RepRisk)

- *Reporting Scope*$_{f,t}$: The percentage of the company's activities covered in its Environmental and Social reporting (TR.CSRReportingScope) (source: Refinitiv Eikon)

- *High-reach bad ESG news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to environmental, social, or governance issues that are sharp and of high reach. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. High-reach incidents imply the news are being published by global media outlets (source: RepRisk)

- *High-reach bad E news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to environmental issues that are sharp and of high reach. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. High-reach incidents imply the news are being published by global media outlets. (source: RepRisk)

- *High-reach bad S news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to social issues that are sharp and of high reach. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. High-reach incidents imply the news are being published by global media outlets. (source: RepRisk)

- *High-reach bad G news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to governance issues that are sharp and of high reach. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. High-reach incidents imply the news are being published by global media outlets. (source: RepRisk)

- *Bad ESG news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to environmental, social, or governance issues that are sharp. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. (source: RepRisk)

- *Bad ES news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to environmental, or social issues that are sharp. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. (source: RepRisk)

- *Bad G news*$_{f,t}$: Natural logarithm of one plus the firm-level count of news related to governance issues that are sharp. Sharp incidents imply that the name of the firm is explicitly mentioned and the underlying ESG incident is precisely specified. (source: RepRisk)

C.2 Overview of Sample Firms

Table C.1: Overview of Sample Firms

Firm	Country	Sample period (FY)	STOXX Europe 50	EURO STOXX 50	Comments
ABB	CH	2013-2020	since 09/2008	-	-
Adidas	DE	2013-2020	since 09/2020	since 09/2016	-
Adyen	NL	2018-2020	since 09/2020	since 09/2020	First annual report published in 2018
Ahold Delhaize	NL	2016-2020	-	since 09/2016	Merger in 07/2016 of "Koninklijke Ahold" (NL) and "Delhaize Group" (BE)
Air Liquide	FR	2013-2020	since 09/2011	since 08/1998	-
Airbus	NL	2013-2020	since 09/2017	since 03/2013	-
Allianz	DE	2013-2020	since 08/1998	since 08/1998	-
Amadeus IT	ES	2013-2020	-	since 09/2018	-
Anheuser-Busch InBev	BE	2013-2020	since 09/2010	since 09/2009	-
ASML	NL	2013-2020	since 08/2016	since 06/2012	-
Assicurazioni Generali	IT	2013-2020	-	08/1998-09/2016	-
AstraZeneca	GB	2013-2020	since 09/2000	-	-
AXA	FR	2013-2020	since 09/2013	since 08/1998	-
Banco Santander	ES	2013-2020	09/1999-09/2020	since 09/1999	-
Barclays	GB	2013-2020	04/1999-12/2018	-	-
BASF	DE	2013-2020	since 09/2002	since 09/1999	-
Bayer	DE	2013-2020	since 09/2008	since 08/1998	-
BBVA	ES	2013-2020	08/1998-09/2019	08/1998-09/2020	-
BG Group	GB	2013-2015	04/2009-02/2016	-	Takeover by Royal Dutch Shell in 2016
BHP Group	AU/GB	2013-2020	09/2006-09/2015 & since 07/2020	-	Renamed from "BHP Billiton" to "BHP Group" in 2017/2018

Continued on next page

Table C.1 – continued from previous page

Firm	Country	Sample period (FY)	STOXX Europe 50	EURO STOXX 50	Comments
BMW	DE	2013-2020	-	since 09/2010	-
BNP Paribas	FR	2013-2020	since 09/2000	since 09/1999	-
BP	GB	2013-2020	since 08/1998	-	-
British American Tobacco	GB	2013-2020	since 09/2008	-	-
BT Group	GB	2013-2020	09/2014-09/2017	-	-
Carrefour	FR	2013-2020	-	08/1998-09/2016	-
Credit Suisse	CH	2013-2020	06/2013-08/2016	-	-
CRH	IE	2013-2020	-	09/2009-09/2014 & since 09/2016	-
Daimler	DE	2013-2020	since 10/1998	since 11/1998	-
Danone	FR	2013-2020	-	since 09/2000	-
Deutsche Bank	DE	2013-2020	08/1998-08/2016	08/1998-09/2018	-
Deutsche Börse	DE	2013-2020	-	since 09/2019	-
Deutsche Post	DE	2013-2020	-	since 09/2013	-
Deutsche Telekom	DE	2013-2020	since 08/1998	since 08/1998	-
Diageo	GB	2013-2020	since 08/1998	-	-
E.ON	DE	2013-2020	08/1998-09/2013	08/1998-09/2018	-
Enel	IT	2013-2020	since 09/2019	since 03/2000	-
ENGIE	FR	2013-2020	07/2008-03/2013	since 07/2008	Renamed from GDF Suez in 2015
Eni	IT	2013-2020	08/1998-07/2020	since 08/1998	-
EssilorLuxottica	FR	2013-2020	-	since 6/2012	Essilor merged in 2018 with Luxottica (IT), Renamed to EssilorLuxottica
Flutter Entertainment	IE	2016-2020	-	since 09/2020	Merged in 2016 from Betfair (GB) and Paddy Power (IE) to Paddy Power Betfair; Renamed to Flutter Entertainment in 2019

Continued on next page

Table C.1 – continued from previous page

Firm	Country	Sample period (FY)	STOXX Europe 50	EURO STOXX 50	Comments
Fresenius	DE	2013-2020	-	09/2015-09/2020	-
GlaxoSmithKline	GB	2013-2020	since 08/1998	-	-
Glencore	JE	2013-2020	06/2013-09/2015 & 05/2017-09/2019	-	-
HSBC	GB	2013-2020	since 09/1999	-	-
Iberdrola	ES	2013-2020	since 09/2019	since 09/2003	-
Imperial Brands	GB	2013-2020	12/2011-09/2013 & 09/2015-09/2018	-	2016 Renaming of Imperial Tobacco to Imperial Brands
Inditex	ES	2014-2020	-	since 09/2011	No data available for FY 2013
ING Groep	NL	2013-2020	08/1998-05/2020	since 08/1998	-
Intesa Sanpaolo	IT	2013-2020	since 09/2015	since 01/2007	-
Kering	FR	2013-2020	since 05/2020	since 09/2018	-
KONE	FI	2013-2020	-	since 09/2020	-
Linde SE	IE	2018-2020	since 12/2018	since 10/2018	Merger in 10/2018 of "Praxair" (US) and "Linde AG" (DE)
Lloyds Banking	GB	2013-2020	09/2013-09/2020	-	-
L'Oréal	FR	2013-2020	since 02/2016	since 08/1998	-
LVMH	FR	2013-2020	since 09/2011	since 08/1998	-
Münchener Rück	DE	2013-2020	-	since 09/1999	-
National Grid	GB	2013-2020	since 09/2011	-	-
Nestlé	CH	2013-2020	-	since 08/1998	-
Nokia	FI	2013-2020	-	08/1998-03/2013 & since 09/2014	-
Novartis	CH	2013-2020	since 08/1998	-	-
Novo Nordisk	DK	2013-2020	since 06/2015	-	-
Orange	FR	2013-2020	-	08/1998-09/2020	-
Pernod Ricard	FR	2013-2020	-	since 09/2020	-
Philips	NL	2013-2020	-	since 08/1998	-

Continued on next page

Table C.1 – continued from previous page

Firm	Country	Sample period (FY)	STOXX Europe 50	EURO STOXX 50	Comments
Prosus	NL	2020	since 11/2020	since 09/2020	Founded in 2019; first annual report for FY 2020
Prudential	GB	2013-2020	since 09/2014	-	-
Reckitt Benckiser	GB	2013-2020	since 06/2012	-	-
RELX Group	GB	2013-2020	since 09/2019	-	-
Repsol	ES	2013-2020	-	08/1998-09/2015	-
Richemont	CH	2013-2020	12/2012-09/2016	-	-
Rio Tinto	AU/GB	2013-2020	since 09/2005	-	-
Roche	CH	2013-2020	since 09/1999	-	-
Royal Dutch Shell	GB	2013-2020	since 12/2005	-	-
RWE	DE	2013-2020	-	08/1998-09/2015	-
Safran	FR	2013-2020	since 09/2018	since 09/2015	-
Saint-Gobain	FR	2013-2020	-	09/2001-09/2018	-
Sanofi	FR	2013-2020	since 09/2007	since 09/1999	-
SAP	DE	2013-2020	since 09/2003	since 07/2004	-
Schneider Electric	FR	2013-2020	since 03/2013	since 09/2007	-
Siemens	DE	2013-2020	since 08/1998	since 08/1998	-
Société Générale	FR	2013-2020	-	08/1998-09/2020 -	
Standard Chartered	GB	2013-2020	09/2010-06/2015	-	-
Syngenta	CH	2013-2016	09/2016-05/2017	-	Takeover by ChemChina in 2017
Telefónica	ES	2013-2020	08/1998-09/2019	08/1998-09/2020	-
Total	FR	2013-2020	since 11/1999	since 06/1999	-
UBS	CH	2013-2020	since 11/2014	-	-
Unibail-Rodamco-Westfield	FR	2018-2020	-	02/2010-09/2019	Takeover of "Westfield" (AUS) in 06/2018
UniCredit	IT	2013-2020	-	08/1998-09/2016	-
Unilever Group	GB	2013-2020	since 09/2011	-	-
Vinci	FR	2013-2020	since 08/2016	since 09/2007	-

Continued on next page

Table C.1 – continued from previous page

Firm	Country	Sample period (FY)	STOXX Europe 50	EURO STOXX 50	Comments
Vivendi	FR	2013-2020	-	since 08/1998	-
Vodafone	GB	2013-2020	since 09/1999	-	-
Volkswagen	DE	2013-2020	-	since 09/2011	-
Vonovia	DE	2015-2020	-	since 09/2020	Merger in 08/2015 of "Deutsche Annington" (DE) and "GAG-FAH" (LU)
Zurich Insurance	CH	2013-2020	since 09/2010	-	-

Notes: Index time specification includes changes up until FY2020

C.3 ESG Classification according to LSEG

The following table shows the underlying taxonomy which we use to classify the non-financial metrics in executive compensation contracts with respect to the E, S, and G dimension and its subcategories.

Table C.2: ESG Classification according to LSEG

Category	Subcategory	Exemplary metrics
Environmental	Emission	Emissions (scope 1-3), GHG protocol
		Emission reduction target
		Internal carbon pricing
		Waste & waste recycling
		Biodiversity
		Environmental management systems
	Innovation	Product innovation (environmental, organic, eco-design products)
		Green revenues (from environmental products)
		Environmental R&D expenditures
		Environmental assets under management
		Environmental project financing
	Resource use	Water
		Energy
		Toxic chemicals
		Renewable energy
		Sustainable packing policy
		Environmental supply chain management
Social	Community	Fair competition
		Bribery and corruption policy
		Whistleblower protection
		Donations
		Product Sales at Discount to Emerging Markets
		Corporate Responsibility Awards
		Crisis Management Systems
	Human rights	Human rights policy
		Child labor policy
		Ethical Trading Initiative
		Freedom of Association policy
	Product responsibility	Responsible marketing
		Customer Health & Safety policy
		Product quality
		Quality management system
		Customer satisfaction
		Healthy foods or products

Continued on next page

Table C.2 – continued from previous page

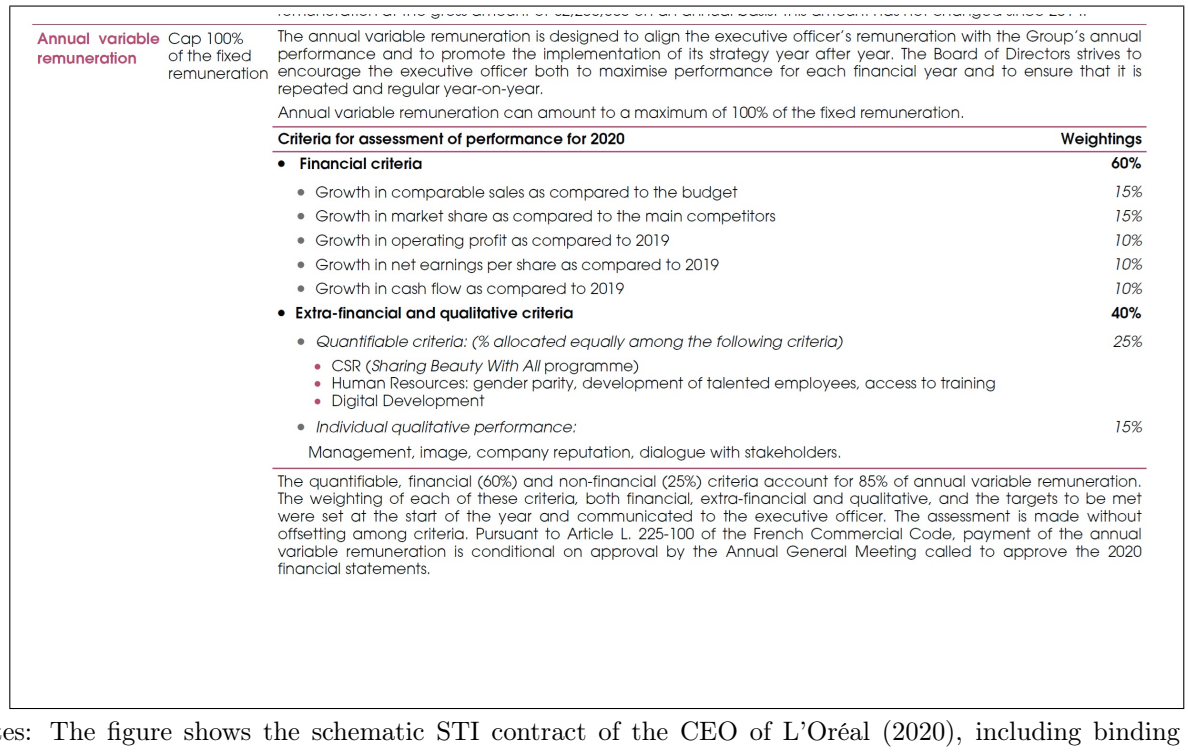
Category	Subcategory	Exemplary metrics
Social		Retailing responsibility
		Revenues from alcohol, tobacco, gambling, armaments
	Workforce	Diversity, equity, and inclusion
		Career development and training
		Employee satisfaction and turnover
		Employees with disabilities
		Female representation (top) management
		Workplace accidents
		Working conditions
		Gender pay gap
		Employee health and safety policy
Governance	CSR strategy	CSR strategy
		Stakeholder engagement
		ESG and GRI reporting scope
		UNPRI Signatory
	Management	Board independence and diversity
		Committees (Audit, compensation, nomination)
		Board size
		Succession plans
		Executive Compensation Policy
	Shareholders	Shareholder rights policy
		Shareholders Vote on Executive Pay
		Shareholder Approval Significant Transactions
		Takeover defences

C.4 Exemplary Disclosure of ESG-related Metrics from Company Reports

Here, various examples of the disclosure of ESG-related metrics in executive STI compensation contracts are highlighted, differentiating between low-disclosure and high-disclosure cases w.r.t. to the richness and salience of information on ESG metrics.

L'Oréal 2019 and 2020 annual financial reports (prototype of an STI contract with numerous ESG-related metrics, which are simultaneously also binding)

Figure C.1: Schematic CEO STI Contract at L'Oréal (2020)



Annual variable remuneration — Cap 100% of the fixed remuneration

The annual variable remuneration is designed to align the executive officer's remuneration with the Group's annual performance and to promote the implementation of its strategy year after year. The Board of Directors strives to encourage the executive officer both to maximise performance for each financial year and to ensure that it is repeated and regular year-on-year.

Annual variable remuneration can amount to a maximum of 100% of the fixed remuneration.

Criteria for assessment of performance for 2020	Weightings
• **Financial criteria**	**60%**
◦ Growth in comparable sales as compared to the budget	*15%*
◦ Growth in market share as compared to the main competitors	*15%*
◦ Growth in operating profit as compared to 2019	*10%*
◦ Growth in net earnings per share as compared to 2019	*10%*
◦ Growth in cash flow as compared to 2019	*10%*
• **Extra-financial and qualitative criteria**	**40%**
◦ *Quantifiable criteria: (% allocated equally among the following criteria)*	*25%*
▪ CSR (*Sharing Beauty With All* programme)	
▪ Human Resources: gender parity, development of talented employees, access to training	
▪ Digital Development	
◦ *Individual qualitative performance:*	*15%*
Management, image, company reputation, dialogue with stakeholders.	

The quantifiable, financial (60%) and non-financial (25%) criteria account for 85% of annual variable remuneration. The weighting of each of these criteria, both financial, extra-financial and qualitative, and the targets to be met were set at the start of the year and communicated to the executive officer. The assessment is made without offsetting among criteria. Pursuant to Article L. 225-100 of the French Commercial Code, payment of the annual variable remuneration is conditional on approval by the Annual General Meeting called to approve the 2020 financial statements.

Notes: The figure shows the schematic STI contract of the CEO of L'Oréal (2020), including binding non-ESG targets, binding ESG targets, discretionary targets. Extracted from the L'Oréal, Universal Registration Document, Annual Financial Report and Integrated Report (2019), p. 96.

Figure C.2: CEO STI Non-Financial and Qualitative Achievement Details at L'Oréal (2020)

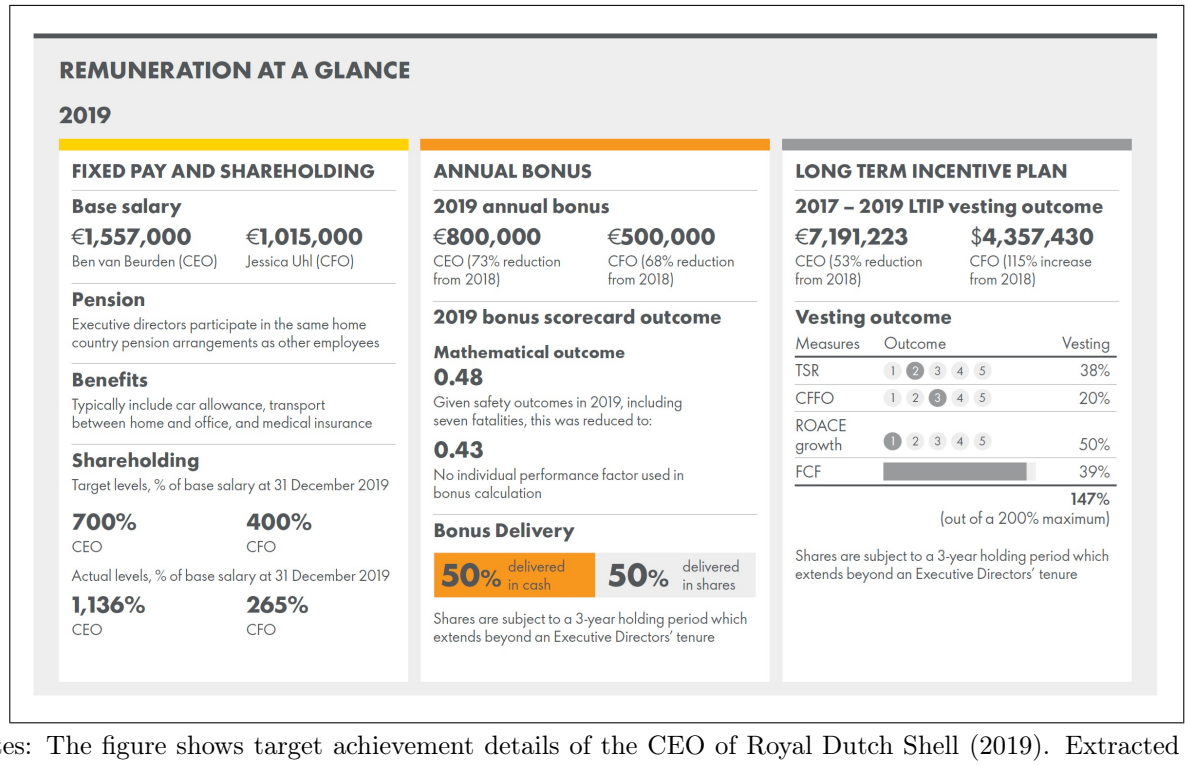
2020 NON-FINANCIAL AND QUALITATIVE TARGETS
TABLE OF NON-FINANCIAL AND QUALITATIVE ACHIEVEMENTS

CSR criteria: Sharing Beauty with All programme	2020 results	2019 results
The Sharing Beauty with All programme was launched in October 2013 by Mr Jean-Paul Agon. It structures the Group's CSR strategy and sets ambitious targets to be met by 2020. This project consists of four pillars, for which the 2020 achievements are set out in detail in Chapter 4 of this document.		
"Innovating Sustainably"		
• Improved environmental or social impact for 100% of our products.	• 96% of new products that have been screened have an improved environmental or social profile.	85%
"Producing Sustainably" (plants and distribution centres)		
• -60% on CO_2 emissions.	• -81% on CO_2 emissions (absolute value).	-78%
• -60% in water consumption.	• -33% water consumption (per finished product unit).	-51%
• -60% waste generation.	• -37% waste production (by finished product unit) The objective of 0% waste from plants and distribution centres sent to landfill was met from 2018.	-35%
"Living Sustainably"		
• Each brand will have assessed its environmental and social footprint.	• 89% of brands have assessed their impact.	89%
• Each brand will have reported on its progress and associated consumers with its commitments.	• 79% of the brands conducted a consumer awareness initiative.	57%
"Developing Sustainably"		
• With the employees (*L'Oréal Share & Care* programme): 100% of L'Oréal employees around the world will have access to healthcare coverage and social protection in 2020.	• 96% of the Group's permanent employees have access to healthcare coverage reflecting the best practices in their country of residence. 92% of the Group's permanent employees benefit from financial protection in the event of life-changing accidents, such as death or permanent disability.	94% 91%
• With strategic suppliers.	• 99% of the Group's strategic suppliers carried out a self-assessment of their Sustainable Development policy.	96%
• With communities.	• Access to work for 100,905 people.	90,635
Human Resources criteria	**2020 results**	**2019 results**
Gender Balance		
• Improving gender balance, in particular at the level of senior management positions.	• 26% of Executive Committee members are women • 49% of key positions held by women • 2020 Equileap ranking: France: No. 1 International: No. 4 • For the fourth consecutive year, L'Oréal is ranked in the Top 100 among the 325 companies of the Bloomberg Gender-Equality Index 2021.	30% 47% No. 2 in Europe Top 100
Talent Development		
• Positive policy results for the recruitment of both experienced and more junior talented employees, and talent development all over the world, in order to favour the emergence of local talent.	• 10th place in the Universum global ranking (business schools). L'Oréal is the only French and European company in the Top 10.	12th place
• Attractive, targeted, digital employer communication.	• Strong presence on social networks: 3.2 million followers on LinkedIn. • 3rd place in the PotentialPark global ranking (recruiting/social media).	2.3 million No. 2
Access to training		
• 100% of employees will receive training once a year starting in 2020.	• 100% of employees received training in 2020. Over 580,000 hours of digital training.	96% 202,000 hours

Notes: The figure shows an extract of the CEO's non-financial and qualitative (in this case ESG) targets and achievement rates of L'Oréal (2020). Extracted from the L'Oréal, Universal Registration Document, Annual Financial Report and Integrated Report (2020), p. 103.

Royal Dutch Shell 2019 annual financial report (prototype of an STI contract with ESG metrics which are all binding, no subjective component)

Figure C.3: CEO Target Achievement Details at Royal Dutch Shell (2019)



Notes: The figure shows target achievement details of the CEO of Royal Dutch Shell (2019). Extracted from the Royal Dutch Shell PLC, Annual Reports and Accounts (2019), p. 140.

Figure C.4: Schematic CEO STI Contract at Royal Dutch Shell (2019)

Notes to the single total figure of remuneration for executive directors table (audited)

Annual bonus

The Annual bonus operated in line with the policy as disclosed on page 156.

Determination of the 2019 annual bonus

The table below summarises the 2019 annual bonus scorecard measures including their weightings, targets and outcomes. The mathematical scorecard outcome for 2019 was 0.48. Please refer to pages 135-136 for a commentary on the scorecard outcome.

After reviewing the mathematical scorecard outcome, and considering the context of wider company performance for the year, the REMCO exercised discretion to adjust the scorecard result downwards to 0.43. This downwards adjustment was to reflect the seven fatalities under Shell operational control during the year.

Accordingly, the REMCO determined a final bonus outcome of €800,000 for the CEO which is 41% of target and 21% of maximum. This is a 73% reduction from 2018. The REMCO determined a final bonus outcome of €500,000 for the CFO which is 41% of target and 21% of maximum. This is a 68% reduction from 2018.

2019 annual bonus outcome (audited) [A][B]

Measures	Weight (% of scorecard)	Threshold	Target set	Outstanding	Result achieved	Score (0-2)
Cash flow from operating activities ($ billion)	**30%**	44	50	56	42	**0**
Operational excellence	**50%**					**0.72**
Production (kboe/d)	12.5%	3,647	3,760	3,873	3,665	0.16
LNG liquefaction volumes (mtpa)	12.5%	35.3	36.4	37.4	35.6	0.23
Refinery and chemical plant availability (%)	12.5%	88.4	90.4	92.4	90.8	1.20
Project delivery on schedule (%)	6.25%	60	80	100	90	1.50
Project delivery on budget (%)	6.25%	105	100	95	99	1.10
Sustainable development	**20%**					**0.59**
Total recordable case frequency (injuries/million hours)	5%	0.9	0.7	0.5	0.9	–
Operational Tier 1 and 2 process safety events (number)	5%	145	115	85	130	0.50
Upstream and Integrated Gas GHG intensity (tonnes of CO_2 equivalent/tonne of hydrocarbon production available for sale)	4%	0.176	0.168	0.160	0.168	1.00
Refining GHG intensity (tonnes CO_2 equivalent per Solomon's Utilized Equivalent Distillation Capacity (UEDC™))	4%	1.11	1.06	1.01	1.06	1.00
Chemicals GHG intensity (tonnes CO_2 equivalent/tonne of petrochemicals production)	2%	1.10	1.00	0.90	1.04	0.60
	100%					
Mathematical scorecard outcome						**0.48**
Adjusted scorecard outcome						**0.43**

[A] These metrics measure the effectiveness with which we operate our assets and portfolio base, assessed against our operational business plan. Shell's longer-term strategic ambitions are measured in the LTIP metrics.
[B] Scorecard targets are based on Shell's annual operating plan and increase or decrease year-on-year. In 2019, target refinery and chemical plant availability was lower and target GHG emission intensities higher than 2018, due to planned business activities, reflecting scheduled maintenance and expected market conditions, and portfolio developments.
[C] In external disclosure, we may use an alternative performance measure, i.e. CFFO excluding Working Capital, to describe the cash flow generation from our operations without the effect of working capital changes.

2019 bonus outcome calculation

Ben van Beurden

Target bonus:

€1,557,000 (base salary)
x 125% = **€1,946,250**

2019 scorecard result 0.43

€800,000 [A]
(51% of base salary)
↓ 73% reduction from 2018

Notes: The figure shows the schematic STI contract of the CEO of Royal Dutch Shell (2019), including binding non-ESG targets, binding ESG targets, and target achievement rates. Extracted from the Royal Dutch Shell PLC, Annual Reports and Accounts (2019), p. 142.

Richemont 2020 annual financial report (prototype of an STI contract with only a few ESG metrics which are also only discretionary, i.e., non-binding metrics)

Figure C.5: STI Executive Compensation Description at Richemont (2020)

Variable compensation of the Senior Executive Committee

Executives also benefit from a short-term cash incentive plan and awards granted under the Group's share option and PSU plans. The Committee considers these components in total to ensure there is an appropriate balance between reward for short-term success and long-term retention. Targets used to determine the payout levels for both the variable short-term incentives and the variable long-term incentives are considered by the Committee on an annual basis prior to the start of the next financial year. The Group does not provide for any transaction-specific success fees for its executives.

Short-term cash incentives

Short-term incentives are paid in cash annually and relate to performance in the previous financial year.

The determination of the level of short-term cash incentive comprises both quantitative and qualitative components, each with a pre-set target and a maximum percentage of base salary. The mix of quantitative and qualitative targets are aligned with the Group's business priorities for the year ahead, encouraging individual creativity and delivering continued profit growth and value creation. The short-term incentive target is set at 75% of base salary, with a maximum cap of 150% of base salary.

The quantitative component of the short-term cash incentive is assessed on actual Group or Maison turnover, operating profit and cash generation, being operating cash flow after capital expenditure and lease payments, compared against the current year's budget. Each of these three measures has equal weighting in the calculation. The impact of the Covid-19 outbreak on results for the year has therefore reduced the percentage achievement of quantitative objectives. The qualitative component is assessed on performance against both individual and collective strategic targets, measuring the contribution to creativity, team-building and succession-planning, among other elements.

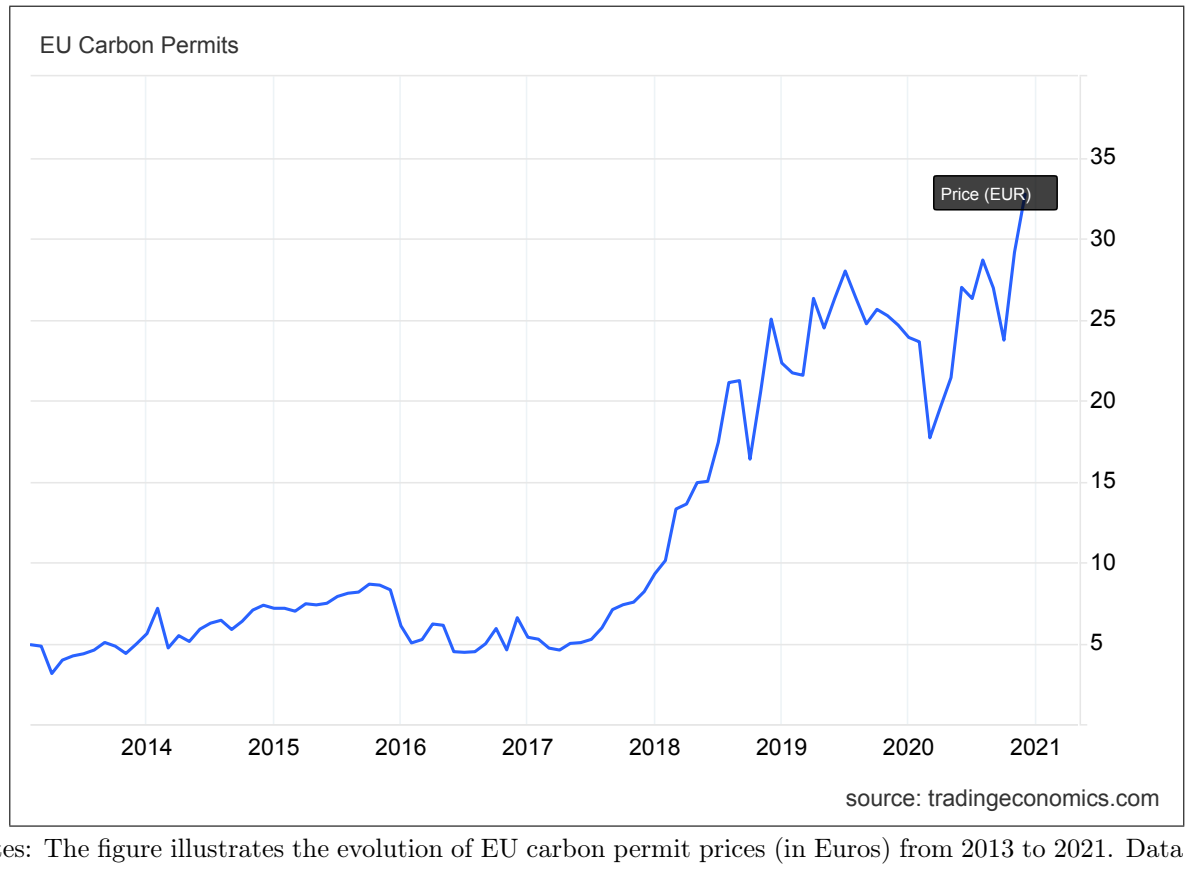
The total incentive awards achieved represented on average 56% of base salary. The individual figures for the Group's executive directors are as follows:

	Quantitative (% of salary)		Qualitative (% of salary)		Total (% of salary)	
	Target	Achieved	Target	Achieved	Target	Achieved
Nicolas Bos	*41%*	37%	*34%*	38%	*75%*	75%
Burkhart Grund	*41%*	12%	*34%*	36%	*75%*	48%
Sophie Guieysse	*41%*	12%	*34%*	35%	*75%*	47%
Jérôme Lambert	*41%*	12%	*34%*	34%	*75%*	46%
Cyrille Vigneron	*41%*	29%	*34%*	32%	*75%*	61%

Notes: The figure shows the description of the STI executive compensation at Richemont (2020), including binding non-ESG targets, discretionary targets, and target achievement rates. Extracted from Richemont, Richemont Annual Report and Accounts (2020), p. 61.

C.5 Internet Appendix

Figure IC.1: Evolution of EU Carbon Permit Prices



Notes: The figure illustrates the evolution of EU carbon permit prices (in Euros) from 2013 to 2021. Data were downloaded from https://tradingeconomics.com/commodity/carbon on July 9, 2024.

Table IC.1: Summary Statistics for Firm-level Variables

		Obs (1)	Mean (2)	S.D. (3)	P5 (4)	P10 (5)	P25 (6)	P50 (7)	P75 (8)	P90 (9)	P95 (10)
Number of executives	(#)	584	4	3	1	1	2	3	6	9	11
Number of employees	(#)	584	112,332	105,617	18,914	26,267	45,202	87,850	137,336	222,000	308,000
Market capitalization	(10^6€)	584	59,818	45,431	17,437	20,872	29,022	44,923	73,939	117,582	169,950
Total assets	(10^6€)	584	321,626	477,263	15,090	24,191	39,492	87,278	394,412	954,882	1,421,095
Net sales	(10^6€)	584	48,224	43,575	8,347	11,820	20,902	35,786	59,273	97,249	133,272
Book to market ratio	(%)	584	69.21	54.48	9.21	13.79	28.16	53.02	97.86	146.92	172.12
Book leverage	(%)	584	24.94	13.69	2.80	3.93	15.64	24.63	34.72	43.81	49.22
Net PPE/total assets	(%)	584	18.94	20.16	0.24	0.40	1.31	11.17	29.41	46.78	64.18
ROA	(%)	558	5.37	6.43	-0.22	0.33	0.94	4.21	7.79	11.97	14.83
DPS	(%)	471	52.30	21.78	10.25	22.22	39.88	52.29	68.18	79.92	86.99
Stock return	(%)	584	8.81	22.70	-31.17	-22.19	-4.68	9.26	23.35	37.04	44.12
Historical stock-to-accounting volatility		568	75.61	138.62	3.46	5.13	8.83	21.18	90.53	176.03	397.27
Log(total assets)		584	25.50	1.43	23.44	23.91	24.40	25.19	26.70	27.58	27.98
Log(book to market ratio)		580	-0.69	0.87	-2.33	-1.95	-1.26	-0.63	-0.02	0.39	0.56
Board independence	(%)	584	71.75	19.84	38.46	45.45	58.33	75.00	87.50	100.00	100.00
Female board membership	(%)	584	31.23	10.25	14.29	17.65	25.00	30.77	38.46	42.86	45.45
Institutional ownership	(%)	584	50.26	15.85	24.79	31.57	38.93	50.88	59.53	70.82	76.48
Block ownership	(%)	584	7.22	16.04	0.00	0.00	0.00	0.00	10.14	26.52	46.57
Emission policy Y/N	(1/0)	584	1.00	0.06	1	1	1	1	1	1	1
Total CO_2 emission equivalents	(tons/10^3€)	580	0.18	0.63	0.00	0.00	0.01	0.02	0.05	0.46	1.01
Historical avg. total CO_2 emission equivalents	(tons/10^3€)	576	0.16	0.39	0.00	0.00	0.01	0.02	0.06	0.53	1.11
Employee turnover	(%)	499	12.12	5.67	3.50	4.80	8.30	11.73	15.30	19.60	22.80
ESG Reporting Scope		584	88.99	20.65	40.74	58.85	87.00	100.00	100.00	100.00	100.00
ESG rating		584	4.97	1.60	2.00	2.76	3.79	5.15	6.04	7.00	7.00
High-reach bad ESG news		584	1.19	1.10	0.00	0.00	0.00	1.10	1.95	2.77	3.30
High-reach bad E news		584	0.30	0.59	0.00	0.00	0.00	0.00	0.69	1.10	1.39
High-reach bad S news		584	0.47	0.65	0.00	0.00	0.00	0.00	0.69	1.39	1.95
High-reach bad G news		584	0.94	1.08	0.00	0.00	0.00	0.69	1.61	2.48	3.26
Bad ESG news		584	2.70	1.29	0.00	1.10	1.79	2.71	3.71	4.41	4.72
Bad ES news		584	2.00	1.24	0.00	0.00	1.10	1.95	2.94	3.69	4.13
Bad G news		584	2.17	1.30	0.00	0.00	1.10	2.20	3.09	3.87	4.42

Notes: Reported are summary statistics on firm characteristics (at the firm level) in a given year. For variable definitions, see Appendix C.1.

Table IC.2: STI and LTI Performance Metrics by ESG Category

	Obs (1)	Mean (2)	S.D. (3)	P5 (4)	P10 (5)	P25 (6)	P50 (7)	P75 (8)	P90 (9)	P95 (10)
Panel A: STI										
Number of ESG metrics:	2,500	1.859	2.797	0	0	0	1	3	5	6
$\#(Environmental)_{i,t}$	2,500	0.147	0.613	0	0	0	0	0	0	1
$\#(Social)_{i,t}$	2,500	1.326	2.012	0	0	0	1	2	4	5
$\#(Governance)_{i,t}$	2,500	0.384	0.784	0	0	0	0	1	1	2
Number of non-ESG metrics:	2,490	7.730	6.619	1	2	3	6	10	16	22
$\#(Financial)_{i,t}$	2,490	3.693	2.897	1	1	2	3	5	7	8
$\#(Non\text{-}financial)_{i,t}$	2,490	4.037	4.795	0	0	1	2	5	10	15
Panel B: LTI										
Number of ESG metrics:	2,405	0.416	0.980	0	0	0	0	0	2	2
$\#(Environmental)_{i,t}$	2,405	0.061	0.368	0	0	0	0	0	0	0
$\#(Social)_{i,t}$	2,405	0.272	0.735	0	0	0	0	0	1	2
$\#(Governance)_{i,t}$	2,405	0.072	0.303	0	0	0	0	0	0	1
Number of non-ESG metrics:	2,334	3.222	2.347	1	1	2	3	4	6	8
$\#(Financial)_{i,t}$	2,334	2.536	1.281	1	1	2	2	3	4	5
$\#(Non\text{-}financial)_{i,t}$	2,334	0.686	1.720	0	0	0	0	0	2	4

Notes: Reported are summary statistics for the number of ESG and non-ESG performance metrics by different categories. In the case of ESG metrics, we show the number of metrics for each of the three categories Environmental, Social, and Governance. In the case of non-ESG metrics, we distinguish between financial and non-financial metrics. We do not distinguish between binding and discretionary metrics in this table. Panel A focuses on short-term incentive contracts (STI), whereas Panel B reports numbers for long-term incentive contracts (LTI).

Table IC.3: ESG Reporting Disclosure

	(1)
Dependent variable:	ESG Reporting Scope
Ln(Number of binding ESG metrics)	-1.796
	(1.832)
Ln(Number of binding non-ESG metrics)	0.605
	(2.685)
Ln(Number of discretionary ESG metrics)	-2.556
	(2.813)
Ln(Number of discretionary non-ESG metrics)	1.443
	(1.716)
Year 2014	-1.256
	(0.982)
Year 2015	-0.681
	(1.497)
Year 2016	-3.009
	(2.491)
Year 2017	-2.930
	(2.889)
Year 2018	-2.508
	(3.309)
Year 2019	-1.932
	(3.740)
Year 2020	-3.065
	(3.588)
Firm F.E.	Yes
Observations	569
Adjusted R^2	0.57

Notes: This table reports a panel regression at the firm-year level. The dependent variable is ESG Reporting Scope, defined as the percentage of a company's activities covered in its environmental and social reporting. The main regressors are the logarithm of one plus the number of binding and discretionary ESG and non-ESG performance metrics in executives' STI contracts. Year fixed effects (base year: 2013) and firm fixed effects are included throughout. Robust standard errors are clustered by firm and reported in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table IC.4: Decomposition of Within-executive STI Variance by Industry

			Binding metrics				Discretionary metrics		Covariance terms (2×cov)		
			$RSTI_{i,t}^{B,ESG}$		$RSTI_{i,t}^{B,nESG}$		$RSTI_{i,t}^{D}$		$(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{B,nESG})$	$(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{D})$	$(RSTI_{i,t}^{B,nESG}, RSTI_{i,t}^{D})$
	obs	$Var(RSTI_{i,t})$ $\times 10^{-6}$	STI weight	Var share	STI weight	Var share	STI weight	Var share	Var share	Var share	Var share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Consumers	283	229,451	0.001	0.000	0.625	1.242	0.374	0.425	0.002	−0.001	−0.668
Energy & Utilities	68	808,789	0.043	0.048	0.828	1.374	0.130	0.636	0.196	0.006	−1.261
Financials	293	449,736	0.005	0.007	0.608	1.479	0.387	1.061	0.012	−0.031	−1.528
Industrials & Materials	292	102,235	0.057	0.036	0.733	0.995	0.210	0.393	0.032	−0.095	−0.361
Health	140	103,511	0.000	0.000	0.732	1.589	0.268	1.863	0.000	0.000	−2.451
ICT	175	335,513	0.067	0.103	0.804	1.963	0.128	2.441	−0.054	−0.039	−3.415

Notes: This table shows the decomposition of the variance of executives' realized short-term incentive compensation $RSTI_{i,t}$ for different industry sectors. The analysis focuses on within-executive variation in pay. To this end, we first compute the residuals of STI parts $RSTI_{i,t}^{B,ESG}$, $RSTI_{i,t}^{B,nESG}$, and $RSTI_{i,t}^{D}$ that are not explained by executive fixed effects. We sum up these residuals, free of cross-sectional variation, to total $RSTI_{i,t}$. Column 2 reports the variance $Var\left(RSTI_{i,t}\right)$ of the remaining within-executive variation in total STI. The variance shares reported in columns 4, 6, 8, and 9 to 11 correspond to the variance and covariance terms in Equation (4.6) divided by total STI variance. For example, the variance share of binding ESG metrics in column 4 is defined as $Var(RSTI_{i,t}^{B,ESG})/Var(RSTI_{i,t})$. Columns 3, 5, and 7 report the ex ante weights that firms assign to the different STI performance metrics, and are calculated as the sample averages of $w_{i,t}^{B,ESG}$, $w_{i,t}^{B,nESG}$, and $w_{i,t}^{D}$. As in previous tables, superscripts B and D designate binding and discretionary performance metrics, and superscripts ESG and $nESG$ designate ESG and non-ESG metrics. For more information on the variance decomposition approach, see Section 4.5.1.3.

Table IC.5: Decomposition of Within-executive STI Variance by Executive Position

			Binding metrics				Discretionary metrics		Covariance terms (2×cov)		
			$RSTI_{i,t}^{B,ESG}$		$RSTI_{i,t}^{B,nESG}$		$RSTI_{i,t}^{D}$		$(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{B,nESG})$	$(RSTI_{i,t}^{B,ESG}, RSTI_{i,t}^{D})$	$(RSTI_{i,t}^{B,nESG}, RSTI_{i,t}^{D})$
	obs	$Var(RSTI_{i,t})$ $\times 10^{-6}$	STI weight	Var share	STI weight	Var share	STI weight	Var share	Var share	Var share	Var share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
CEO	327	227,838	0.036	0.020	0.725	1.292	0.239	0.651	0.013	−0.032	−0.945
CFO	219	77,472	0.033	0.019	0.748	1.296	0.219	0.818	0.046	−0.035	−1.143
CHRO	81	49,668	0.015	0.002	0.710	0.921	0.276	0.316	0.013	−0.004	−0.248
COO	57	106,292	0.007	0.007	0.713	1.091	0.280	0.364	0.002	−0.005	−0.459
Division/Region Head	392	65,499	0.022	0.005	0.670	1.230	0.308	0.874	0.010	−0.015	−1.104
Other specialist	163	36,514	0.020	0.006	0.586	1.840	0.393	1.621	0.004	−0.011	−2.460

Notes: This table shows the decomposition of the variance of executives' realized short-term incentive compensation $RSTI_{i,t}$ for different executive positions. The analysis focuses on within-executive variation in pay. To this end, we first compute the residuals of STI parts $RSTI_{i,t}^{B,ESG}$, $RSTI_{i,t}^{B,nESG}$, and $RSTI_{i,t}^{D}$ that are not explained by executive fixed effects. We sum up these residuals, free of cross-sectional variation, to total $RSTI_{i,t}$. Column 2 reports the variance $Var\left(RSTI_{i,t}\right)$ of the remaining within-executive variation in total STI. The variance shares reported in columns 4, 6, 8, and 9 to 11 correspond to the variance and covariance terms in Equation (4.6) divided by total STI variance. For example, the variance share of binding ESG metrics in column 4 is defined as $Var(RSTI_{i,t}^{B,ESG})/Var(RSTI_{i,t})$. Columns 3, 5, and 7 report the ex ante weights that firms assign to the different STI performance metrics, and are calculated as the sample averages of $w_{i,t}^{B,ESG}$, $w_{i,t}^{B,nESG}$, and $w_{i,t}^{D}$. As in previous tables, superscripts B and D designate binding and discretionary performance metrics, and superscripts ESG and $nESG$ designate ESG and non-ESG metrics. For more information on the variance decomposition approach, see Section 4.5.1.3.

Table IC.6: ESG Pay and Firm and Executive Characteristics

	Discretionary ESG metrics (ESG_D)				Binding ESG metrics (ESG_B)			
	At least 1 ESG_D (1)	Log(1+No. ESG_D) (2)	No. ESG_D / No. all metrics (3)	Weight All D (4)	At least 1 ESG_B (5)	Log(1+No. ESG_B) (6)	No. ESG_B / No. all metrics (7)	Weight ESG_B (8)
Historical avg. total CO_2 emission equivalents	0.131	0.356**	0.050**	-2.590	-0.227**	-0.262	-0.066	-2.674
	(0.106)	(0.146)	(0.022)	(6.291)	(0.106)	(0.162)	(0.040)	(2.128)
Historical stock-to-accounting volatility ($\times 10^{-2}$)	0.028	-0.023	0.008	1.231	0.037***	0.045***	0.021***	0.781**
	(0.032)	(0.049)	(0.006)	(1.839)	(0.012)	(0.015)	(0.007)	(0.378)
Log(total assets)	0.056	0.109	0.024	8.852**	-0.064	-0.053	-0.016	-1.106
	(0.054)	(0.085)	(0.021)	(3.993)	(0.046)	(0.061)	(0.017)	(0.927)
Log(book to market ratio)	0.132**	0.239**	0.053**	6.378	0.047	0.037	0.009	0.862
	(0.061)	(0.092)	(0.025)	(4.885)	(0.050)	(0.075)	(0.022)	(1.173)
ROA	0.005	0.014	0.001	-0.044	-0.001	-0.001	-0.000	-0.039
	(0.006)	(0.009)	(0.002)	(0.420)	(0.003)	(0.004)	(0.001)	(0.067)
Stock return	0.002**	0.002*	0.001**	0.146*	0.001	0.000	0.000	0.018
	(0.001)	(0.001)	(0.000)	(0.082)	(0.001)	(0.001)	(0.000)	(0.013)
Book leverage	0.003	0.004	0.001	-0.057	0.004	0.004	0.001	0.049
	(0.004)	(0.006)	(0.001)	(0.251)	(0.003)	(0.005)	(0.001)	(0.057)
Net PPE/total assets	0.003	0.003	0.001	0.132	0.000	0.003	0.001	0.054
	(0.003)	(0.005)	(0.001)	(0.136)	(0.002)	(0.003)	(0.001)	(0.040)
DPS	-0.004	0.011	-0.002	0.538	0.048**	0.067*	0.022**	0.989*
	(0.034)	(0.045)	(0.010)	(2.276)	(0.022)	(0.036)	(0.010)	(0.525)
Emissions policy	0.724***	0.896***	0.142***	27.159***	0.223***	0.178*	0.055**	3.236**
	(0.229)	(0.274)	(0.049)	(9.117)	(0.062)	(0.093)	(0.026)	(1.349)
ESG rating	0.022	0.046	-0.000	-1.424	0.006	0.025	0.004	0.402
	(0.027)	(0.038)	(0.008)	(1.794)	(0.016)	(0.021)	(0.005)	(0.291)
Institutional ownership	0.006**	0.009***	0.001**	0.185	-0.006***	-0.004**	-0.001**	-0.052*
	(0.003)	(0.003)	(0.001)	(0.184)	(0.002)	(0.002)	(0.000)	(0.027)
Block ownership	-0.007**	-0.008	-0.002	-0.435	0.003*	0.001	0.000	0.014
	(0.003)	(0.006)	(0.002)	(0.292)	(0.002)	(0.002)	(0.001)	(0.035)
Board independence	0.003	0.006**	0.001	0.171	-0.002	-0.003	-0.001	-0.043
	(0.002)	(0.003)	(0.001)	(0.112)	(0.001)	(0.002)	(0.001)	(0.034)
Female board membership	-0.003	-0.000	0.000	-0.142	0.001	0.003	0.001	0.034
	(0.005)	(0.006)	(0.001)	(0.358)	(0.003)	(0.003)	(0.001)	(0.046)
Tenure	-0.001	-0.000	0.001	0.053	-0.002	-0.000	0.000	-0.018
	(0.005)	(0.008)	(0.002)	(0.403)	(0.003)	(0.003)	(0.001)	(0.055)
Age	0.004	0.003	-0.000	0.095	0.001	0.002	0.001	0.083**
	(0.003)	(0.004)	(0.001)	(0.222)	(0.002)	(0.003)	(0.001)	(0.039)
Female	-0.074**	-0.069	-0.006	-2.572	-0.014	-0.013	0.003	0.134
	(0.031)	(0.043)	(0.011)	(1.921)	(0.020)	(0.031)	(0.009)	(0.499)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Executive position F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,091	2,091	2,090	2,090	2,091	2,091	2,090	2,090
Adjusted R^2	0.21	0.27	0.23	0.16	0.19	0.17	0.18	0.16

Notes: We report panel regressions for various measures of ESG adoption in executive STI as function of firm and executive characteristics as well as fixed effects for industry sectors, executive positions, and years. The dependent variables are defined as follows. Column 1: At least 1 $ESG_D = \mathbb{1}(\#_{i,t}^{D,ESG} \geq 1)$; column 2: Log(1+No. ESG_D) $= Ln(1 + \#_{i,t}^{D,ESG})$; column 3: No. ESG_D/No. all metrics $= \#_{i,t}^{D,ESG}/(\#_{i,t}^{D} + \#_{i,t}^{B})$; column 4: Weight All $D = w_{i,t}^{D}$; column 5: At least 1 $ESG_B = \mathbb{1}(\#_{i,t}^{B,ESG} \geq 1)$; column 6: Log(1+No. ESG_B) $= Ln(1 + \#_{i,t}^{B,ESG})$; column 7: No. ESG_B/No. all metrics $= \#_{i,t}^{B,ESG}/(\#_{i,t}^{D} + \#_{i,t}^{B})$; column 8: Weight $ESG_B = w_{i,t}^{B,ESG}$. The independent variables are defined in Appendix C.1. Robust standard errors are clustered by firm and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Chapter 5

Communicating through Crisis - Nonanswers in Earnings Calls during the Covid Pandemic[§]

This study examines changes in management communication during the Covid pandemic, with a focus on nonanswers in question and answer sessions of conference calls and the capital market reactions that accompany them. The results indicate that greater crisis exposure of companies, as revealed in their management presentations, is associated with a higher frequency of nonanswers from management. Furthermore, nonanswers continue to be associated with statistically and economically significant negative market responses, also under crisis conditions. However, especially when the management presentation reveals a comparatively constructive or opportunity-oriented view of the company's situation with respect to the pandemic, nonanswers tend to be associated with even stronger market penalties, potentially indicating elevated expectations for transparency and clear communication. The findings reveal notable changes in corporate communication and emphasize its crucial role in maintaining investor confidence during periods of heightened uncertainty.

Key Words: Earnings Conference Calls; Conference Calls; Nonanswers; Textual Analysis; Crisis; Corporate Communication

JEL Classification: D83, G14, M41

[§]This chapter is based upon an unpublished manuscript by Ehmann (2026).

5.1 Introduction

"*We don't know what other measures from the politics will come. So customer reaction, the lengths and duration of the coronavirus situation is difficult to predict.*"
Frank Witter, CFO at Volkswagen AG, in the company's conference call held on 17 March, 2020

In early 2020, the world was struck by the Covid pandemic. Across the globe, the crisis affected social and economic structures alike, disrupting nearly every aspect of life. As a result, instability and uncertainty rose sharply, a trend also captured by the World Uncertainty Index. This index, which is designed to measure uncertainty in the global economy, reached unprecedented levels at the onset of the crisis in early 2020 (Ahir et al. (2022)). Amid these global challenges, also corporate managements were forced to find ways to navigate through and cope with change and uncertainty. Conference calls, sometimes also called earnings calls or earnings conference calls, are a crucial setting for timely and effective management reactions to the crisis. Conference calls are regularly scheduled meetings where management provides updates and discusses company affairs. During these calls, management provides information on past activities and the company's outlook, while participants, such as analysts and investors, are given the opportunity to ask questions. Thus, conference calls create a space for dialogue between executives and the market and are intended as a means to decrease existing information asymmetry (Hollander et al. (2010); Matsumoto et al. (2011)). Scholars increasingly examine how management navigates interactions with conference call participants, such as investors and analysts, and how markets respond. Various features of conference calls have been the subject of research, ranging from sentiment conveyed at a general level (Price et al. (2012)) to rather specific language characteristics, as for instance the use of euphemisms (Suslava (2021)) or humor (Call et al. (2024)) in conference calls.

This paper examines a distinctive response pattern exhibited by management during conference calls, specifically nonanswers (Hollander et al. (2010); Gow et al. (2021); Barth et al. (2023); De Kok (2025)). Nonanswers are management answers in which the question posed is not answered, for instance, by blathering, avoiding the answer, or admitting (or pretending) to not knowing. Nonanswers may be of particular importance during times of crisis. Such periods are often accompanied by heightened uncertainty and, in some cases, even fear among investors (Kale and Kale (2023)). At the same time, reliable and immediate knowledge about current or future developments is inherently limited, making clear and reassuring statements more difficult to provide. This paper analyzes nonanswers within the crisis context in greater detail. Specifically, it investigates whether the frequency of nonanswers changes with various levels of exposure to the Covid pandemic. Additionally, subsequent market reactions are explored. Typically, nonanswers are viewed negatively by the market, as they can suggest or be interpreted as withholding information, and are therefore met with adverse market reactions.

However, given the exceptional nature of the Covid crisis, it is not clear whether market responses to nonanswers differ under crisis conditions and thus, inherent unpredictability.

Using a sample of 16,978 conference calls between 2016 and 2023 from both the US and Europe, I find that greater crisis exposure, as revealed by the management, is associated with higher levels of nonanswers from management. However, there is no consistent evidence that this relationship is particularly driven by any specific Covid-related sentiment, such as a comparatively negative, optimistic, or uncertain outlook of the firm on the Covid crisis. Additionally, I find that they remain statistically and economically negatively associated with market responses, even in the context of (average) crisis exposure. These findings indicate that also when facing significant uncertainty and when situational knowledge is inherently limited, the market does not fully discount nonanswers as merely a reflection of the broader context. By contrast, one distinctive pattern emerges: Especially when a company is comparatively positively exposed to the crisis, i.e., when the presentation statement of the management reflects a rather constructive or opportunity-oriented view of the company's situation than average, nonanswers are associated with comparatively greater market penalties. This suggests that, particularly under these conditions, the market may be more inclined to demand greater transparency and less willing to accept ambiguity or evasive communication. The findings point to the possibility that firms with a more constructive or opportunity-oriented view of the crisis face increased market pressure to communicate clearly.

This research offers valuable insights for the literature. First of all, this paper sheds new light on nonanswer patterns in corporate conference calls, particularly considering differences in crisis exposure. Prior literature already uncovered and analyzed the frequency and characteristics of nonanswers at a general level (Hollander et al. (2010); Gow et al. (2021); Barth et al. (2023)) and during the time frame of a particular crisis, the Covid pandemic (De Kok (2025)). This paper extends existing research by shifting from a generalized view of crisis periods to a more nuanced analysis that considers variation in crisis exposure and sentiment conveyed, in order to examine how nonanswer patterns relate to differing crisis contexts. Specifically, it investigates nonanswer behavior across distinct crisis settings, contributing to the broader literature on corporate communication under uncertainty (Crawford Camiciottoli (2018); Suslava (2021); Falconieri and De Amicis (2023); Howe et al. (2024)). Thereby, the analysis centers on the Covid pandemic as a particularly severe and globally relevant crisis.

Secondly, this paper analyzes changes in market reactions in response to crisis periods and shifts in communication patterns. Previous studies examined market responses during periods of crisis and heightened uncertainty (see, for instance, Altig et al. (2020); Szczygielski et al. (2022); Kale and Kale (2023)), as well as reactions to various elements and changes in confer-

ence call communication throughout crises (Wang and Xing (2020); Falconieri and De Amicis (2023); Hassan et al. (2023); Fang et al. (2024); Howe et al. (2024); Jory et al. (2025)). This paper adds to the literature by specifically focusing on how perceptions of the construct of nonanswers evolve during the Covid pandemic. Thus, the paper yields important insights into corporate communication strategies in times of great uncertainty through the lens of conference calls. This focus is particularly relevant as effective communication contributes to successful crisis management. The Covid pandemic provides an ideal context for this analysis, given its unprecedented global scope and the magnitude of its economic disruption, as reflected in a historically significant 2.7% decline in the world's GDP in 2020 (Statista (2025)). This is a particularly noteworthy exception for the GDP growth rate, which usually fluctuates around positive two to five percent growth per year, aside from the other significant exception of the global financial crisis, when growth declined, even though only by 0.3%. Additionally, the pandemic also introduced pervasive turbulence across non-economic domains, as for instance indicated by the specifically developed stringency index that captures differences and fluctuations in the overall strictness of government responses to Covid, based on (lockdown) measures like school closures or movement restrictions (Hale et al. (2021)). Furthermore, the Covid pandemic generated a high degree of systemic volatility, fostering an environment which was characterized by pronounced informational ambiguity. Within this context, the strategic use and interpretation of nonanswers become especially salient, presenting a unique opportunity to investigate corporate communication practices under conditions of severe crisis.

The rest of the paper proceeds as follows. Section 5.2 reviews the literature and develops the hypotheses. In Section 5.3, the data is described. Section 5.4 outlines the statistical analysis. Section 5.5 presents the results and Section 5.6 concludes.

5.2 Literature Review & Hypotheses Development

Conference calls are widely acknowledged as a significant source of information for the capital market (Bowen et al. (2002); Matsumoto et al. (2011); Mayew et al. (2013)). They represent a voluntary form of disclosure by the management and are usually held shortly after the release of financial results, typically at the end of each financial quarter. Conference calls generally start with a carefully prepared management presentation delivered by the company's top executives. The management presentation often includes a performance summary, an outlook on future developments, and other important current topics. It is usually rather formal and preplanned (Matsumoto et al. (2011); Crawford Camiciottoli (2018)). As such, conference calls provide management with an opportunity to present quarterly results and outline future planning. Afterwards, a more interactive question and answer session follows. Here, other call participants are also involved, mainly financial analysts and investors. These participants get

to ask questions which management then addresses. Due to its interactive nature, this part is sometimes also called discussion session. Thus, it becomes apparent that the scripted presentation by management and the spontaneous question and answer session differ significantly in both structure and content. However, both parts of a conference call have been found to contain additional information beyond that incorporated in the written official press release of the company (Matsumoto et al. (2011)). Findings from prior research indicate that especially the second interactive discussion between management and other call participants is of particular informational value, suggesting that the participation and active involvement of (informed) external parties, such as analysts and investors, is particularly beneficial with regard to informational relevance (Matsumoto et al. (2011); Mayew et al. (2020)).

These aspects highlight that conference calls have the potential to contribute to reducing the information asymmetry between management of a company and external parties. Not only are conference calls particularly insightful due to the information directly revealed, but their unique real-time and bilateral disclosure format also allows external parties to derive additional information from verbal and non-verbal cues (Matsumoto et al. (2011); Gow et al. (2021); Call et al. (2024)). A multitude of verbal and non-verbal cues have been investigated in prior research. This includes, among others, the use and impact of the general sentiment or tone (Price et al. (2012)), emotional signals (Mayew and Venkatachalam (2012)), scripted answers (Lee (2016)), linguistic complexity (Bushee et al. (2018)), euphemisms (Suslava (2021)), blame behavior (Noh and Zhou (2022)), humor (Call et al. (2024)), and vocal quality (Baik et al. (2024)) in conference calls.

The literature stream at the center of this paper analyzes management nonanswers in conference calls (Hollander et al. (2010); Gow et al. (2021); Barth et al. (2023); De Kok (2025)). Nonanswers occur when managers do not supply the requested information in response to questions posed in the question and answer session of a conference call (Barth et al. (2023)). Among others, this includes managers refusing to answer, managers being unable to answer (Hollander et al. (2010); Gow et al. (2021); Barth et al. (2023)), managers blathering (Barth et al. (2023)) and managers deferring inquiries until after the call (Gow et al. (2021)). Nonanswers are often evaluated as a sign of nondisclosure.

Typically, nonanswers are highly context-dependent and at the same time, they may also be rather subtle, which makes their identification inherently challenging (De Kok (2025)). There are multiple possibilities to measure and quantify nonanswers in large text corpora. Hollander et al. (2010) manually code nonanswers in 681 conference calls held by publicly listed firms in the US between January 1, 2004 and December 31, 2004. Then, they separate calls with nonanswers, i.e., calls in which information was withheld, and calls without nonanswers. Gow

et al. (2021) allowed the analysis of larger datasets by developing a set of regular expressions indicative of nonanswers.[63] If there is a nonanswer regular expression in a response of an executive, the whole answer is classified as a nonanswer. Gow et al. (2021) then calculate a nonanswer rate by relating the amount of nonanswers to the total amount of responses in the question and answer part of a conference call transcript. They study conference calls from 2003 to 2016, also from firms incorporated in the US. Another approach brought forward by Barth et al. (2023) makes use of a supervised machine learning framework to identify 1,364 trigrams characteristic of nonanswers. A trigram is a sequence of three consecutive words, commonly used in text analysis to capture local linguistic patterns. They use the trigrams and their respective weightings (loadings) to calculate a so-called nonanswer score. The score relates the occurring nonanswer trigrams in management answers to the amount of total words spoken by the management. Thus, they measure nonanswers on a continuous spectrum. Barth et al. (2023) study nonanswers in (non-)financial companies listed in the S&P500 from 2002 to 2019. Most recently, De Kok (2025) uses generative large language models to classify nonanswers with respect to different dimensions, as for instance, refusal to answer, provisions of alternative qualitative answer, and others. The study analyzes 63,959 conference calls from 2013 to 2022.

Prior evidence suggests that nonanswers are rather prevalent. Hollander et al. (2010) indicate that about 65% of conference calls contain at least one question which remains unanswered. Gow et al. (2021) show that about 11% of questions result in nonanswers by the management, a rate which they find to be rather stable over time. These results largely match more recent findings by De Kok (2025).

Moreover, nonanswers can affect the capital market as omitted responses, whether intentional or not, can implicitly signal information to which the market reacts (Hollander et al. (2010); Barth et al. (2023); Liu et al. (2024); Bamber and Nappert (2025)). In line with the interpretation of nonanswers as a sign of nondisclosure, empirical evidence by Hollander et al. (2010) and, Barth et al. (2023) suggests that the capital market generally perceives nonanswers as a negative signal concerning future company performance. For example, Barth et al. (2023) show that higher (non-) weighted nonanswer scores are associated with significantly lower cumulative abnormal stock returns and additionally also with higher implied volatility of the stock. Although markets also react to other related verbal and non-verbal cues by management in conference calls (see, for instance, Price et al. (2012)), Barth et al. (2023) point out that their developed nonanswers score is a distinct and new dimension of information.

[63] Gow et al. (2021) make their list of regular expressions available at https://github.com/iangow/non_answer_perf/blob/main/non_answer_perf.ipynb.

While the approaches outlined above all aim to depict the same construct of nonanswers, there are slight differences in focus, and there is no unambiguous universal definition for categorization (Barth et al. (2023)). Boundaries between other existing concepts in the literature are fluent. Related concepts focus on scripted language (Lee (2016)), managerial evasiveness expressed, for example, by topic shifting in managerial answers (Dai et al. (2022); Zhou and Huang (2023)), or providing irrelevant answers (Yang et al. (2024)), expected answers (Bai et al. (2023)), dodging questions (Barcellos (2025)), or overall unusualness in financial communication (Beckmann et al. (2024)).

In times of crisis, high uncertainty, and fast changing dynamics, information frictions on markets are particularly intense (Maslar et al. (2021); Kale and Kale (2023)). Under these circumstances, conference calls as interactive real-time events with insights beyond historical financial information can be especially valuable to investors (Falconieri and De Amicis (2023)). There are several studies investigating conference calls during crisis periods, as for instance, health crises including the recent Covid pandemic, or the SARS crisis in the early 2000s as well as economic crises as, for instance, the global financial crisis starting in 2007 (Crawford Camiciottoli (2018); Wang and Xing (2020); Suslava (2021); Falconieri and De Amicis (2023); Fang et al. (2024); Hassan et al. (2023); Kale and Kale (2023); Howe et al. (2024)).

Firstly, prior empirical evidence indicates that the style of communication in conference calls changes in times of crisis. This is, for example, reflected in shifts in the tone or sentiment of a conference call. Falconieri and De Amicis (2023) investigate communication in conference calls held by US listed firms around the global financial crisis (2006-2013). Their findings suggest that communication in both the management presentation and the question and answer part tends to be less positive during a crisis. This result is also supported by Wang and Xing (2020) who investigate conference calls during the start of the Covid pandemic (2020Q1) in the US and analyze a more fine-grained text-based Covid exposure metric. More specifically, Falconieri and De Amicis (2023) outline that the change is especially pronounced for the question and answer sessions. They attribute this finding to the strain placed on executives, which becomes particularly evident during interactions with analysts. Moreover, there is also evidence suggesting that the level of communicated uncertainty in both the management presentation and/or the question and answer session rises with the crisis (exposure) (Wang and Xing (2020); Falconieri and De Amicis (2023)). Furthermore, Falconieri and De Amicis (2023) point out that in a crisis, the presentation statements tend to get longer while the answers provided in the question and answer sessions are comparatively shorter, indicating more scripted information and less spontaneous interactions.

There are also other noticeable differences in communication for crisis periods. For instance, Suslava (2021) highlights a considerable increase in the use of euphemisms in economic downturns, as, for instance, during the financial crisis. Her results are based on data from US firms between 2002 to 2016.

In line with the focus of this paper, times of crisis could also affect the amount of nonanswers given in conference calls. In a crisis situation, an abnormally high level of uncertainty prevails. This also particularly applies to the Covid pandemic, during the early stage of which the World Uncertainty Index rose to a previously unmatched high (Ahir et al. (2022)). The pandemic, characterized by unstable economic conditions and frequently changing regulations, makes clear communication increasingly difficult, even with the best intentions. This could also be reflected in a comparatively higher level of nonanswers provided by management during conference calls. The notion can be supported by prior literature as, for instance, Barth et al. (2023, p. 6339) highlight that there is "*a peak of nonanswers during and in the immediate aftermath of the financial crisis*". In line, the descriptive findings by De Kok (2025) show an increase in nonanswers also for the time period of the Covid crisis, especially in relation to forward-looking questions. Additionally, nonanswers are particularly likely when questions carry great uncertainty (Gow et al. (2021)) and negativity (Barth et al. (2023)), both of which are potentially amplified with crisis exposure.

Thus, the arguments put forward suggest a generally positive relationship between nonanswers and the Covid crisis. However, the strength of the relationship could differ according to the prevalent perceptions regarding the pandemic within each company. In line with the literature, the relationship is likely to be especially strong for companies that are hit by the pandemic in a particularly negative way or for companies who find themselves in an exceptionally uncertain environment due to the crisis.

Moreover, an exceptional event like the Covid crisis may alter the communication dynamics surrounding nonanswers. Prior literature (Hollander et al. (2010); Gow et al. (2021); Barth et al. (2023); De Kok (2025)) primarily focuses on external analysts' and investors' role, namely, the questions they ask, to identify patterns of nonanswers. As such, one could argue that also in the context of the Covid crisis, the Covid exposure at the question level, which reflects analysts' perceptions of the pandemic's impact on the company, could provide informative insights into nonanswers. On the other hand, during the pandemic, the information revealed in the presentation statement may be particularly relevant, also with respect to nonanswers. As noted by Matsumoto et al. (2011), there are generally two reasons for managers to initially omit relevant information that later becomes the focus of questions. Management might either deliberately withhold information to obscure certain details, particularly bad news, or they may be un-

aware of its significance for the capital market (Matsumoto et al. (2011)). In the context of a global crisis, disclosing pandemic-related information is unlikely to surprise the capital market much. Instead, it might generally be anticipated. As a result, the incentive to withhold such information may be diminished, leading to greater transparency from management early on. Accordingly, Covid exposure disclosed in the presentation statement will also be examined, as it is likely to be especially relevant during the crisis period.

These considerations lead to the following hypothesis:

Hypothesis 1: There is a positive relationship between management nonanswers and a company's exposure to the Covid pandemic, as indicated by both the management presentation and the content of analysts' questions. This relationship is particularly strong when exposure to the pandemic is especially severe or accompanied by substantial uncertainty.

Secondly, times of crisis can not only alter corporate communication itself but also how markets respond to the communication. Results from Falconieri and De Amicis (2023) indicate that conference calls become even more valuable to the capital market during crises, likely because they help reducing the typically high levels of information asymmetry in such situations. Findings by Wang and Xing (2020) point in the same direction as they find a positive relationship between Covid references in the presentation statements of conference calls and absolute abnormal returns around the call dates. They argue that firm transparency helps the capital market manage uncertainty and is therefore perceived as valuable.

Fang et al. (2024) and Hassan et al. (2023) contribute to this literature stream by exploring more nuanced (con)text-based Covid metrics. They examine market reactions to corporate communication of publicly listed firms subject to US securities regulation (Fang et al. (2024)) and globally (Hassan et al. (2023)) during various periods of economic downturns, including the Covid pandemic. The findings by Fang et al. (2024) and Hassan et al. (2023) indicate that higher levels of Covid exposure are associated with negative short-term abnormal returns around conference calls. This aligns with the notion that firms more severely impacted by the pandemic experience lower short-term returns. These results are also reflected in their analysis of the sentiment expressed towards the pandemic, with negative sentiment, in particular, driving adverse market reactions (Hassan et al. (2023); Fang et al. (2024)). However, there is also evidence that general communication features, such as overall tone or sentiment within a call, historically significantly influencing market reactions, became less relevant during the early stages of the Covid crisis (Wang and Xing (2020)). Nevertheless, Howe et al. (2024) show that the market is particularly sensitive to specific forms of crisis communication, such as statements by the CEO that convey empathy in light of the Covid crisis. Howe et al. (2024)

investigate conference calls from January to March 2020 within the Russell 3000 Index and focus especially on human care statements made by CEOs in response to the pandemic. Their findings suggest that, despite being largely generic and vague, such statements were associated with positive market reactions.

Beyond specific conference calls reactions, Kale and Kale (2023) and Wang and Xing (2020) provide valuable information on market reactions to general corporate conference announcements during Covid and to the application of Covid-specific SEC extension on filing deadlines, respectively. First, Kale and Kale (2023) study Covid-related differences in investor responses to conference announcements in the US between January 2018 and December 2020. They find that market reactions to earning announcements are generally exacerbated within the crisis time frame, irrespective whether there are positive or negative earning surprises. However, investor reactions to significant positive conference surprises were found to be quantitatively larger than to negative ones, suggesting that the favorable surprises were considerably less anticipated (Kale and Kale (2023)). Second, at the beginning of the Covid pandemic, the SEC granted extensions to filing deadlines (U.S. Securities and Exchange Commission (2020)). Typically, prior research suggests that late filings are associated with negative market reactions, as delay signals underlying issues and increases uncertainty (Impink et al. (2012); Wang and Xing (2020)). However, in the case of the Covid pandemic, the late-filing relief is intended to address the exceptional disruptions and restrictions associated with the crisis. It could be argued that in this situation, companies should not be penalized for circumstances beyond their control (Wang and Xing (2020)). The empirical analysis of market reactions by Wang and Xing (2020) reveals that late filings, despite being explicitly granted through the SEC, are still associated with significant negative market reactions. Thus, they are not excused given the exceptional nature of the pandemic.

The discussion of empirical findings on capital market reactions during Covid offers valuable insights for assessing market responses to nonanswers. Firstly, prior evidence (e.g., Wang and Xing (2020)) suggests that mechanisms derived from non-crisis periods may not necessarily apply during a crisis. Thus, while nonanswers are generally linked to negative market reactions, a crisis-focused evaluation is critical. On the one hand, nonanswers could signal nondisclosure or the withholding of information, as explained by, for instance, Hollander et al. (2010), Gow et al. (2021), and Barth et al. (2023). However, during a crisis, when uncertainty is heightened, nonanswers may also be interpreted as a natural response to an unpredictable environment. This perspective could be especially relevant for companies particularly exposed to the crisis. In such cases, nonanswers might be viewed as honest and transparent, as they acknowledge the extraordinary circumstances that make precise information unavailable. As a result, the capital market might not penalize nonanswers as much. On the other hand, nonanswers could

exacerbate investors' concerns and uncertainties, particularly regarding the Covid pandemic (see also Wang and Xing (2020)). This could be particularly relevant for companies that are comparatively more severely impacted by the crisis. Furthermore, findings by Wang and Xing (2020) on market reactions to SEC late filings indicate that the market tends to be unforgiving of unusual behaviors, even during exceptional circumstances. Similarly, the results from Kale and Kale (2023) suggest that market reactions are amplified during crises, further emphasizing the potentially heightened sensitivity of investors in such periods. Thus, ultimately, the market's perception of nonanswers in the context of the crisis remains an open empirical question. To explore this, I propose the following hypothesis:

Hypothesis 2: Nonanswers in conference calls in the context of the Covid pandemic are associated with altered market reactions, with the strength of these reactions potentially varying depending on the extent or prevalent nature of the company's exposure to the crisis.

5.3 Data

5.3.1 Sample

The statistical analysis is based on conference call transcripts that are retrieved via the Thomson Reuters StreetEvents data base by LSEG Data & Analytics, a commonly used database in academic research on conference calls (Brochet et al. (2018); Chen et al. (2018); Gow et al. (2021); Rennekamp et al. (2022); Barth et al. (2023)). To maintain high data quality, only edited versions of transcripts are considered. Most prior research on nonanswers and conference calls during crises focuses on US firms (e.g., Hollander et al. (2010); Wang and Xing (2020); Gow et al. (2021); Barth et al. (2023); Falconieri and De Amicis (2023); Fang et al. (2024)), with only a few exceptions, such as Hassan et al. (2023), who examine conference calls globally during the Covid pandemic. Given the global relevance of Covid, this paper extends the analysis to include both US and European firms, offering a broader international perspective on nonanswers and market reactions. I access all calls that have been held in companies that were listed in either the S&P500 (.SPX), Stoxx (.STOXX), or Eurostoxx (.STOXXE). The sample includes all transcripts of conference calls that have been held between 01/01/2016 and 31/12/2023 and that correspond to a financial quarter between 2016Q1 and 2023Q4. I focus on regular quarterly conference calls, as this is the standard disclosure frequency. For consistency, calls with different frequencies (e.g., half year calls) or preliminary calls are excluded from the analysis. Furthermore, I omit fixed income calls as they are a specific subtype of conference calls targeted at debt investors (De Franco et al. (2023)) and therefore lack adequate comparability.

For the analysis, the transcripts are divided into two segments: a presentation statement, delivered by the company's management, and a question and answer session, where management interacts with other participants. As the study relies on measures extracted from both parts of a conference call, I exclude all transcripts where a clear separation between these parts is not possible. Moreover, transcripts with other data provision problems (e.g., missing RIC) and duplicates are omitted. The textual data measures are then merged with company-level information on financial and industry information from LSEG Data & Analytics described in subsection 5.3.2. Balance sheet items and income statement items are matched on a quarter-year level. As the stock price development of firms around call dates is examined, daily stock prices and returns are accessed and matched.[64] The final sample consists of 16,978 quarterly conference calls from 656 firms.[65]

5.3.2 Variables & Measures

To extract relevant information, I make use of the structured format of the edited transcripts provided by the Thomson Reuters StreetEvents database from LSEG Data & Analytics. First, I extract key information regarding the identity of the company and the timing of the call, as for instance, the RIC (company identifier), the date of the call, and the economic quarter to which the call refers to. Then, I extract textual information. Therefore, I disentangle the paragraphs *corporate participants*, *conference call participants*, *presentation statement*, and *questions & answers* according to the dividers in place (=). Within the question and answer session, I separate questions and answers by initially extracting the names of the *corporate representatives* and the *other call participants* listed at the beginning of the transcript. Subsequently, I classify all statements made by corporate participants as answers and all statements from other call participants as questions. For sentence-based analyses, I split the text of a conference call transcript into sentences using punctuation (e.g., period, question mark, exclamation mark). All textual data are analyzed in lower case. One of my main variables of focus, the nonanswer measure by Barth et al. (2023), relies on trigrams. Their identification is only possible if the original wording is maintained. Hence, I forgo any stemming of words.

The unit of analysis is one conference call transcript of company i in quarter t on conference call date d. In addition, per call, three different segments v are analyzed. Firstly, there is the

[64] Table D.1 in Appendix D.1 presents an exact breakdown of the sample restriction.

[65] Table D.2 in Appendix D.1 presents the distribution of the sample by year and index. Generally, the ratio of suitable transcripts from US firms is higher compared to those from European firms, reflecting the generally higher quality of (textual) US data in the database. Also, the amount of transcripts in 2023 is smaller than in the previous years. Data was collected only for calls held up to December 31, 2023. As a result, calls related to 2023Q4 were often unavailable at the time of data collection and are therefore not included in the sample.

management presentation $v = P$. Then, I separate the questions and answer session into two parts: the questions $v = Q$ asked by call participants, and the answers $v = A$ provided by the management. Within the segments, a range of measures is extracted, and their definitions are described below. A comprehensive overview of all variables, including their sources and definitions, is provided in Appendix D.2.

5.3.2.1 Nonanswers

Nonanswers are calculated following the approach presented by Barth et al. (2023). In their research, they applied a supervised machine learning framework to identify a glossary of 1,364 trigrams that signal nonanswers given by management in conference call question and answer sessions.[66] Their glossary contains trigrams (trigram tokens) with their corresponding weights (loadings).[67] Examples of indicative tokens include *"back to you", "not sure i", "off the top", "not know the"*, and *"to wait and"*. By definition, this measure can only be applied to one segment of a conference call, namely the answers provided by the management. For the calculation, I count the amount of nonanswer tokens k used in the management answers in the answer segment $v = A$ of a transcript from company i in quarter t on conference call date d. Then, I weigh the identified nonanswer tokens by their respective factor loadings ϕ_k provided by Barth et al. (2023) that allow a more refined analysis. For comparability, the weighted count is divided by the total amount of words in the management answers in the answer segment $v = A$ of a transcript from company i in quarter t on conference call date d.

Thus, following Barth et al. (2023), I calculate the weighted nonanswer score:

$$Nonanswer^{\phi}_{i,t,d,v=A} = \frac{\sum_{k=1}^{K} \phi_k \, \text{Nonanswer Glossary Tokens}_{i,t,d,k,v=A}}{\text{Total Words}_{i,t,d,v=A}} \tag{5.1}$$

For completeness, non-weighted nonanswer scores are also calculated. These scores are based solely on the amount of nonanswer tokens by Barth et al. (2023) relative to the amount of total words in the management answers in the answer segment $v = A$ of a transcript from company i in quarter t on conference call date d.[68]

[66] The full list of trigrams can be downloaded from Barth et al. (2023) at https://doi.org/10.1287/mnsc.2022.4597.

[67] As explained above, other approaches to quantify nonanswers were brought forward by Gow et al. (2021) and De Kok (2025). All approaches aim to describe similar constructs and have common origins. However, the approach developed by Barth et al. (2023) stands out by explicitly targeting market reactions and examining them empirically. Thus, it is most relevant for the focus of this paper. Hence, for the following analysis, I will rely on the measure constructed by Barth et al. (2023).

[68] See Section 5.5.4 and Appendix D.4.

5.3.2.2 Covid Metrics

In the following analysis, two different approaches are used to measure whether firms are subject to Covid constraints. First, the descriptive analysis will follow a strictly time-based approach, incorporating an indicator variable to capture the Covid period (Falconieri and De Amicis (2023); Kale and Kale (2023)). Secondly, the frequency and tone of verbal references to Covid made by management and other participants in conference calls will serve as a proxy for the extent of Covid-related constraints faced by firms (Wang and Xing (2020); Hassan et al. (2023)). The second approach considers potential heterogeneity concerning the exposure of individual companies within the Covid time frame and thus, allows for a more comprehensive analysis.

As a first step, I will conduct a descriptive analysis of the prevalence of nonanswers during the Covid crisis, employing an indicator variable for the time frame of the Covid pandemic. Given that the indicator variable is purely time-based, it provides only an initial and coarse step, which I will use for a descriptive analysis. It cannot account for the heterogeneous Covid exposure or the associated sentiment expressed by management, investors, and analysts.

Multiple alternative time frames exist for defining the Covid indicator variable. Following Huang et al. (2020), and Szczygielski et al. (2022), December 1, 2019 can be considered as the starting point of the pandemic. Alternatively, other literature suggests emphasizing the beginning of the lockdowns as the decisive point in time (Kale and Kale (2023)) and thus, applies February 29, 2020 as the starting point of the Covid pandemic. Fang et al. (2024) consider December 31, 2020 the end of the pandemic as this was the point in time at which the vaccine was approved and distributed. Yet, only on May 5, 2023, the WHO officially declared the end of the Covid pandemic (World Health Organization (2023)).[69] Taking into account these important events concerning the development of the pandemic, I calculate various indicator variables, that rely on the financial quarters in which they took place: Covid (2019Q4 - 2023Q2), Covid (2020Q1 - 2023Q2), Covid (2019Q4 - 2020Q4), and Covid (2020Q1 - 2020Q4) that equal one in the respective financial quarters to which the calls correspond, and zero, otherwise. Furthermore, I also evaluate the impact of the indicator variables that make use of the specific event dates. The variables Covid (Call Date 01/12/19 - 05/05/23), Covid (Call Date 29/02/20 - 05/05/23), Covid (Call Date 01/12/19 - 31/12/20), Covid (Call Date 29/02/20 - 31/12/20) equal one for conference calls that are held in the respective time periods, and are zero, otherwise.

[69]The start and end dates presented here are illustrative rather than exhaustive, other definitions exist that vary slightly around the points in time considered here.

Secondly, further analyses will exploit textual data from conference calls transcripts to quantify the impact of the pandemic on companies. There are various methods that quantify textual data. Prior literature often relies on the so-called bag-of-words (BoW) method. Thereby, the quantification of textual data is based on the count of word occurrences in texts regardless of their context (see, for instance, Loughran and McDonald (2016)). In line with previous literature, some measures used in this study also follow this approach. The main textual measures indicative of the Covid Exposure of companies rely on a sentence-based design similar to Noh and Zhou (2022). In contrast to applying the BoW method to all words of a transcript or a transcript segment as a whole, a sentence-based design makes use of the idea that a sentence reflects an entire line of thought (Noh and Zhou (2022)). Thus, constructing measures on a sentence level instead of a transcript level can allow for more comprehensive insights. Alternatively, some modifications of the BoW method also focus on word proximity. Thereby, only words around a certain word of interest (e.g., Covid) enter the calculations (Hassan et al. (2023); Jory et al. (2025)). However, in order to circumvent setting an artificial limit of neighboring words, I follow the sentence-based design (Noh and Zhou (2022)).

First of all, the Covid Exposure of a company is determined. Building on the work of Hassan et al. (2023), Fang et al. (2024), and Howe et al. (2024), I compile a glossary that is informative of Covid references. Whenever a sentence contains a word of the Covid glossary, the sentence is flagged. This Covid glossary includes words as, for instance, *"coronavirus", "covid", "pandemic"*, and more.[70] To calculate the Covid Exposure score, the amount of sentences in which at least one Covid reference occurs is divided by the total amount of sentences in the respective segment v of a transcript from company i in quarter t on conference call date d.

$$Covid\ Exposure_{i,t,d,v} = \frac{Covid\ Sentences_{i,t,d,v}}{Total\ Sentences_{i,t,d,v}} \tag{5.2}$$

In order to evaluate how companies perceive the impact of Covid, I also extract the sentiment in which the pandemic is discussed. The measures used align closely with those employed by Noh and Zhou (2022) and Hassan et al. (2023). Again, a sentence-based design is applied. Each sentence that contains a word of the Covid glossary is classified as either positive (Covid Optimism), negative (Covid Blame), or neutral towards the pandemic. Thereby, I make use of the sentiment glossary provided by Loughran and McDonald (2011), which is commonly used

[70]The complete list of Covid-related words used in this analysis is: *"coronavirus", "corona virus", "covid", "covid-19", "covid19", "2019-ncov", "sars-cov-2", "sarscov", "ncov", "wuhan virus", "virus", "pandemic", "pandemics", "epidemic", "epidemics", "infectious disease", "infectious diseases", "outbreak", "outbreaks", "plague", "plagues", "contagious disease", "contagious diseases", "contagious illness", "contagious illnesses", "infectious outbreak", "infectious outbreaks"*. The glossary is based on the word lists for Covid-related research applied by Hassan et al. (2023); Fang et al. (2024) and Howe et al. (2024).

in financial textual analysis research (Mayew and Venkatachalam (2012); Davis et al. (2015); Wang and Xing (2020); Hassan et al. (2023); Barth et al. (2023); Fang et al. (2024); Howe et al. (2024)). If more positive words than negative words appear in a Covid-related sentence, the sentence is classified as positive towards Covid (Covid Optimism). If more negative words than positive words appear in a Covid-related sentence, the sentence is classified as negative towards Covid (Covid Blame). If there are exactly as many positive as negative words, the sentence is categorized as neutral. Since this result is inconclusive and does not allow for reliable conclusions with respect to the accompanying sentiment, neutral sentences are not included in the final analysis. To calculate the Covid Blame (Covid Optimism) score, the amount of sentences with predominantly negative (positive) Covid sentiment is divided by the total amount of sentences in the respective segment v of a transcript from company i in quarter t on conference call date d.

$$Covid\ Blame_{i,t,d,v} = \frac{Predominantly\ Negative\ Covid\ Sentences_{i,t,d,v}}{Total\ Sentences_{i,t,d,v}} \tag{5.3}$$

$$Covid\ Optimism_{i,t,d,v} = \frac{Predominantly\ Positive\ Covid\ Sentences_{i,t,d,v}}{Total\ Sentences_{i,t,d,v}} \tag{5.4}$$

Lastly, I evaluate the degree of uncertainty and risk that accompanies Covid discussions. In line with the other Covid Metrics, the analysis is carried out on a sentence level. Again, the Covid glossary is applied. Each Covid-related sentence is evaluated for uncertainty. Uncertainty-indicating words are drawn from the glossary provided by Loughran and McDonald (2011). If a sentence that contains a Covid glossary word also contains a word that signals uncertainty, the sentence is flagged. To calculate the Covid Uncertainty score, the amount of sentences with Covid Uncertainty is divided by the total amount of sentences in the respective segment v of a transcript from company i in quarter t on conference call date d.

$$Covid\ Uncertainty_{i,t,d,v} = \frac{Uncertain\ Covid\ Sentences_{i,t,d,v}}{Total\ Sentences_{i,t,d,v}} \tag{5.5}$$

The respective Covid textual measures are calculated separately for each of the three segments: the presentation statements, the questions posed, and the answers provided. As described in the literature, the three variables carry different informational value (Matsumoto et al. (2011)) with regard to Covid Exposure. To ensure a comprehensive analysis, all three

segments are evaluated.

Firstly, as outlined earlier, the management's presentation statement is prepared in advance of the conference call. Thus, it offers insight into their ex-ante disclosure decisions (Matsumoto et al. (2011)). When the management addresses the pandemic in their presentation, the textual Covid metrics reflect this focus, capturing not only the pandemic's impact on the specific company and its current situation but also additional insights into the management's assessment of the pandemic, particularly regarding sentiment and uncertainty (Wang and Xing (2020); Fang et al. (2024); Howe et al. (2024)). Covid metrics at the question level reflect the priorities of other call participants, i.e., analysts or investors, and highlight their considerations with respect to the crisis (Wang and Xing (2020)). The management is then confronted with these concerns and can react to them in their answers.

5.3.2.3 Other Textual Measures

In addition to the main variables of interest, other relevant textual measures are calculated (Noh and Zhou (2022); Barth et al. (2023)). Firstly, I analyze the degree of negativity in the transcripts. Following the literature, I apply the glossary provided by Loughran and McDonald (2011). The Negativity Rate$_{i,t,d,v}$ is measured as the total amount of negative words relative to the total number of words in the respective segment v of a transcript from company i in quarter t on conference call date d (Call et al. (2021)).

Secondly, the degree of uncertainty is evaluated. The determination of the Uncertainty Rate$_{i,t,d,v}$ in the respective segment v of a transcript from company i in quarter t on conference call date d follows the one described for the Negativity Rate$_{i,t,d,v}$. However, instead of the amount of negative words, the amount of uncertain words as defined by Loughran and McDonald (2011) is considered (Call et al. (2021)). Some literature also refers to this measure as vagueness (e.g., Falconieri and De Amicis (2023)).

In line with Noh and Zhou (2022), I want to ensure that the variables Covid Blame$_{i,t,d,v}$, Covid Optimism$_{i,t,d,v}$, and Covid Uncertainty$_{i,t,d,v}$ capture unique and distinct information. Therefore, I exclude all sentences containing Covid references in the calculation of the variables Negativity Rate$_{i,t,d,v}$ and Uncertainty Rate$_{i,t,d,v}$.

The procedure outlined with respect to the determination of the Negativity Rate$_{i,t,d,v}$ is also applied to explore the degree of complexity in the transcripts. For the calculation of the Complexity Rate$_{i,t,d,v}$ in the respective segment v of a transcript from company i in quarter t

on conference call date d, I rely on the list of complex words by Loughran and McDonald (2024).

Similarly, I measure the extent to which future outlook and (financial) performance are emphasized in the respective segment v of a transcript from company i in quarter t on conference call date d. Again, I apply the same procedure as described above with respect to the Negativity Rate$_{i,t,d,v}$. However, for the calculation of the variables Future Orientation$_{i,t,d,v}$ and Financial Orientation$_{i,t,d,v}$, I implement the word lists for forward-looking words and financial words provided by Matsumoto et al. (2011) (for a similar approach, see, for instance Bozanic et al. (2018)).

Lastly, I assess the proportion of quantitative hard information provided in conference calls (Campbell et al. (2021)). Following Campbell et al. (2021), I start by extracting the amount of numerical values that occur within speech. Thereby, only numerical values that begin with a space or dollar sign and are composed of numeric characters are counted. Numerical values with separators (period, comma) are only counted once (e.g., monetary values). As year references are not as informative, I exclude all whole numbers between 1950-2024. The numerical values are then related to the amount of meaningful context words. In line with Campbell et al. (2021), I count all words in the respective segment v of a transcript from company i in quarter t on conference call date d that are also part of the financial glossary provided by Loughran and McDonald (2011, 2024), excluding stop words. Thus, I measure the variable Numbers$_{i,t,d,v}$ as the amount of numerical values used relative to the sum of both the total amount of numbers and the amount of glossary words in the glossaries provided in Loughran and McDonald (2011, 2024) in the respective segment v of a transcript from company i in quarter t on conference call date d.

5.3.2.4 Market Reactions

To assess market reactions to conference calls and their content, I examine cumulative abnormal returns (CARs) surrounding the call date (Fuller et al. (2002)). First, I access daily stock price and return data for each individual company i and each index via LSEG Data & Analytics. I use a single market model for estimation with a 149-day estimation window (MacKinlay (1997); Kreidl and Scholz (2021); Wang and Xing (2020)). If a company is listed on more than one index, expected returns are calculated with respect to the largest spanning index. For example, if a firm is listed in both the Stoxx and the Eurostoxx indices, the Stoxx index is considered. For each company i, I regress daily total returns on daily index returns over 149 days, i.e., a rolling window of 175-26 days before the conference call day d (event date). Then, I calculate the expected return of the stock of a company i for each day in the event

window. The abnormal return is defined as the difference between the expected return based on the estimation and the actual return on day d. I use abnormal returns from one day before the call date d to one day after the call date d, resulting in a three-day cumulative abnormal return window around the call date of [-1;1] (Price et al. (2012); Wang and Xing (2020); Campbell et al. (2021); Noh and Zhou (2022); Bai et al. (2023); Hassan et al. (2023); Allee et al. (2023); Fang et al. (2024); Allee et al. (2024)).

5.3.2.5 Other Financial Measures

Other financial control variables at the company level include the natural logarithm of total assets (ln(Assets)), the Book-To-Market Ratio, and Tobin's Q. As a firm's performance can affect the sentiment of a conference call, I control for certain performance measures. These include a company i's Return on Assets in quarter t and quarterly Revenue Growth relative to the previous quarter (Falconieri and De Amicis (2023)). I also control for the Earnings Surprise. Earnings Surprise is defined as the difference between the consensus earnings forecast for the earnings per share and the actually realized earnings per share in quarter t, scaled by the stock price of company i on the call date d (Price et al. (2012); Campbell et al. (2021); Frankel et al. (2022)). In addition, I include industry information, more precisely the GICS sector code of a company i.

All financial variables are retrieved via LSEG Data & Analytics. For each company i, I extract the relevant information in Euro.[71] A complete list of variables can be found in Appendix D.2. Complete descriptive statistics can be found in Table D.3 in Appendix D.3.

5.4 Statistical Analysis

For the analysis, all data is winsorized at the 1% level of each tail of the distribution to ensure that results are not driven by extreme observations. To improve comparability of empirical results in the empirical analyses, I compute the z-score (std.) with zero mean and standard deviation of one of all textual variables in the dataset. All standard errors are clustered on the company level. Furthermore, I include a range of fixed effects (FE) that absorb unobserved, time-invariant heterogeneity. First, Industry FE capture within-industry variation, absorbing unobserved time-invariant industry heterogeneity potentially correlated with both text-derived pandemic exposure and nonanswers. For instance, this matters when industries that are structurally more evasive also exhibit higher pandemic exposure as measured in the textual data. In the same manner, I also add Index FE capturing unobserved time-constant index charac-

[71] Currency conversion does not materially affect the results.

teristics. Lastly, I include Quarter×Year FE to account for period specificities. In additional analyses, Firm FE are included. This controls for firm-specific, time-invariant heterogeneity, such as a company's persistent communication style. As only within firm variation is exploited, this is the strictest approach.

To capture the context and correlations associated with nonanswers, I incorporate a rich set of textual control variables. Firstly, the occurrence of nonanswers is analyzed. Gow et al. (2021) and Barth et al. (2023) show that more negative, uncertain, and complex questions and questions that ask for detailed information, as is frequently the case for numerical values, are associated with a higher level of nonanswers. Hollander et al. (2010) also indicate that questions asking about a company's financial situation are likely to not be answered. Moreover, questions concerning future developments are inherently more difficult to answer and may therefore be more frequently associated with nonanswers. This might even be exacerbated in times of crisis (De Kok (2025)). Thus, I include the textual control variables, Negativity Rate, Uncertainty Rate, Complexity Rate, Numbers, Future Orientation, and Financial Orientation, measured using the question segment. Additionally, I account for management's current outlook on the specified characteristics by including the respective textual control variables measured in the presentation segment. Furthermore, as firm performance may be related the occurrence of nonanswers, I control for various financial variables, such as ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth, and Earnings Surprise, to ensure comparability in firms' economic conditions.

Secondly, market reactions are affected by a wide range of textual and financial factors beyond, but potentially correlated with, the main variable of interest, nonanswers (also see Hollander et al. (2010); Gow et al. (2021); Barth et al. (2023); Falconieri and De Amicis (2023)). To address potential omitted-variable bias, I make use of a rich set of control variables. I employ the same set of textual and financial control variables as outlined above. However, to emphasize the relevance of managerial speech for market responses, the textual controls are measured within both the presentation segment and the answer segment (see, e.g., Falconieri and De Amicis (2023)).

5.5 Results

5.5.1 Descriptive Results

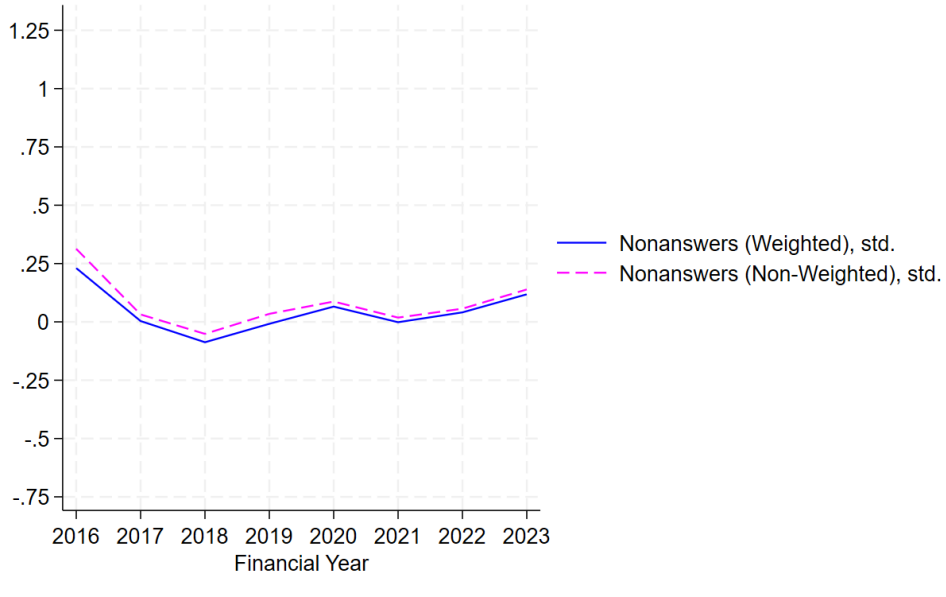
The descriptive analysis first presents Figures 5.1 to 5.7 that illustrate the development of nonanswers and the main textual (Covid) measures by segment, i.e., the presentation, ques-

tion, and answer segment, followed by Table 5.1 comparing these variables across the Covid and non-Covid time periods.

Firstly, Figure 5.1 shows the development of weighted and non-weighted nonanswers following Barth et al. (2023) over the sample time period. Consistent with the literature, nonanswers are generally at a comparatively stable level (Gow et al. (2021); Barth et al. (2023)). Nonetheless, prior literature already indicated potential sensitivity of nonanswers to both economic crises (Barth et al. (2023)), and health crises (De Kok (2025)). In line, in the time frame on display, there seems to be a small increase in (non-)weighted nonanswers around the financial year 2020, the peak year of the Covid pandemic's global impact. Yet, the rise in nonanswers is relatively modest, particularly when compared to the other fluctuations in the measure over the sample time period.

Therefore, there appears to be no clear indication for difficulties in answering questions during the crisis, for instance, due to uncertainties with respect to future developments of the economy or potential restrictions.

Figure 5.1: Development of (Non-)Weighted Nonanswers following Barth et al. (2023)

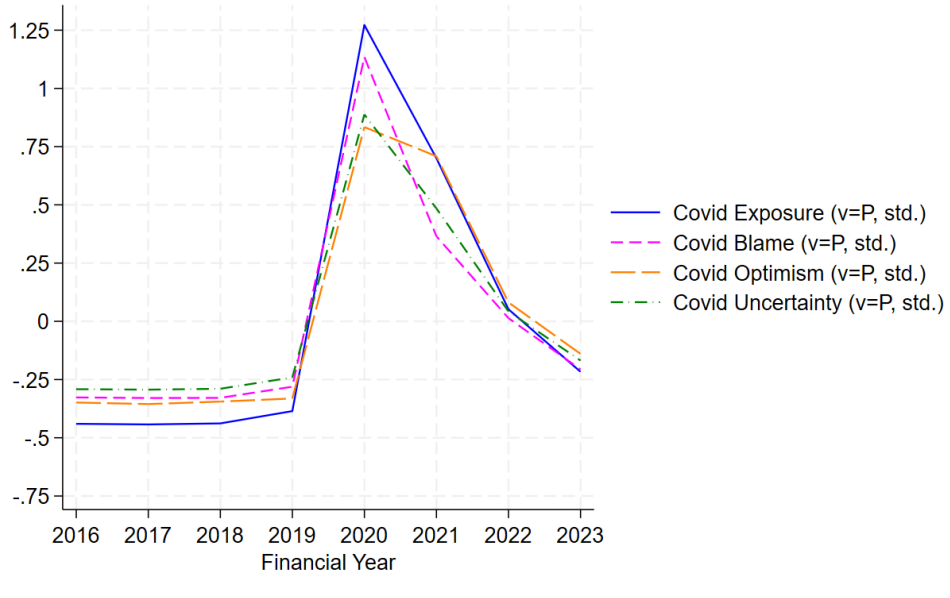


Notes: N = 16,978. The Figure illustrates the yearly average development of both weighted and non-weighted nonanswers, calculated using the methodology proposed by Barth et al. (2023). All variables are standardized.

Figures 5.2 to 5.4 plot the evolution of the textual Covid Metrics within the presentation, question, and answer segments, respectively. Not surprisingly, all textual Covid Metrics peak at the onset of the Covid crisis in the financial year 2020 and decline thereafter. Among the different measures, the most general measure Covid Exposure rises highest, as all sentences that somehow refer to Covid are taken into account (blue solid line). This applies to all three different segments alike. Generally, Covid Blame (pink short-dashed line) is more pronounced than Covid Optimism (orange long-dashed line), especially at the onset of the crisis. This may indicate that the substantial challenges posed by the crisis outweighed potential opportunities, particularly during its initial phase. Over the course of the crisis, Covid Optimism gradually takes over and prevails. Again, all segments show roughly comparable developments.

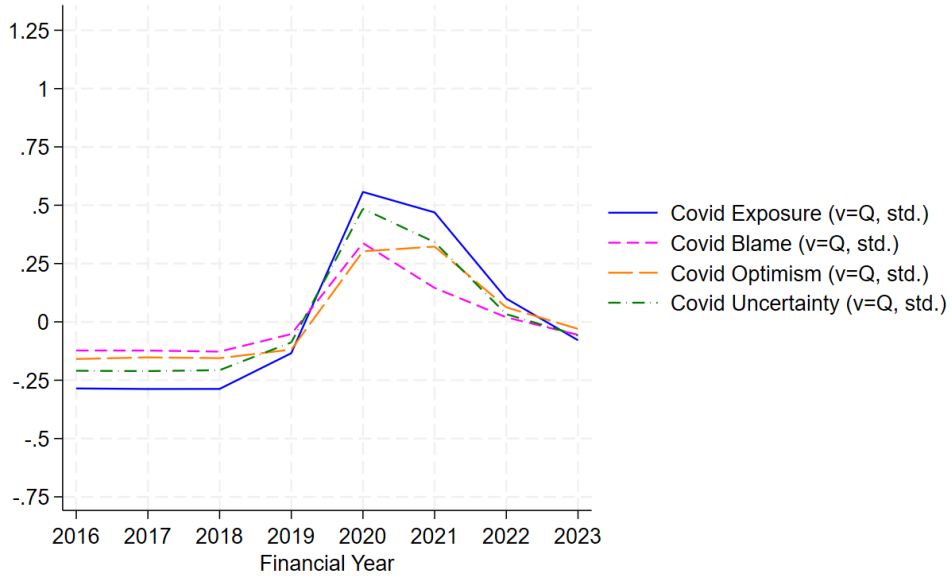
Furthermore, particularly in the questions raised during conference calls and, to a lesser extent, also in the corresponding responses, Covid Uncertainty (green dash-dotted line) is comparatively pronounced. By contrast, the presentation segment is characterized by a greater emphasis on Covid Blame than on Covid Uncertainty. These patterns may indicate that, from the perspective of analysts and investors, the crisis is predominantly framed in terms of uncertainty, which is reflected in the content of their questions. In contrast, management's initial presentation statements appear to focus more on articulating the challenges posed by the crisis. Still, it should be noted, that this observation is based exclusively on descriptive data.

Figure 5.2: Development of Textual Covid Metrics in the Presentation Segments $v = P$



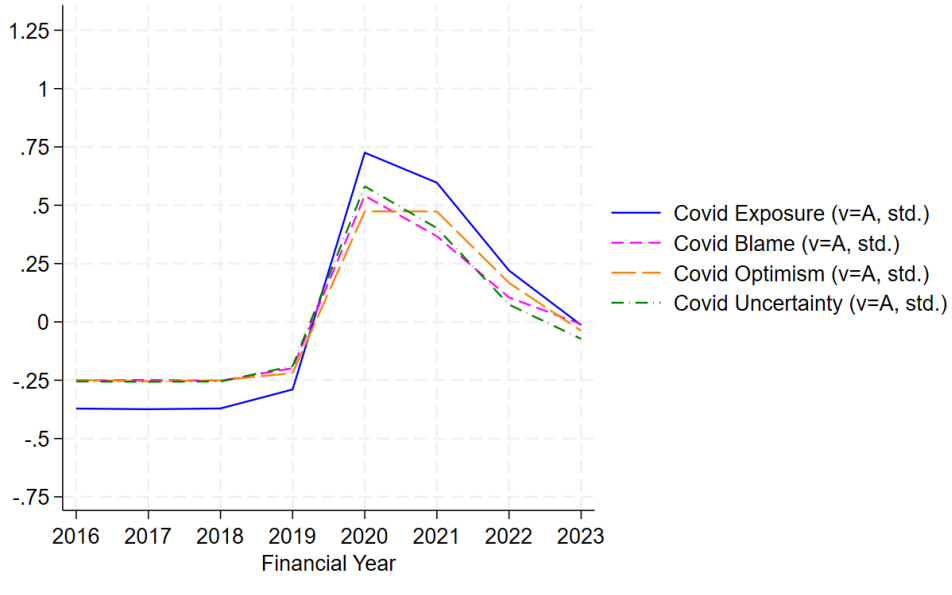
Notes: N = 16,978. The Figure illustrates the yearly average development of the textual Covid Metrics Covid Exposure, Covid Blame, Covid Optimism, and Covid Uncertainty based exclusively on the presentation segments of the conference calls. All variables are standardized.

Figure 5.3: Development of Textual Covid Metrics in the Question Segments $v = Q$



Notes: N = 16,978. The Figure illustrates the yearly average development of the textual Covid Metrics Covid Exposure, Covid Blame, Covid Optimism, and Covid Uncertainty based exclusively on the question segments of the conference calls. All variables are standardized.

Figure 5.4: Development of Textual Covid Metrics in the Answer Segments $v = A$



Notes: N = 16,978. The Figure illustrates the yearly average development of the textual Covid Metrics Covid Exposure, Covid Blame, Covid Optimism, and Covid Uncertainty based exclusively on the answer segments of the conference calls. All variables are standardized.

Additionally, it is important to note that the Covid Metrics slightly fluctuate, also before the Covid pandemic started. This is due to the fact that the measures are triggered by a variety of words related to Covid, but also to other infectious diseases and their implications. Therefore, firms that e.g., research other infectious diseases or firms that are affected by infectious diseases in some other way and hence mention some of the keywords before the official beginning of the pandemic are picked up by the measure.

Figures 5.5 to 5.7 illustrate the development of other textual measures over time for the presentation, the question, and the answer segments.

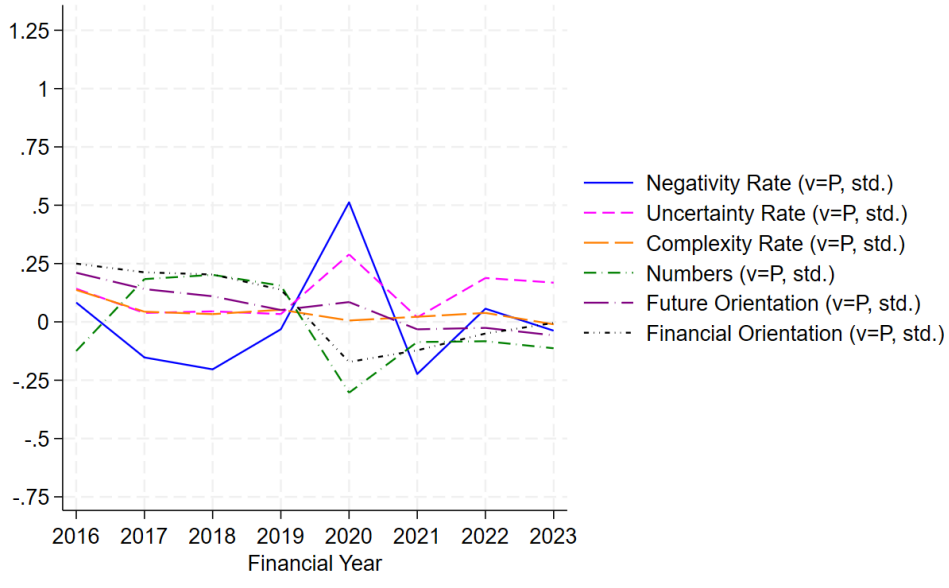
Even though sentences that reference Covid in some way are not considered for the depicted Negativity Rate, there is a substantial increase in general negativity across all segments for the financial year 2020 (blue solid line). The increase is particularly strong for managerial speech, i.e., the presentation and the answer segments. The Uncertainty Rate also experiences an upward shift at the onset of the crisis, even though the rise seems to be slightly less pronounced (pink short-dashed line). This is broadly in line with research by Falconieri and De Amicis (2023) who find similar textual patterns in the time frame of the financial crisis. While both the Negativity Rate and the Uncertainty Rate in the presentation and answer segments quickly bounce back thereafter, there seems to be a lingering increased level of uncertainty for the questions asked.

As the crisis emerges, Numbers, i.e., specific information conveyed in conference calls, are used less frequently, especially when considering the presentation statement (green short dash-dotted line). Given that concrete information is harder to acquire and provide during a crisis, this is in line with expectations.

While the Financial Orientation (black dot–dot–dashed line) in conference calls seems to be weaker during the crisis, the opposite is true with respect to the Future Orientation (purple long dash-dotted line). This phenomenon is most evident in managerial speech. These observations correspond closely with prior findings by Matsumoto et al. (2011). They show that managers focus more on future-oriented statements and less on financial information when firm performance is comparatively poor. This pattern appears to persist even during periods of crisis, i.e., periods potentially posing significant challenges for firms.

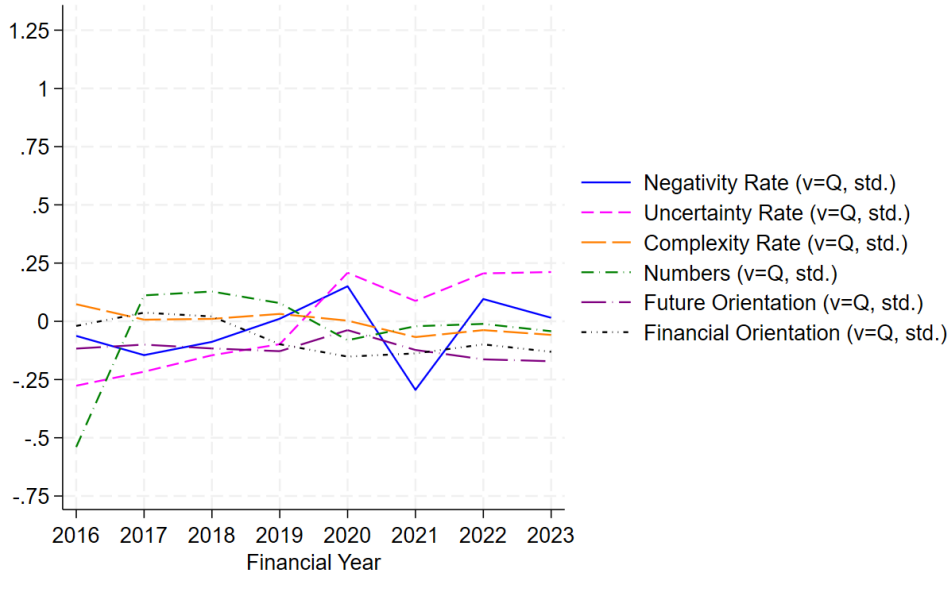
The level of the Complexity Rate stays relatively stable over the sample period for all three segments (orange long-dashed line).

Figure 5.5: Development of Other Textual Characteristics in the Presentation Segments $v = P$



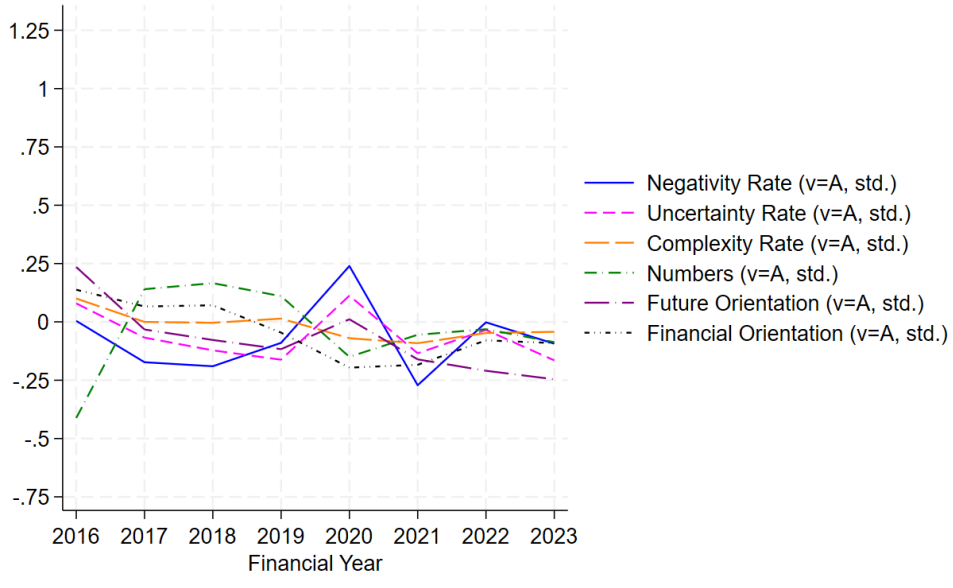
Notes: N = 16,978. The Figure illustrates the yearly average development of the textual measures Negativity Rate, Uncertainty Rate, Complexity Rate, Numbers, Future Orientation, and Financial Orientation based exclusively on the presentation segments of the conference calls. All variables are standardized.

Figure 5.6: Development of Other Textual Characteristics in the Question Segments $v = Q$



Notes: N = 16,978. The Figure illustrates the yearly average development of the textual measures Negativity Rate, Uncertainty Rate, Complexity Rate, Numbers, Future Orientation, and Financial Orientation based exclusively on the question segments of the conference calls. All variables are standardized.

Figure 5.7: Development of Other Textual Characteristics in the Answer Segments $v = A$



Notes: N = 16,978. The Figure illustrates the yearly average development of the textual measures Negativity Rate, Uncertainty Rate, Complexity Rate, Numbers, Future Orientation, and Financial Orientation based exclusively on the answer segments of the conference calls. All variables are standardized.

In addition to the Figures discussed above, Table 5.1 compares the textual variables for the Covid and the Non-Covid time frame quantitatively and with respect to significant differences. Column 1 of Table 5.1 displays mean values of the textual variables for the Non-Covid time frame, while column 2 focuses on the Covid time frame. The split is based upon the time-based Covid variable Covid (2019Q4 - 2020Q4).[72] Column 3 reports the difference between the two values (column 1 - column 2) and indicates statistical significance based on a two-sample t-test with equal variances.

Firstly, the mean of both nonanswers and non-weighted nonanswers is quantitatively very similar for crisis and non-crisis periods. The difference is not statistically different on conventional levels. Thus, the results presented in Table 5.1 are broadly consistent with the patterns observed in the corresponding Figure 5.1. Still, it is somewhat inconsistent with the observations in the literature discussed above that hint towards elevated levels of nonanswers for a crisis period, suggesting the need for further investigation.

[72]In Table 5.1, the results with respect to the Covid variable (2019Q4–2020Q4) is highlighted. Still, all results are largely robust to an alternative time-based Covid data split that makes use of the time-based Covid Metrics Covid (2019Q4-2023Q2) (see Appendix D.5), Covid (2020Q1-2023Q2), Covid (2020Q1-2020Q4), and to the Covid date specific indicator variables (Covid (Call Date 01/12/19-05/05/23), Covid (Call Date 29/02/20-05/05/23), Covid (Call Date 01/12/19-31/12/20), Covid (Call Date 29/02/20-31/12/20) (untabulated).

Table 5.1: Descriptive Statistics

	Non-Covid Period Mean (1)	**Covid Period** Mean (2)	**Difference** (1)-(2) (3)
Nonanswers (Weighted, std.)	0.04236	0.03484	0.00752
Nonanswers (Non-Weighted, std.)	0.07632	0.06409	0.01222
Negativity $\text{Rate}_{\text{v=P}}$ (std.)	-0.07608	0.40085	-0.47693***
Negativity $\text{Rate}_{\text{v=Q}}$ (std.)	-0.06927	0.10727	-0.17654***
Negativity $\text{Rate}_{\text{v=A}}$ (std.)	-0.11774	0.16537	-0.28311***
Uncertainty $\text{Rate}_{\text{v=P}}$ (std.)	0.08855	0.24475	-0.15620***
Uncertainty $\text{Rate}_{\text{v=Q}}$ (std.)	-0.03691	0.15940	-0.19631***
Uncertainty $\text{Rate}_{\text{v=A}}$ (std.)	-0.08381	0.06234	-0.14616***
Complexity $\text{Rate}_{\text{v=P}}$ (std.)	0.04530	0.01677	0.02853
Complexity $\text{Rate}_{\text{v=Q}}$ (std.)	-0.00741	0.01300	-0.02040
Complexity $\text{Rate}_{\text{v=A}}$ (std.)	-0.01200	-0.05184	0.03984*
$\text{Numbers}_{\text{v=P}}$ (std.)	0.01595	-0.19495	0.21090***
$\text{Numbers}_{\text{v=Q}}$ (std.)	-0.04617	-0.02223	-0.02394
$\text{Numbers}_{\text{v=A}}$ (std.)	-0.02669	-0.07776	0.05107**
Future $\text{Orientation}_{\text{v=P}}$ (std.)	0.05301	0.10565	-0.05265**
Future $\text{Orientation}_{\text{v=Q}}$ (std.)	-0.13243	-0.04466	-0.08777***
Future $\text{Orientation}_{\text{v=A}}$ (std.)	-0.08789	0.00167	-0.08956***
Financial $\text{Orientation}_{\text{v=P}}$ (std.)	0.08605	-0.10128	0.18733***
Financial $\text{Orientation}_{\text{v=Q}}$ (std.)	-0.06183	-0.12274	0.06091***
Financial $\text{Orientation}_{\text{v=A}}$ (std.)	-0.01778	-0.16235	0.14457***
Covid $\text{Exposure}_{\text{v=P}}$ (std.)	-0.15696	0.97898	-1.13594***
Covid $\text{Exposure}_{\text{v=Q}}$ (std.)	-0.08286	0.51635	-0.59921***
Covid $\text{Exposure}_{\text{v=A}}$ (std.)	-0.08409	0.57548	-0.65957***
Covid $\text{Blame}_{\text{v=P}}$ (std.)	-0.15078	0.88498	-1.03576***
Covid $\text{Blame}_{\text{v=Q}}$ (std.)	-0.05144	0.30674	-0.35818***
Covid $\text{Blame}_{\text{v=A}}$ (std.)	-0.06892	0.42654	-0.49546***
Covid $\text{Optimism}_{\text{v=P}}$ (std.)	-0.09050	0.61597	-0.70647***
Covid $\text{Optimism}_{\text{v=Q}}$ (std.)	-0.03187	0.24565	-0.27752***
Covid $\text{Optimism}_{\text{v=A}}$ (std.)	-0.04632	0.35670	-0.40302***
Covid $\text{Uncertainty}_{\text{v=P}}$ (std.)	-0.10208	0.69561	-0.79769***
Covid $\text{Uncertainty}_{\text{v=Q}}$ (std.)	-0.06658	0.44692	-0.51350***
Covid $\text{Uncertainty}_{\text{v=A}}$ (std.)	-0.07895	0.47095	-0.54991***

Notes: N=16,978. For definition of variables, see Section 5.3.2. The data is split based on the dummy variable *Covid (2019Q4 - 2020Q4)*. Column 3 shows results of a two-sample t-test with equal variances. All textual variables are standardized. Although coefficient magnitudes differ depending on standardization, the statistical significance remains unchanged. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Regarding the other textual variables, the findings from the visual evidence are broadly consistent. The Negativity Rate and the Uncertainty Rate exhibit statistically and economically significant differences between crisis and non-crisis periods. Similar to above, times of crisis are characterized by higher levels of negativity in conference calls. This phenomenon is particularly pronounced in the presentation statement. Uncertainty also intensifies during the crisis, with the increase especially pronounced at the question level. The use of quantitative hard facts, here proxied by the variable Numbers, is notably reduced in management's discourse during the Covid period across both the presentation and answer segments. During times of crisis, Future Orientation present in conference calls tends to increase, while Financial Orientation rather decreases. The Complexity Rate seems to be fairly consistent.

For the sake of completeness, also the textual-based Covid Metrics are reported. In addition to the factors previously discussed explaining movements in Covid-related metrics prior to the official onset of the crisis, this definition employs a relatively narrow Covid time frame, with the pandemic ending in 2020Q4. Even beyond this period, some firms continue to be affected by the pandemic, albeit to a lesser degree. Nonetheless, the immense differences in Covid Exposure on all levels between the Covid and the Non-Covid time frames suggest that the measurement process works as intended.

5.5.2 Nonanswers & Covid Metrics

Table 5.2 reports the regression results for weighted nonanswers in relation to textual-based Covid Metrics corresponding to both the presentation and the question segments. Textual controls address both relevant question content and management presentation characteristics to ensure completeness.[73]

Covid references in the prepared presentation serve as a proxy for management's revealed Covid Exposure at the point in time of the conference call. Results in all columns, 1 to 4, illustrate a statistically significant positive association between Covid references by the management and the respective level of nonanswers. In detail, an increase in Covid Exposure in the presentation statement by one SD is associated, ceteris paribus, with an increase in weighted nonanswers of about 5% of an SD (column 1). The positive relation persists for all three different Covid sentiments explored (columns 2 to 4). This implies that when management discloses Covid Exposure in their presentation statement, it seems to correlate with a lack of information, knowledge, or willingness, to adequately respond to subsequent questions in the question and answer session.

[73]The results stay robust to only including textual control characteristics on the question level.

Table 5.2: Nonanswers & Covid Metrics

Dependent Variable:	Nonanswers (Weighted, std.)			
	(1)	(2)	(3)	(4)
Covid Exposure$_P$ (std.)	0.0499***			
	(0.0129)			
Covid Blame$_P$ (std.)		0.0260***		
		(0.0101)		
Covid Optimism$_P$ (std.)			0.0242**	
			(0.0116)	
Covid Uncertainty$_P$ (std.)				0.0273***
				(0.0094)
Covid Exposure$_Q$ (std.)	-0.0143			
	(0.0089)			
Covid Blame$_Q$ (std.)		-0.0036		
		(0.0068)		
Covid Optimism$_Q$ (std.)			-0.0190***	
			(0.0072)	
Covid Uncertainty$_Q$ (std.)				-0.0067
				(0.0079)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,978	16,978	16,978	16,978
R-squared	0.111	0.110	0.110	0.110

Notes: Textual control variables include Negativity Rate$_{P,Q}$, Uncertainty Rate$_{P,Q}$, Complexity Rate$_{P,Q}$, Numbers$_{P,Q}$, Future Orientation$_{P,Q}$, Financial Orientation$_{P,Q}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 5.3: Nonanswers & Covid Metrics: Firm FE

Dependent Variable:	Nonanswers (Weighted, std.)			
	(1)	(2)	(3)	(4)
Covid Exposure$_P$ (std.)	0.0305***			
	(0.0101)			
Covid Blame$_P$ (std.)		0.0154**		
		(0.0077)		
Covid Optimism$_P$ (std.)			0.0185**	
			(0.0080)	
Covid Uncertainty$_P$ (std.)				0.0133*
				(0.0078)
Covid Exposure$_Q$ (std.)	-0.0101			
	(0.0070)			
Covid Blame$_Q$ (std.)		-0.0007		
		(0.0050)		
Covid Optimism$_Q$ (std.)			-0.0085	
			(0.0055)	
Covid Uncertainty$_Q$ (std.)				-0.0062
				(0.0063)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,959	16,959	16,959	16,959
Within R-squared	0.007	0.007	0.007	0.007

Notes: Textual control variables include Negativity Rate$_{P,Q}$, Uncertainty Rate$_{P,Q}$, Complexity Rate$_{P,Q}$, Numbers$_{P,Q}$, Future Orientation$_{P,Q}$, Financial Orientation$_{P,Q}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Thus, Covid Exposure as revealed in the presentation statement of a company may be perceived as signal of general uncertainty about the current situation.[74]

References to Covid by other conference call participants during the Q&A segment (e.g., analysts or investors) generally do not appear to be significantly related to nonanswers. Column 1 shows a statistically insignificant coefficient of the variable Covid Exposure on the question level. However, considering the more detailed sentiment-related Covid Metrics, column 3 indicates that there is a negative correlation between nonanswers and questions that reflect a more constructive or opportunity-oriented view of the company's situation for the pandemic. In detail, an increase in Covid Optimism on the question level by one SD is associated, ceteris paribus, with a decrease in weighted nonanswers of about 1.9% of an SD. Questions in which the pandemic is discussed rather negatively or in a way highlighting accompanying uncertainty are not significantly related to nonanswers (columns 2 and 4). Thus, the analysis suggests that especially when analysts speak optimistically about the pandemic, firms are less inclined to resort to nonanswers.[75]

In summary, the presented empirical results point towards a significant positive association between nonanswers and Covid Exposure in the management's presentation statement. This finding suggests that greater emphasis on Covid Exposure in the management's presentation is associated with more challenges in addressing questions from other participants, and thus points to limited knowledge, ability, or willingness to answer questions, for instance about future developments. Furthermore, Covid Exposure as revealed by the analysts on the question level is less indicative of nonanswers. One possible explanation for the generally less informative role of questions is that analysts' references to Covid Exposure in the question and answer segment may not add substantial incremental information beyond what is already conveyed in the management presentation. Due to the high visibility and significant impact of the pandemic, the management often directly acknowledges the extraordinary situation in the presentation. The global and severe nature of the pandemic likely removes much of the incentive for management to obscure exposure, thus, holding everything constant, Covid Exposure as revealed in the management presentation is informative of nonanswers. Still, a negative relation between Covid references and nonanswers is found for questions that reflect a comparatively

[74] In additional analyses, I run the same regressions as above for all three cases: including all Covid sentiment metrics simultaneously, including only the presentation level sentiment metrics, and including only the question level sentiment metrics. Results remain largely consistent, in a sense that they do not provide clear evidence that any specific Covid sentiment metric is driving the findings. While the coefficient estimates are similar, statistical significance is lost in some specifications, likely due to limited unique variation remaining among the metrics.

[75] It could also be argued that meaningful variation in the Covid Exposure measures only becomes apparent after the pandemic began. To address this, I estimate an alternative regression that includes only observations after the crisis started. Results remain consistent.

more opportunity-oriented view of the company's position with respect to Covid. This finding also seems plausible and suggests that when even the analysts and investors adopt a relatively positive framing of the pandemic, there is a reduced need for nonanswers.

All in all, the results provide partial support for Hypothesis 1. However, there is no consistent evidence that the findings are driven by a particular prevailing Covid-related sentiment, which points to the pandemic's substantial and system-wide impact across contexts.

Table 5.3 replicates Table 5.2, but includes Firm FE instead of Industry FE and Index FE.[76] Thus, the focus lies on exploiting within-firm variation (Matsumoto et al. (2011); Call et al. (2024)). This allows for an estimation in which unobserved time-invariant characteristics on the firm-level are controlled for, such as a persistent firm-specific communication style that correlates with both the degree of Covid Exposure in the textual data and the frequency of nonanswers.

While the coefficients in Table 5.3 generally are smaller, the results on the association between textual-based Covid Metrics and nonanswers on the presentation level are qualitatively similar to above. Again, a higher level of Covid Exposure as revealed by the management presentation correlates with a higher level of nonanswers, suggesting they can be understood as a sign for a potential lack of sufficient information, expertise, or willingness to answer. However, when evaluating textual-based Covid Metrics in the question segment, none of the relevant coefficients are statistically different from zero anymore. Thus, the results emphasize the key role of Covid Exposure disclosed in presentation statements with respect to nonanswers and the relatively less prominent incremental role of analysts' questions. Additionally, this suggests that the previously significant negative relation between nonanswers and question-level Covid Optimism in Table 5.2 seems to have captured mostly variation between firms.

5.5.3 Market Reactions on Nonanswers & Covid Metrics

In the following, market reactions to nonanswers are examined, with a focus on changes in this relationship across (sentiment-related) Covid Exposure.

Table 5.4 reports regression results. Market reactions, proxied by cumulative abnormal returns CARs[-1;1] around the respective call dates, are explained by weighted nonanswers that are interacted with various Covid Metrics measured on the presentation level. Thus, I specifically focus on Covid references made by the management in their prepared presentation, which reflect the Covid Exposure of the company at the particular point in time of the call date

[76] The latter are abandoned to avoid near perfect collinearity due to the lack of industry or index switchers.

as revealed by the management. Following Falconieri and De Amicis (2023), textual controls include both elements of the management's speech, the presentation and the answer segments, for completeness.[77]

The first row in Table 5.4 displays the association between nonanswers and CARs[-1;1] when Covid Metrics are set to zero, i.e., due to their standardization to their mean. Thus the coefficients in columns 1 to 4 display the relation between nonanswers and market reaction in form of CARs[-1;1] for average Covid Metrics. All coefficients are significantly different from zero at the 1% level and negative. Thus, the results in row 1 of Table 5.4 confirm a generally negative statistically significant association between weighted nonanswers and CARs[-1;1] around call dates, also for average levels of Covid Metrics.[78] In terms of economic significance, the results in Table 5.4 (row 1, column 1) imply that a one SD increase in nonanswers is associated with a decrease of approximate 2.4% of an SD in CARs[-1;1] around the call date.[79] This result is in line with the original findings by Barth et al. (2023) in terms of both sign and economic magnitude. Thus, consistent with the interpretation by Barth et al. (2023), the relationship carries economic significance.[80]

In the remainder of Table 5.4, deviations from average Covid Metrics are explored. Therefore, interaction terms between nonanswers and Covid Metrics are analyzed. Column 1 shows results for a change in the association of nonanswers for comparatively high levels of Covid Exposure. The coefficient is negative, however, as it is not statistically significant from zero, the sign should not be interpreted. This suggests that the nonanswers by management, also in the context of a relatively high level of Covid Exposure for the company at the time of the conference call, is not considered overly harsh by the market. While nonanswers generally are penalized, there is no clear evidence to suggest that punishment increases with particularly high levels of Covid Exposure.

[77] Barth et al. (2023) focused on textual variables in the management's answers only. The results stay robust to this modification.

[78] Additionally, in untabulated results, I evaluated nonanswers in a non-standardized scaling of crisis variables. Then, setting the non-standardized Covid Metrics to zero indicates the absence of the Covid crisis and results in the interaction term being omitted. A statistically and economically significant negative association between weighted nonanswers and CARs[-1;1] persists, also in the absence of the crisis. In detail, a one SD increase in nonanswers is associated with a decrease of approximately 1.9 - 2.2% of an SD in CARs[-1;1].

[79] The SD of CARs[-1;1] in the regression sample amounts to 5.73677.

[80] Barth et al. (2023) measure the sensitivity between (weighted) nonanswers and a related, though not identical, CAR measure over [0;1], calculated as the sum of abnormal returns on the call date and the following day, using either the three-factor model of Fama and French (1993) or the five-factor model of Fama and French (2015). Quantitatively, their results show that FF3/FF5 CARs[0,1] decrease by 2.7% of an SD for a one SD increase in weighted nonanswers. Though not directly reported, this association can be calculated by multiplying the SD of weighted nonanswers (0.037) with the estimated coefficient of the relationship (-0.038) in Table 3, column 1 in Barth et al. (2023). Relating the result (-0.001406) to their SD of FF3 CARs[0,1] (0.052) amounts to a negative association of approximately 2.7% of an SD. Thus, although the exact measurement of the constructs and the sample (period) differ, the estimated magnitudes are roughly consistent with each other.

Table 5.4: Market Reactions on Nonanswers & Covid Metrics: CARs

Dependent Variable:	CARs[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Weighted, std.)	-0.1363***	-0.1358***	-0.1373***	-0.1358***
	(0.0523)	(0.0521)	(0.0521)	(0.0521)
Nonanswers (Weighted, std.)	-0.0657			
×Covid Exposure$_P$ (std.)	(0.0418)			
Nonanswers (Weighted, std.)		-0.0229		
×Covid Blame$_P$ (std.)		(0.0411)		
Nonanswers (Weighted, std.)			-0.0920**	
×Covid Optimism$_P$ (std.)			(0.0453)	
Nonanswers (Weighted, std.)				-0.0460
×Covid Uncertainty$_P$ (std.)				(0.0406)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,729	16,729	16,729	16,729
R-squared	0.027	0.027	0.028	0.027

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 5.5: Market Reactions on Nonanswers & Covid Metrics: CARs & Firm FE

Dependent Variable:	CARs[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Weighted, std.)	-0.1464**	-0.1459**	-0.1479**	-0.1458**
	(0.0634)	(0.0634)	(0.0635)	(0.0633)
Nonanswers (Weighted, std.)	-0.0912**			
×Covid Exposure$_P$ (std.)	(0.0434)			
Nonanswers (Weighted, std.)		-0.0402		
×Covid Blame$_P$ (std.)		(0.0432)		
Nonanswers (Weighted, std.)			-0.1074**	
×Covid Optimism$_P$ (std.)			(0.0478)	
Nonanswers (Weighted, std.)				-0.0504
×Covid Uncertainty$_P$ (std.)				(0.0412)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,706	16,706	16,706	16,706
Within R-squared	0.030	0.030	0.031	0.030

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The same applies to columns 2 and 4, focusing on the interaction terms of nonanswers and both Covid Blame, and Covid Uncertainty. Again, the negative, but insignificant coefficients suggest that the market does not view nonanswers as excessively negative, also for relatively high levels of negative and uncertain Covid Exposure faced by the company. However, column 3 reveals a noteworthy peculiarity. Here, the interaction term of nonanswers and Covid Optimism is statistically significantly negative and comparatively large. More precisely, the coefficient implies an increase in the total association of approximately two thirds and is therefore economically significant as well. This finding suggests that nonanswers are penalized more heavily by the market when management presents the company's situation during Covid in a comparatively favorable light.[81] A potential explanation for this phenomenon is that when a company reveals a more constructive or opportunity-oriented view of the crisis, the market might place higher expectations on management for transparency and clarity. Consequently, in these cases, nonanswers might be met with stronger market penalties, and investors might be less tolerant of ambiguity or evasive responses. Conversely, when the crisis exposure is comparatively negative or exposure-related uncertainty comparatively high, nonanswers may rather be interpreted as a reflection of the challenging circumstances, and thus, comparatively greater market leniency. Thus, Hypothesis 2 is partially supported.

Table 5.5 replicates the analysis of market reactions and nonanswers for different Covid Metrics on the presentation level with a specific focus on applying Firm FE. Again, nonanswers generally are associated with negative market reactions around the conference call dates, also for average Covid Measurements (row 1, columns 1 to 4).[82] However, when including Firm FE, column 1 indicates a significant negative interaction term for nonanswers and Covid Exposure, pointing to an association in which higher Covid Exposure generally coincides with a more pronounced negative response to nonanswers. This suggests that nonanswers in general are punished more heavily by the market. Yet, similar to above, the interactions in columns 2 and 4 for Covid Blame and Covid Uncertainty and nonanswers are not statistically significant, while the negative and statistically significant coefficient for the interaction term for Covid Optimism in column 3 persists. It needs to be noted that columns 2 to 4 decompose the overall Covid Exposure measure of column 1 into its underlying sentiment components, at least to the extent to which Covid is discussed with an accompanying sentiment. Thus, the results from the fixed effects estimation in Table 5.5 provide tentative evidence, that nonanswers generally are punished more harshly with rising Covid exposure. Nevertheless, in line with the results already

[81] Similar to above, in additional results, I also include all Covid sentiment metrics simultaneously in the regression specification. Results remain qualitatively unchanged. In line with the discussion above, meaningful variation in Covid Exposure measures may only materialize after the start of the pandemic. To test this, I estimate an alternative regression using only observations after the crisis began. The results remain robust.

[82] Again, I also evaluated the same regressions with non-standardized Covid Metrics. The results stay qualitatively similar.

discussed, especially when the company's exposure is framed in a comparatively opportunity-oriented sentiment, nonanswers tend to be associated with stronger negative market reactions. This may reflect higher expectations for transparency and consistency.

5.5.4 Robustness Checks

The analysis is supplemented by a series of robustness tests. First, as documented in Section 5.3.2.1, Barth et al. (2023) provide loadings to assign appropriate weights to trigrams indicative of nonanswers. In their approach to capture nonanswers, they carefully evaluated the loadings and found them to be meaningful. Therefore, the main analysis builds upon them. Nevertheless, to ensure that these weightings are not solely responsible for the results reported above, all tables in Section 5.5.2 and Section 5.5.3 are replicated using non-weighted nonanswer scores. As indicated by Appendix D.4, the results stay largely similar.

Secondly, regarding the sentiment lists used above, the full glossary provided by Loughran and McDonald (2011) is applied. However, prior research has highlighted certain framing specificities in conference call transcripts that may potentially be problematic (Allee and DeAngelis (2015); Wang and Xing (2020)). For example, the words "question" and "questions" are considered as negative words in the glossary by Loughran and McDonald (2016). However, in conference calls, they are not necessarily associated with a negative meaning but rather commonly used words, for instance, when referring to the questions asked in the call. Thus, I exclude them from the negative word list (Allee and DeAngelis (2015); Call et al. (2024)). Furthermore, there are certain typical word combinations to which the same applies, as for instance, "good morning". These word combinations are also excluded from evaluation. This includes "good" in case it is followed by "morning", "afternoon", or "day"; "effective" in case it is followed by "income", "tax", and "rate"; "efficiency" in case it is followed by "ratio"; and "closing" in case it is followed by "remark" or "remarks". This approach closely follows Allee and DeAngelis (2015). Therefore, I apply a refined version of the dictionary as a robustness check. The actual changes in the affected variables are comparatively small. For the majority of affected textual variables, changes in the summary statistics are observable only from the third decimal place onward. There are no substantive changes in results.

Similarly, as a robustness check, the textual sentiment measure $\text{Tone}_{i,t,d,v}$ is calculated, replacing the previously used textual control measure $\text{Negativity}_{i,t,d,v}$. Tone, sometimes also called net negativity or net positivity (Wang and Xing (2020); De Amicis et al. (2021); Call et al. (2024); Falconieri and De Amicis (2023)), is a textual sentiment measure that focuses on the amount of negative words in comparison to the amount of positive words in a given text. I

compute $\text{Tone}_{i,t,d,v}$ by dividing the difference between positive words and negative words by the total amount of positive and negative words in the respective segment v of a transcript from company i in quarter t on conference call date d. Again, the glossary provided by Loughran and McDonald (2011) is employed. I also exclude all sentences containing Covid references in the calculation (Noh and Zhou (2022)). As presented in Appendix D.6, results are qualitatively very similar. Note that the sample is slightly smaller ($N = 16{,}938$ in Table D.9; a reduction of 40 observations) because some transcript segments contain neither positive nor negative words, which would result in division by zero.

Furthermore, a robustness check is conducted for the market reaction measure used in the analysis in Section 5.5.3. Instead of focusing on CARs, I investigate market-adjusted returns. A market-adjusted $\text{return}_{i,t,d}$ of company i on conference call day d is defined as the difference between the individual return of the stock of company i on conference call date d and the index return on that day d. Again, in case a company is listed on more than one index, I calculate the market-adjusted return with respect to the largest spanning index. As documented in Appendix D.7, the results stay robust.

5.6 Conclusion

This paper examines management communication, specifically the provision of nonanswers, in conference calls in an extreme crisis situation, namely the Covid pandemic, and explores corresponding capital market responses. Thus, it sheds light on the potential changes in corporate communication during periods of high uncertainty and emphasizes the critical role of effective messaging in crisis management. The role of nonanswers might be particularly relevant under crisis conditions, where uncertainty is high and clear information is scarce.

The analysis provides several important insights. Firstly, greater crisis exposure as revealed by the management presentation is associated with an increased frequency of nonanswers by management. However, there is no consistent evidence that singles out a particular strong association for a specific Covid-related sentiment driving the relationship. Thus, if management discloses comparatively high levels of Covid Exposure in their presentation, the incidence of nonanswers also tends to be higher, possibly reflecting challenges in assessing future developments across the full spectrum of managerial sentiment.

Secondly, nonanswers remain statistically and economically negatively associated with market responses, also in the context of the Covid pandemic. Notably, especially for firms conveying a comparatively positive and constructive view of the crisis, the negative relation between nonanswers and market responses seems to be even stronger. These findings could indicate

that, under such conditions, investors place elevated value on clear and transparent communication. In terms of managerial implications, this suggests that firms benefiting from a crisis should place particular emphasis on transparency, as in these cases, the market might be less tolerant of nonanswers.

The study is not without limitations. Firstly, text analysis presents a complex unit of measurement, requiring careful consideration of linguistic nuances and contextual variability (Loughran and McDonald (2016)). The identification of nonanswers in this paper is based on trigram patterns. Although this method has demonstrated strong results in prior literature (e.g., Barth et al. (2023)), it may fail to detect instances in which slight deviations from the predefined expressions occur. Given that the analysis is based on spoken language, where variations in phrasing are common, it cannot be definitively concluded that all nonanswers are captured. This is particularly relevant as nonanswers often manifest as subtle elements of speech. Moreover, the study focuses on a unique time period characterized by crisis. If the nature of nonanswers is itself affected by the crisis context, it is possible that their identification is further complicated. Therefore, future research is needed to ensure that nonanswers are comprehensively and accurately detected. Similar concerns also apply to the measurement of Covid Exposure, as it is based on textual elements derived from spoken language. While prior studies (e.g., Noh and Zhou (2022); Hassan et al. (2023)) have successfully employed comparable approaches, future research could refine and validate these measures further, potentially by incorporating large language model techniques. Additionally, it may be insightful to further differentiate between different types of nonanswers, for instance, those in which management genuinely does not know the answer and those in which information is deliberately obscured. However, evaluating this systematically is challenging in general, particularly for large samples, and may become even more complicated when considering Covid-specific dynamics.

Secondly, the scope and nature of the Covid pandemic are unprecedented in both scale and impact. Daily life, global systems, and economic activity were significantly disrupted over an extended period, spanning months and even years. Given the extraordinary nature of this crisis, it remains uncertain to what extent the findings presented here are generalizable to other types of crises or periods of heightened uncertainty. Although prior research (e.g., Fang et al. (2024)) indicates that economic and health-related crises have comparable impacts, the specific implications in this context remain to be fully explored.

Thirdly, the results described above reveal meaningful correlations that shed light on how firms and markets interact in different contexts. Nevertheless, the study should not be interpreted as providing robust causal evidence. Future work could build on these patterns to explore underlying causal mechanisms more directly.

Appendix D

D.1 Sample Composition & Restriction

Table D.1: Sample Restriction

Transcripts available on StreetEvents per index held within 01/01/2016-31/12/2023 and corresponding to financial quarters 2016Q1-2023Q4	**33,999**
Data Cleaning	
Calls with inconsistencies in dividers between presentation and question and answer part removed	(883)
Calls with formal inconsistencies removed (e.g., missing RIC identification, keep only edited version)	(96)
Duplicate entries removed (e.g., cases of multiple listings across indices)	(6,492)
Call Filtering	
Fixed income calls removed	(93)
Non-quarterly conference calls removed (e.g., half-year/preliminary calls)	(3,743)
Calls with no full financial data match removed (i.e., missing financial data)	(5,714)
Final Sample	**16,978**

Table D.2: Sample Composition

Year	Firms	Number of Transcripts			
		Overall	S&P500	Stoxx	Eurostoxx
2016	564	2,023	1,619	404	241
2017	593	2,165	1,713	452	253
2018	598	2,188	1,719	467	265
2019	596	2,223	1,735	488	287
2020	592	2,209	1,738	471	276
2021	599	2,230	1,746	484	280
2022	607	2,246	1,768	478	276
2023	596	1,694	1,337	354	201
Total	656	16,978	13,375	3,598	2,079

Notes: The year indicated reflects the financial year to which the conference call corresponds. The total number of transcripts by index does not equal the overall number of observations, as some companies are listed in multiple indexes. The amount of transcripts in 2023 is smaller than in the previous years. Data was collected only for calls held up to December 31, 2023. As a result, calls related to 2023Q4 were often unavailable at the time of data collection and are therefore not included in the sample.

D.2 Variable Definitions

Textual Variables:

- *Weighted Nonanswer Score following Barth et al. (2023)*$_{i,t,d,k,v=A}$:

$$Nonanswer^{\phi}_{i,t,d,v=A} = \frac{\sum_{k=1}^{K} \phi_k \text{ Nonanswer Glossary Tokens}_{i,t,d,k,v=A}}{\text{Total Words}_{i,t,d,v=A}} \tag{5.1}$$

- *Non-Weighted Nonanswer Score following Barth et al. (2023)*$_{i,t,d,k,v=A}$:

$$Nonanswer_{i,t,d,v=A} = \frac{Nonanswer\ Glossary\ Tokens_{i,t,d,v=A}}{Total\ Words_{i,t,d,v=A}} \tag{5.6}$$

- *Covid Exposure*$_{i,t,d,v}$:

$$CovidExposure_{i,t,d,v} = \frac{CovidSentences_{i,t,d,v}}{TotalSentences_{i,t,d,v}} \tag{5.2}$$

  The respective transcript segment is segmented into sentences based on punctuation, with the total number of sentences serving as the denominator. A sentence is classified as a Covid sentence if it contains at least one term from the Covid dictionary. This dictionary, based on the work of Hassan et al. (2023), Fang et al. (2024), and Howe et al. (2024), is reproduced in full in Section 5.3.2.2.

- *Covid Blame*$_{i,t,d,v}$:

$$CovidBlame_{i,t,d,v} = \frac{PredominantlyNegativeCovidSentences_{i,t,d,v}}{TotalSentences_{i,t,d,v}} \tag{5.3}$$

  The respective transcript segment is segmented into sentences based on punctuation, with the total number of sentences serving as the denominator. A sentence is classified as a predominantly negative Covid sentence if it contains at least one term from the Covid dictionary and the number of negative words exceeds the number of positive words, as defined by the sentiment word glossary published by Loughran and McDonald (2011). The Covid dictionary, based on the work of Hassan et al. (2023), Fang et al. (2024), and Howe et al. (2024), is reproduced in full in Section 5.3.2.2.

- *Covid Optimism*$_{i,t,d,v}$:

$$CovidOptimism_{i,t,d,v} = \frac{PredominantlyPositiveCovidSentences_{i,t,d,v}}{TotalSentences_{i,t,d,v}} \tag{5.4}$$

  The respective transcript segment is segmented into sentences based on punctuation, with

the total number of sentences serving as the denominator. A sentence is classified as a predominantly positive Covid sentence if it contains at least one term from the Covid dictionary and the number of positive words exceeds the number of negative words, as defined by the sentiment word glossary published by Loughran and McDonald (2011). The Covid dictionary, based on the work of Hassan et al. (2023), Fang et al. (2024), and Howe et al. (2024), is reproduced in full in Section 5.3.2.2.

- *Covid Uncertainty*$_{i,t,d,v}$:

$$CovidUncertainty_{i,t,d,v} = \frac{UncertainCovidSentences_{i,t,d,v}}{TotalSentences_{i,t,d,v}} \tag{5.5}$$

The respective transcript segment is segmented into sentences based on punctuation, with the total number of sentences serving as the denominator. A sentence is classified as an uncertain Covid sentence if it contains at least one term from the Covid dictionary and at least one word of the list of uncertain words following the glossary published by Loughran and McDonald (2011). The Covid dictionary, based on the work of Hassan et al. (2023), Fang et al. (2024), and Howe et al. (2024), is reproduced in full in Section 5.3.2.2.

- *Negativity*$_{i,t,d,v}$:

$$Negativity_{i,t,d,v} = \frac{NegativeWords_{i,t,d,v}}{TotalWords_{i,t,d,v}} \tag{5.7}$$

The list of negative words follows the glossary published by Loughran and McDonald (2011).

- *Uncertainty*$_{i,t,d,v}$:

$$Uncertainty_{i,t,d,v} = \frac{UncertaintyWords_{i,t,d,v}}{TotalWords_{i,t,d,v}} \tag{5.8}$$

The list of uncertain words follows the glossary published by Loughran and McDonald (2011).

- *Complexity*$_{i,t,d,v}$:

$$Complexity_{i,t,d,v} = \frac{ComplexWords_{i,t,d,v}}{TotalWords_{i,t,d,v}} \tag{5.9}$$

The list of complex words follows the glossary published by Loughran and McDonald (2024).

- *Numbers*$_{i,t,d,v}$:

$$Numbers_{i,t,d,v} = \frac{NumericalValues_{i,t,d,v}}{ContextGlossaryWords_{i,t,d,v} + NumericalValues_{i,t,d,v}} \tag{5.10}$$

The approach taken here follows Campbell et al. (2021). Firstly, the number of numerical

values that occur within speech is extracted. Thereby, all numerical values that begin with a space or dollar sign and are composed of numeric characters only are counted. Numerical values with separators (period, comma) are only counted once. As year references are not as informative, I exclude all whole numbers between 1950-2024. The numerical values are related to the amount of meaningful context words. In line with Campbell et al. (2021), I count all words in the respective section v in a conference call transcript of company i in quarter t that are also part of the financial glossary provided by Loughran and McDonald (2011, 2024), excluding stop words.

- *Future Orientation*$_{i,t,d,v}$:

$$FutureOrientation_{i,t,d,v} = \frac{Forward - lookingWords_{i,t,d,v}}{TotalWords_{i,t,d,v}} \tag{5.11}$$

The list of forward-looking words follows the glossary published by Matsumoto et al. (2011).

- *Financial Orientation*$_{i,t,d,v}$:

$$FinancialOrientation_{i,t,d,v} = \frac{FinancialWords_{i,t,d,v}}{TotalWords_{i,t,d,v}} \tag{5.12}$$

The list of financial words follows the glossary published by Matsumoto et al. (2011).

- $Tone_{i,t,d,v}$:

$$Tone_{i,t,d,v} = \frac{PositiveWords_{i,t,d,v} - NegativeWords_{i,t,d,v}}{PositiveWords_{i,t,d,v} + NegativeWords_{i,t,d,v}} \tag{5.13}$$

The list of negative and positive words follows the glossary published by Loughran and McDonald (2011).

Time-Related Variables:

- *Covid (2019Q4-2020Q4)*: Indicator variable equal to one if the respective financial quarters to which the conference call corresponds is between 2019Q4-2020Q4.
- *Covid (2020Q1-2020Q4)*: Indicator variable equal to one if the respective financial quarters to which the conference call corresponds is between 2020Q1-2020Q4.
- *Covid (2019Q4-2023Q2)*: Indicator variable equal to one if the respective financial quarters to which the conference call corresponds is between 2019Q4-2023Q2.
- *Covid (2020Q1-2023Q2)*: Indicator variable equal to one if the respective financial quarters to which the conference call corresponds is between 2020Q1-2023Q2.

- *Covid (Call Date 29/02/20-05/05/23)*: Indicator variable equal to one if the conference call takes place between February 29, 2020 and May 5, 2023.

- *Covid (Call Date 01/12/19-05/05/23)*: Indicator variable equal to one if the conference call takes place between December 1, 2019 and May 5, 2023.

- *Covid (Call Date 29/02/20-31/12/20)*: Indicator variable equal to one if the conference call takes place between February 29, 2020 and December 31, 2020.

- *Covid (Call Date 01/12/19-31/12/20)*: Indicator variable equal to one if the conference call takes place between December 1, 2019 and December 31, 2020.

Financial Variables:

- *CARs [-1;1]$_{id}$*: Cumulative abnormal returns (CARs) within a three-day window [-1;1] around the conference call date, estimated via a single market model over a 149-day estimation window using daily stock and index returns from LSEG Data & Analytics (Thomson Reuters).

- *Market-adjusted Returns [-1;1]$_{id}$*: Market-adjusted return within a three-day window [-1;1] around the conference call date, defined as the difference between the stock's daily return and the corresponding market index return.

- *ln(Assets)$_{it}$*: Natural logarithm of total assets (TR.F.TotAssets) retrieved via LSEG Data & Analytics (Thomson Reuters).

- *Tobin's Q_{it}*: Market capitalization (TR.F.MktCap) divided by total assets (TR.F.TotAssets), both retrieved via LSEG Data & Analytics (Thomson Reuters).

$$Tobin'sQ_{it} = \frac{TR.F.MktCap_{it}}{TR.F.TotAssets_{it}} \tag{5.14}$$

- *Return on Assets$_{it}$*: Return on Assets (TR.F.ReturnAvgTotAssetsPctTTM) retrieved via LSEG Data & Analytics (Thomson Reuters).

- *Revenue Growth$_{it}$*: Calculated revenue growth based on the total revenue data item (TR.F.TotRevenue) retrieved via LSEG Data & Analytics (Thomson Reuters).

$$\text{Revenue Growth}_{i,t} = \frac{\text{TR.F.TotRevenue}_{i,t} - \text{TR.F.TotRevenue}_{i,t-1}}{\text{TR.F.TotRevenue}_{i,t-1}} \tag{5.15}$$

- *Book-To-Market Ratio*$_{it}$: Common Equity (TR.F.ComEqTot) divided by market capitalization (TR.F.MktCap), both retrieved via LSEG Data & Analytics (Thomson Reuters).

$$Book - To - MarketRatio_{it} = \frac{TR.F.ComEqTot_{it}}{TR.F.MktCap_{it}} \tag{5.16}$$

- *Earnings Surprise*$_{it}$: Earnings Surprise is defined as the difference between the consensus earnings forecast for the earnings per share (TR.EPSMean) and the actually realized earnings per share (TR.EPSActValue) in quarter t, scaled by the closing stock price of company i on the call date d (CloseStockPrice). All variables retrieved via LSEG Data & Analytics (Thomson Reuters).

$$\text{Earnings Surprise}_{i,t,d} = \frac{\text{TR.EPSActValue}_{i,t,d} - \text{TR.EPSMean}_{i,t,d}}{\text{CloseStockPrice}_{i,d}} \tag{5.17}$$

- *GICSCode*$_i$: GICS Sector Code Name (TR.GICSSectorCode) retrieved via LSEG Data & Analytics (Thomson Reuters).

D.3 Descriptive Statistics - Overall Sample

Table D.3: Descriptive Statistics - Overall Sample

	Mean	SD	Min	Median	Max
Textual Variables:					
Nonanswers (Weighted)	0.11134	0.02639	0.05562	0.10941	0.18702
Nonanswers (Non-Weighted)	0.05163	0.01096	0.02740	0.05110	0.08097
Covid Exposure$_{P}$	0.00905	0.01948	0.00000	0.00000	0.23932
Covid Exposure$_{Q}$	0.00321	0.01047	0.00000	0.00000	0.33333
Covid Exposure$_{A}$	0.00407	0.01009	0.00000	0.00000	0.15686
Covid Blame$_{P}$	0.00228	0.00655	0.00000	0.00000	0.11111
Covid Blame$_{Q}$	0.00057	0.00423	0.00000	0.00000	0.33333
Covid Blame$_{A}$	0.00075	0.00271	0.00000	0.00000	0.05641
Covid Optimism$_{P}$	0.00241	0.00647	0.00000	0.00000	0.08163
Covid Optimism$_{Q}$	0.00063	0.00373	0.00000	0.00000	0.08696
Covid Optimism$_{A}$	0.00086	0.00304	0.00000	0.00000	0.04575
Covid Uncertainty$_{P}$	0.00217	0.00712	0.00000	0.00000	0.13333
Covid Uncertainty$_{Q}$	0.00121	0.00552	0.00000	0.00000	0.12195
Covid Uncertainty$_{A}$	0.00085	0.00308	0.00000	0.00000	0.06471
Negativity Rate$_{P}$	0.00842	0.00453	0.00058	0.00759	0.02291
Negativity Rate$_{Q}$	0.00721	0.00389	0.00000	0.00663	0.02007
Negativity Rate$_{A}$	0.00640	0.00304	0.00139	0.00589	0.01736
Uncertainty Rate$_{P}$	0.00838	0.00380	0.00115	0.00790	0.02119
Uncertainty Rate$_{Q}$	0.01786	0.00486	0.00688	0.01756	0.03141
Uncertainty Rate$_{A}$	0.00820	0.00315	0.00223	0.00783	0.01792
Complexity Rate$_{P}$	0.00231	0.00186	0.00000	0.00190	0.00847
Complexity Rate$_{Q}$	0.00096	0.00125	0.00000	0.00062	0.00599
Complexity Rate$_{A}$	0.00095	0.00092	0.00000	0.00072	0.00444
Numbers$_{P}$	0.05478	0.01896	0.00990	0.05324	0.11220
Numbers$_{Q}$	0.02361	0.01148	0.00245	0.02216	0.05935
Numbers$_{A}$	0.02306	0.01077	0.00416	0.02153	0.05711
Future Orientation$_{P}$	0.01623	0.00541	0.00504	0.01569	0.03218
Future Orientation$_{Q}$	0.02272	0.00677	0.00956	0.02197	0.04552
Future Orientation$_{A}$	0.02060	0.00633	0.00852	0.01992	0.04115
Financial Orientation$_{P}$	0.03570	0.01228	0.00948	0.03464	0.06989

Descriptive Statistics - Overall Sample (continued)

	Mean	SD	Min	Median	Max
Financial Orientation$_{Q}$	0.01696	0.00772	0.00254	0.01598	0.04024
Financial Orientation$_{A}$	0.01350	0.00620	0.00236	0.01250	0.03310
Tone$_{P}$	0.46984	0.24723	-0.24138	0.50000	0.93798
Tone$_{Q}$	0.24897	0.32274	-0.60000	0.26667	1.00000
Tone$_{A}$	0.42652	0.25287	-0.33333	0.46032	0.87500
Financial Variables:					
CARs [-1;1]	0.19757	5.73677	-39.88557	0.17166	58.89388
Market-adjusted Returns [-1;1]	0.26354	5.74862	-39.41781	0.26155	59.42744
ln(Assets)	23.72446	1.49663	20.30251	23.61590	27.78867
Earnings Surprise	0.00111	0.00408	-0.01621	0.00048	0.02217
Tobin's Q	1.87754	2.16584	0.02980	1.14968	11.96267
Book-To-Market Ratio	0.40721	0.38950	-0.11367	0.30121	2.28158
Return on Assets	6.60535	7.00820	-12.80711	5.29162	31.00156
Revenue growth	3.06841	16.71854	-48.64945	2.03107	84.26599

Notes: N = 16,978. Variables not standardized. A description of all variables is provided in Appendix D.2.

D.4 Non-Weighted Nonanswers

Table D.4: Nonanswers (Non-Weighted) & Textual-Based Covid Analysis

Dependent Variable:	Nonanswers (Non-Weighted, std.)			
	(1)	(2)	(3)	(4)
Covid Exposure$_P$ (std.)	0.0502***			
	(0.0136)			
Covid Blame$_P$ (std.)		0.0252**		
		(0.0103)		
Covid Optimism$_P$ (std.)			0.0222*	
			(0.0122)	
Covid Uncertainty$_P$ (std.)				0.0272***
				(0.0094)
Covid Exposure$_Q$ (std.)	-0.0148			
	(0.0092)			
Covid Blame$_Q$ (std.)		-0.0030		
		(0.0060)		
Covid Optimism$_Q$ (std.)			-0.0150**	
			(0.0076)	
Covid Uncertainty$_Q$ (std.)				-0.0078
				(0.0077)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,978	16,978	16,978	16,978
R-squared	0.153	0.152	0.152	0.152

Notes: Textual control variables include Negativity Rate$_{P,Q}$, Uncertainty Rate$_{P,Q}$, Complexity Rate$_{P,Q}$, Numbers$_{P,Q}$, Future Orientation$_{P,Q}$, Financial Orientation$_{P,Q}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.5: Nonanswers (Non-Weighted) & Textual-Based Covid Analysis: Firm FE

Dependent Variable:	Nonanswers (Non-Weighted, std.)			
	(1)	(2)	(3)	(4)
Covid Exposure$_P$ (std.)	0.0304***			
	(0.0090)			
Covid Blame$_P$ (std.)		0.0151**		
		(0.0069)		
Covid Optimism$_P$ (std.)			0.0181**	
			(0.0074)	
Covid Uncertainty$_P$ (std.)				0.0118*
				(0.0065)
Covid Exposure$_Q$ (std.)	-0.0109*			
	(0.0065)			
Covid Blame$_Q$ (std.)		-0.0003		
		(0.0043)		
Covid Optimism$_Q$ (std.)			-0.0061	
			(0.0051)	
Covid Uncertainty$_Q$ (std.)				-0.0067
				(0.0055)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,959	16,959	16,959	16,959
Within R-squared	0.006	0.005	0.005	0.005

Notes: Textual control variables include Negativity Rate$_{P,Q}$, Uncertainty Rate$_{P,Q}$, Complexity Rate$_{P,Q}$, Numbers$_{P,Q}$, Future Orientation$_{P,Q}$, Financial Orientation$_{P,Q}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.6: Market Reactions on Nonanswers (Non-Weighted) & Covid Metrics: CARs

Dependent Variable:	CARs[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Non-Weighted, std.)	-0.1069*	-0.1062*	-0.1076**	-0.1067*
	(0.0545)	(0.0543)	(0.0543)	(0.0543)
Nonanswers (Non-Weighted, std.)	-0.0320			
×Covid Exposure$_P$ (std.)	(0.0443)			
Nonanswers (Non-Weighted, std.)		0.0166		
×Covid Blame$_P$ (std.)		(0.0417)		
Nonanswers (Non-Weighted, std.)			-0.0701	
×Covid Optimism$_P$ (std.)			(0.0433)	
Nonanswers (Non-Weighted, std.)				-0.0197
×Covid Uncertainty$_P$ (std.)				(0.0415)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,729	16,729	16,729	16,729
R-squared	0.027	0.027	0.027	0.027

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.7: Market Reactions on Nonanswers (Non-Weighted) & Covid Metrics: CARs & Firm FE

Dependent Variable:	CARs[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Non-Weighted, std.)	-0.1334*	-0.1315*	-0.1353*	-0.1324*
	(0.0689)	(0.0690)	(0.0691)	(0.0689)
Nonanswers (Non-Weighted, std.)	-0.0526			
×Covid Exposure$_P$ (std.)	(0.0466)			
Nonanswers (Non-Weighted, std.)		0.0066		
×Covid Blame$_P$ (std.)		(0.0434)		
Nonanswers (Non-Weighted, std.)			-0.0845*	
×Covid Optimism$_P$ (std.)			(0.0468)	
Nonanswers (Non-Weighted, std.)				-0.0187
×Covid Uncertainty$_P$ (std.)				(0.0416)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,706	16,706	16,706	16,706
Within R-squared	0.030	0.030	0.030	0.030

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

D.5 Alternative Covid Time Frame - Covid 2019Q4 - 2023Q2

Table D.8: Descriptive Statistics: Alternative Covid time frame (2019Q4 - 2023Q2)

	Non-Covid Period Mean (1)	**Covid Period** Mean (2)	**Difference** (1)-(2) (3)
Nonanswers (Weighted, std.)	0.03018	0.05248	-0.02230
Nonanswers (Non-Weighted, std.)	0.07873	0.06980	0.00892
Negativity Rate$_{v=P}$ (std.)	-0.07517	0.08003	-0.15520***
Negativity Rate$_{v=Q}$ (std.)	-0.06650	-0.01400	-0.05250***
Negativity Rate$_{v=A}$ (std.)	-0.11245	-0.02998	-0.08247***
Uncertainty Rate$_{v=P}$ (std.)	0.06809	0.16114	-0.09306***
Uncertainty Rate$_{v=Q}$ (std.)	-0.16096	0.15597	-0.31692***
Uncertainty Rate$_{v=A}$ (std.)	-0.07232	-0.04757	-0.02475*
Complexity Rate$_{v=P}$ (std.)	0.06028	0.02042	0.03986
Complexity Rate$_{v=Q}$ (std.)	0.02486	-0.03405	0.05891**
Complexity Rate$_{v=A}$ (std.)	0.02146	-0.05970	0.08116***
Numbers$_{v=P}$ (std.)	0.08765	-0.12761	0.21526***
Numbers$_{v=Q}$ (std.)	-0.06072	-0.02324	-0.03747***
Numbers$_{v=A}$ (std.)	-0.00787	-0.06296	0.05509***
Future Orientation$_{v=P}$ (std.)	0.11146	0.00992	0.10154***
Future Orientation$_{v=Q}$ (std.)	-0.11722	-0.11926	0.00204
Future Orientation$_{v=A}$ (std.)	-0.01381	-0.13498	0.12117***
Financial Orientation$_{v=P}$ (std.)	0.18836	-0.08141	0.26977***
Financial Orientation$_{v=Q}$ (std.)	-0.01734	-0.12787	0.11053***
Financial Orientation$_{v=A}$ (std.)	0.05530	-0.14094	0.19624***
Covid Exposure$_{v=P}$ (std.)	-0.42852	0.49785	-0.92637***
Covid Exposure$_{v=Q}$ (std.)	-0.27619	0.31433	-0.59052***
Covid Exposure$_{v=A}$ (std.)	-0.35188	0.40994	-0.76182***
Covid Blame$_{v=P}$ (std.)	-0.32245	0.36776	-0.69021***
Covid Blame$_{v=Q}$ (std.)	-0.12120	0.13862	-0.25982***
Covid Blame$_{v=A}$ (std.)	-0.23497	0.26589	-0.50086***
Covid Optimism$_{v=P}$ (std.)	-0.33912	0.39917	-0.73829***
Covid Optimism$_{v=Q}$ (std.)	-0.14967	0.18131	-0.33098***
Covid Optimism$_{v=A}$ (std.)	-0.24106	0.28772	-0.52879***
Covid Uncertainty$_{v=P}$ (std.)	-0.28492	0.34961	-0.63453***
Covid Uncertainty$_{v=Q}$ (std.)	-0.19918	0.23960	-0.43878***
Covid Uncertainty$_{v=A}$ (std.)	-0.24480	0.27358	-0.51837***

Notes: N=16,978. For definition of variables, see Section 5.3.2. The data is split based on the dummy variable *Covid (2019Q4 - 2023Q2)*. Column 3 shows results of a two-sample t-test with equal variances. All textual variables are standardized. Although coefficient magnitudes differ depending on standardization, the statistical significance remains unchanged. * p<0.10, ** p<0.05, *** p<0.01.

D.6 Alternative Control Measure - Tone

Table D.9: Nonanswers & Covid Metrics *Control: Tone*

Dependent Variable:	Nonanswers (Weighted, std.)			
	(1)	(2)	(3)	(4)
Covid Exposure$_P$ (std.)	0.0495***			
	(0.0130)			
Covid Blame$_P$ (std.)		0.0270***		
		(0.0102)		
Covid Optimism$_P$ (std.)			0.0246**	
			(0.0116)	
Covid Uncertainty$_P$ (std.)				0.0262***
				(0.0097)
Covid Exposure$_Q$ (std.)	-0.0141			
	(0.0089)			
Covid Blame$_Q$ (std.)		-0.0036		
		(0.0068)		
Covid Optimism$_Q$ (std.)			-0.0190***	
			(0.0072)	
Covid Uncertainty$_Q$ (std.)				-0.0060
				(0.0078)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,938	16,938	16,938	16,938
R-squared	0.110	0.109	0.109	0.109

Notes: Textual control variables include Tone$_{P,Q}$, Uncertainty Rate$_{P,Q}$, Complexity Rate$_{P,Q}$, Numbers$_{P,Q}$, Future Orientation$_{P,Q}$, Financial Orientation$_{P,Q}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.10: Nonanswers & Covid Metrics: Firm FE *Control: Tone*

Dependent Variable:	Nonanswers (Weighted, std.)			
	(1)	(2)	(3)	(4)
Covid Exposure$_P$ (std.)	0.0295***			
	(0.0101)			
Covid Blame$_P$ (std.)		0.0147*		
		(0.0077)		
Covid Optimism$_P$ (std.)			0.0188**	
			(0.0080)	
Covid Uncertainty$_P$ (std.)				0.0121
				(0.0080)
Covid Exposure$_Q$ (std.)	-0.0095			
	(0.0070)			
Covid Blame$_Q$ (std.)		-0.0005		
		(0.0050)		
Covid Optimism$_Q$ (std.)			-0.0082	
			(0.0055)	
Covid Uncertainty$_Q$ (std.)				-0.0057
				(0.0063)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,919	16,919	16,919	16,919
Within R-squared	0.008	0.007	0.008	0.007

Notes: Textual control variables include Tone$_{P,Q}$, Uncertainty Rate$_{P,Q}$, Complexity Rate$_{P,Q}$, Numbers$_{P,Q}$, Future Orientation$_{P,Q}$, Financial Orientation$_{P,Q}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.11: Market Reactions on Nonanswers & Covid Metrics: CARs *Control: Tone*

Dependent Variable:	CARs[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Weighted, std.)	-0.1323**	-0.1318**	-0.1328**	-0.1319**
	(0.0524)	(0.0522)	(0.0522)	(0.0522)
Nonanswers (Weighted, std.)	-0.0590			
×Covid Exposure$_P$ (std.)	(0.0417)			
Nonanswers (Weighted, std.)		-0.0168		
×Covid Blame$_P$ (std.)		(0.0410)		
Nonanswers (Weighted, std.)			-0.0905**	
×Covid Optimism$_P$ (std.)			(0.0452)	
Nonanswers (Weighted, std.)				-0.0414
×Covid Uncertainty$_P$ (std.)				(0.0392)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,704	16,704	16,704	16,704
R-squared	0.031	0.031	0.031	0.031

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.12: Market Reactions on Nonanswers & Covid Metrics: CARs & Firm FE *Control: Tone*

Dependent Variable:	CARs[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Weighted, std.)	-0.1288**	-0.1283**	-0.1300**	-0.1282**
	(0.0637)	(0.0637)	(0.0639)	(0.0637)
Nonanswers (Weighted, std.)	-0.0821*			
×Covid Exposure$_P$ (std.)	(0.0437)			
Nonanswers (Weighted, std.)		-0.0329		
×Covid Blame$_P$ (std.)		(0.0435)		
Nonanswers (Weighted, std.)			-0.1013**	
×Covid Optimism$_P$ (std.)			(0.0481)	
Nonanswers (Weighted, std.)				-0.0428
×Covid Uncertainty$_P$ (std.)				(0.0393)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,681	16,681	16,681	16,681
Within R-squared	0.036	0.036	0.036	0.036

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

D.7 Alternative Measure of Market Reactions on Nonanswers & Covid Metrics - Market-Adjusted Returns

Table D.13: Market Reactions on Nonanswers & Covid Metrics: Market-Adjusted Returns

Dependent Variable:	Market-Adjusted Returns[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Weighted, std.)	-0.1474***	-0.1470***	-0.1488***	-0.1473***
	(0.0530)	(0.0528)	(0.0528)	(0.0528)
Nonanswers (Weighted, std.)	-0.0618			
×Covid Exposure$_P$ (std.)	(0.0419)			
Nonanswers (Weighted, std.)		-0.0245		
×Covid Blame$_P$ (std.)		(0.0423)		
Nonanswers (Weighted, std.)			-0.0808*	
×Covid Optimism$_P$ (std.)			(0.0451)	
Nonanswers (Weighted, std.)				-0.0457
×Covid Uncertainty$_P$ (std.)				(0.0399)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Index FE	yes	yes	yes	yes
Observations	16,977	16,977	16,977	16,977
R-squared	0.029	0.029	0.029	0.029

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table D.14: Market Reactions on Nonanswers & Covid Metrics: Market-Adjusted Returns & Firm FE

Dependent Variable:	Market-Adjusted Returns[-1;1]			
	(1)	(2)	(3)	(4)
Nonanswers (Weighted, std.)	-0.1484**	-0.1478**	-0.1493**	-0.1478**
	(0.0629)	(0.0628)	(0.0630)	(0.0628)
Nonanswers (Weighted, std.)	-0.0873**			
×Covid Exposure$_P$ (std.)	(0.0432)			
Nonanswers (Weighted, std.)		-0.0429		
×Covid Blame$_P$ (std.)		(0.0437)		
Nonanswers (Weighted, std.)			-0.0913*	
×Covid Optimism$_P$ (std.)			(0.0479)	
Nonanswers (Weighted, std.)				-0.0521
×Covid Uncertainty$_P$ (std.)				(0.0405)
Textual Controls	yes	yes	yes	yes
Financial Controls	yes	yes	yes	yes
Quarter × Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Observations	16,958	16,958	16,958	16,958
Within R-squared	0.030	0.030	0.030	0.030

Notes: Corresponding Covid Metric$_P$ main term also included in the model (Covid Exposure$_P$, Covid Blame$_P$, Covid Optimism$_P$, Covid Uncertainty$_P$, respectively). Textual control variables include Negativity Rate$_{P,A}$, Uncertainty Rate$_{P,A}$, Complexity Rate$_{P,A}$, Numbers$_{P,A}$, Future Orientation$_{P,A}$, Financial Orientation$_{P,A}$. All textual controls are standardized. Financial controls include ln(Assets), Book-To-Market Ratio, Tobin's Q, Return on Assets, Revenue Growth and Earnings Surprise. Constant not displayed. Standard errors are clustered on firm level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6

Conclusion

In conclusion, this thesis demonstrates that organizational effectiveness and resilience in complex and dynamic environments depend on the careful design and implementation of management practices, incentive systems, and communication strategies. Structured management practices enhance employee engagement across a wide range of organizational settings. Still, the exact strength of the positive effect depends on firm-specific contingencies such as industry or size. In family firms, informal cultural mechanisms can partly compensate for fewer formal management practices. However, this works best when family involvement is primarily in ownership rather than management. Additionally, heterogeneity across skill levels and supervisory positions needs to be considered. At the executive level, ESG-linked incentives vary in their binding power and prevalence, with stronger, performance-sensitive metrics supporting alignment and credibility in sustainability commitments. In practice, however, these incentives are currently applied in a largely symbolic manner and rarely translate into meaningful performance-based exposure to ESG outcomes. Finally, in crises such as the Covid pandemic, managerial communication in conference calls is also affected. Specifically, greater crisis exposure is associated with a rise in managerial nonanswers. Yet markets continue to penalize such nonanswers, and seem to do so even more severely when companies have a rather optimistic outlook on the crisis, highlighting the need for clear communication in times of uncertainty.

Taken together, the results highlight that effective management requires a fine balance between formal systems, informal organizational mechanisms, and transparent communication, with the relative importance of each shaped by contextual factors such as firm size, industry, ownership, and external environmental pressures. By integrating evidence across multiple organizational levels and using diverse, complementary datasets, this thesis not only advances academic understanding of management, governance, and sustainability practices but also offers actionable insights for managers seeking to foster engagement and credibility in the face of evolving internal and external pressures.

References

Abdel-Kader, M. and R. Luther (2008). The impact of firm characteristics on management accounting practices: A uk-based empirical analysis. *The British Accounting Review 40*(1), 2–27.

Abernethy, M. A., H. C. Dekker, and J. Grafton (2021). The influence of performance measurement on the processual dynamics of strategic change. *Management Science 67*(1), 640–659.

Abeysekera, A. P. and C. S. Fernando (2020). Corporate social responsibility versus corporate shareholder responsibility: A family firm perspective. *Journal of Corporate Finance 61*, 101370.

Afonso Alves, C., A. P. Matias Gama, and M. Augusto (2021). Family influence and firm performance: the mediating role of stewardship. *Journal of Small Business and Enterprise Development 28*(2), 185–204.

Agarwal, V., J.-P. Gómez, K. Hosseini, and M. Jha (2025). Esg metrics in executive compensation: A multitasking approach. Technical report, CFR Working Paper No. 25-10.

Aguinis, H., R. K. Gottfredson, and H. Joo (2013). Avoiding a "me" versus "we" dilemma: Using performance management to turn teams into a source of competitive advantage. *Business Horizons 56*(4), 503–512.

Ahir, H., N. Bloom, and D. Furceri (2022). The world uncertainty index. *National Bureau of Economic Research Working Paper Series*.

Al-Shaer, H., K. Albitar, and J. Liu (2023). Ceo power and csr-linked compensation for corporate environmental responsibility: Uk evidence. *Review of Quantitative Finance and Accounting 60*(3), 1025–1063.

Al-Shaer, H. and M. Zaman (2019). Ceo compensation and sustainability reporting assurance: Evidence from the uk. *Journal of Business Ethics 158*, 233–252.

Allee, K. D. and M. D. DeAngelis (2015). The structure of voluntary disclosure narratives: Evidence from tone dispersion. *Journal of Accounting Research 53*(2), 241–274.

Allee, K. D., C. Do, and H. Do (2023). Contextualization of numbers in earnings conference calls. *Available at SSRN 3867023*.

Allee, K. D., C. Do, and H. Do (2024). Can analysts elicit useful information by asking unique questions in earnings conference calls? *Available at SSRN 4763812*.

Altig, D., S. Baker, J. M. Barrero, N. Bloom, P. Bunn, S. Chen, S. J. Davis, J. Leather, B. Meyer, E. Mihaylov, et al. (2020). Economic uncertainty before and during the covid-19 pandemic. *Journal of Public Economics 191*, 104274.

Aminadav, G. and E. Papaioannou (2020). Corporate control around the world. *The Journal of Finance 75*(3), 1191–1246.

Amore, M. D., M. Bennedsen, I. Le Breton-Miller, and D. Miller (2021). Back to the future: The effect of returning family successions on firm performance. *Strategic Management Journal 42*(8), 1432–1458.

Amore, M. D., V. D'Angelo, L. Breton-Miller, D. Miller, V. Pelucco, M. v. Essen, et al. (2024). Using artificial intelligence to measure the family control of companies. *Available at SSRN 4683794*.

Amore, M. D., A. Minichilli, and G. Corbetta (2011). How do managerial successions shape corporate financial policies in family firms? *Journal of Corporate Finance 17*(4), 1016–1027.

Amore, M. D., V. Pelucco, and F. Quarato (2022). Family ownership during the covid-19 pandemic. *Journal of Banking & Finance 135*, 106385.

Anderson, R. C. and D. M. Reeb (2003). Founding-family ownership and firm performance: evidence from the s&p 500. *The Journal of Finance 58*(3), 1301–1328.

Anderson, S. W., H. C. Dekker, and K. L. Sedatole (2010). An empirical examination of goals and performance-to-goal following the introduction of an incentive bonus plan with participative goal setting. *Management Science 56*(1), 90–109.

Aresu, S., R. Hooghiemstra, and A. Melis (2023). Integration of csr criteria into executive compensation contracts: A cross-country analysis. *Journal of Management 49*(8), 2766–2804.

Armstrong, M. (2022). *Armstrong's Handbook of Performance Management: An Evidence-Based Guide to Performance Leadership*. Kogan Page Publishers.

Arnold, M. and M. Artz (2015). Target difficulty, target flexibility, and firm performance: Evidence from business units' targets. *Accounting, Organizations and Society 40*, 61–77.

Arnold, M. and M. Artz (2019). The use of a single budget or separate budgets for planning and performance evaluation. *Accounting, Organizations and Society 73*, 50–67.

Artz, B., D. G. Blanchflower, and A. Bryson (2022). Unions increase job satisfaction in the united states. *Journal of Economic Behavior & Organization 203*, 173–188.

Asmuß, B. (2008). Performance appraisal interviews: Preference organization in assessment sequences. *The Journal of Business Communication 45*(4), 408–429.

Badawi, A. B. and R. P. Bartlett (2024). Esg overperformance? assessing the use of esg targets in executive compensation plans. *Available at SSRN 4941016*.

Bai, J. J., N. M. Boyson, Y. Cao, M. Liu, and C. Wan (2023). Executives vs. chatbots: Unmasking insights through human-ai differences in earnings conference q&a. *Northeastern U. D'Amore-McKim School of Business Research Paper* (4480056).

Baik, B., A. G. Kim, D. S. Kim, and S. Yoon (2024). Vocal delivery quality in earnings conference calls. *Journal of Accounting and Economics*, 101763.

Bakker, A. B. (2017). Strategic and proactive approaches to work engagement. *Organizational Dynamics 46*(2), 67–75.

Bakker, A. B. and E. Demerouti (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology 22*(3), 309–328.

Bakker, A. B. and M. P. Leiter (2010). *Work Engagement: A Handbook of Essential Theory and Research*. Taylor & Francis.

Bamber, M. and P.-L. Nappert (2025). Can we explain managerial non-answers during conference call q&as? *Contemporary Accounting Research 42*(2), 1079–1105.

Bandiera, O., G. Fischer, A. Prat, and E. Ytsma (2021). Do women respond less to performance pay? building evidence from multiple experiments. *American Economic Review: Insights 3*(4), 435–454.

Bandiera, O., R. Lemos, A. Prat, and R. Sadun (2018). Managing the family firm: Evidence from ceos at work. *The Review of Financial Studies 31*(5), 1605–1653.

Barcellos, L. P. (2025). Overcoming deceptive evasions in earnings calls: The role of investor suspicion. *Management Science - Articles In Advance*.

Barontini, R. and J. G. Hill (2024). Sustainability and executive compensation. In K. Alexander, M. Gargantini, and M. Siri (Eds.), *The Cambridge Handbook of EU Sustainable Finance: Regulation, Supervision, and Governance*, Cambridge Law Handbooks. Cambridge University Press.

Barth, A., S. Mansouri, and F. Woebbeking (2023). "let me get back to you"—a machine learning approach to measuring nonanswers. *Management Science 69*(10), 6333–6348.

Bassanini, A., T. Breda, E. Caroli, and A. Rebérioux (2013). Working in family firms: Paid less but more secure? evidence from french matched employer-employee data. *ILR Review 66*(2), 433–466.

Baum, C. F. and M. E. Schaffer (2012, November). IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments. Statistical Software Components, Boston College Department of Economics.

Bebchuk, L. A. and R. Tallarita (2022). The perils and questionable promise of esg-based compensation. *Journal of Corporation Law 48*, 37.

Beckmann, L., H. Beckmeyer, I. Filippou, S. Menze, and G. Zhou (2024). Unusual financial communication-evidence from chatgpt, earnings calls, and the stock market. *Olin Business School Center for Finance & Accounting Research Paper No. 2024/02*.

Bedford, D. S., T. Malmi, and M. Sandelin (2016). Management control effectiveness and strategy: An empirical analysis of packages and systems. *Accounting, Organizations and Society 51*, 12–28.

Bender, S., N. Bloom, D. Card, J. Van Reenen, and S. Wolter (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics 36*(S1), S371–S409.

Bennedsen, M., K. M. Nielsen, F. Pérez-González, and D. Wolfenzon (2007). Inside the family firm: The role of families in succession decisions and performance. *The Quarterly Journal of Economics 122*(2), 647–691.

Bennedsen, M., M. Tsoutsoura, and D. Wolfenzon (2019). Drivers of effort: Evidence from employee absenteeism. *Journal of Financial Economics 133*(3), 658–684.

Berg, F., K. Fabisik, and Z. Sautner (2020). Rewriting history ii: The (un) predictable past of esg ratings. *European Corporate Governance Institute–Finance Working Paper 708*(2020), 10–2139.

Berger, J., C. Harbring, and D. Sliwka (2013). Performance appraisals and the impact of forced distribution - an experimental investigation. *Management Science 59*(1), 54–68.

Berrone, P., C. Cruz, L. R. Gomez-Mejia, and M. Larraza-Kintana (2010). Socioemotional wealth and corporate responses to institutional pressures: Do family-controlled firms pollute less? *Administrative Science Quarterly 55*(1), 82–113.

Bertrand, M., S. Johnson, K. Samphantharak, and A. Schoar (2008). Mixing family with business: A study of thai business groups and the families behind them. *Journal of Financial Economics 88*(3), 466–498.

Birnberg, J. G., J. Luft, and M. D. Shields (2006). Psychology theory in management accounting research. *Handbooks of Management Accounting Research 1*, 113–135.

Bjuggren, C. M. (2015). Sensitivity to shocks and implicit employment protection in family firms. *Journal of Economic Behavior & Organization 119*, 18–31.

Block, J. H., C. Hansen, and H. Steinmetz (2023). Are family firms doing more innovation output with less innovation input? a replication and extension. *Entrepreneurship Theory and Practice 47*(4), 1496–1520.

Block, J. H. and M. Wagner (2014). The effect of family ownership on different dimensions of corporate social responsibility: Evidence from large us firms. *Business Strategy and the Environment 23*(7), 475–492.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2019). What drives differences in management practices? *American Economic Review 109*(5), 1648–83.

Bloom, N., C. Genakos, R. Sadun, and J. Van Reenen (2012). Management practices across firms and countries. *Academy of Management Perspectives 26*(1), 12–33.

Bloom, N., T. Kretschmer, and J. Van Reenen (2011). Are family-friendly workplace practices a valuable firm resource? *Strategic Management Journal 32*(4), 343–367.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics 122*(4), 1351–1408.

Bloom, N. and J. Van Reenen (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives 24*(1), 203–224.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review 86*(5), 1549–1575.

Bonham, J. and A. Riggs-Cragun (2024). Motivating esg activities through contracts, taxes and disclosure regulation. *Chicago Booth Research Paper* (22-05).

Bormann, K. C., S. Backs, and C. Hoon (2021). What makes nonfamily employees act as good stewards? emotions and the moderating roles of stewardship culture and gender roles in family firms. *Family Business Review 34*(3), 251–269.

Boselie, P., G. Dietz, and C. Boon (2005). Commonalities and contradictions in hrm and performance research. *Human Resource Management Journal 15*(3), 67–94.

Boswell, W. R. and J. W. Boudreau (2002). Separating the developmental and evaluative performance appraisal uses. *Journal of Business and Psychology 16*(3), 391–412.

Bowen, R. M., A. K. Davis, and D. A. Matsumoto (2002). Do conference calls affect analysts' forecasts? *The Accounting Review 77*(2), 285–316.

Bozanic, Z., D. T. Roulstone, and A. Van Buskirk (2018). Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics 65*(1), 1–20.

Brickley, J., C. Smith, and J. Zimmerman (2017). *Managerial Economics and Organizational Architecture*. McGraw-Hill Education.

Broadbent, J. and R. Laughlin (2009). Performance management systems: A conceptual model. *Management Accounting Research 20*(4), 283–295.

Brochet, F., K. Kolev, and A. Lerman (2018). Information transfer and conference calls. *Review of Accounting Studies 23*, 907–957.

Broszeit, S., M.-C. Laible, U. Fritsch, and H. Görg (2019). Management practices and productivity in germany. *German Economic Review 20*(4), e657–e705.

Budworth, M.-H., G. P. Latham, and L. Manroop (2015). Looking forward to performance improvement: A field test of the feedforward interview for performance management. *Human Resource Management 54*(1), 45–54.

Burkart, M., F. Panunzi, and A. Shleifer (2003). Family firms. *The Journal of Finance 58*(5), 2167–2201.

Burke, R. J., W. Weitzel, and T. Weir (1978). Characteristics of effective employee performance review and development interviews: replication and extension 1. *Personnel Psychology 31*(4), 903–919.

Bushee, B. J., I. D. Gow, and D. J. Taylor (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research 56*(1), 85–121.

Cabeza-García, L., M. Sacristán-Navarro, and S. Gómez-Ansón (2017). Family involvement and corporate social responsibility disclosure. *Journal of Family Business Strategy 8*(2), 109–122.

Call, A. C., R. W. Flam, J. A. Lee, and N. Y. Sharp (2024). Managers' use of humor on public earnings conference calls. *Review of Accounting Studies 29*(3), 2650–2687.

Call, A. C., N. Y. Sharp, and T. D. Shohfi (2021). Which buy-side institutions participate in public earnings conference calls? implications for capital markets and sell-side coverage. *Journal of Corporate Finance 68*, 101964.

Campbell, J. L., X. Zheng, and D. Zhou (2021). Number of numbers: Does quantitative textual disclosure reduce information risk? *Available at SSRN 3775905*.

Canavati, S. (2018). Corporate social performance in family firms: a meta-analysis. *Journal of Family Business Management 8*(3), 235–273.

Carter, M. E., A. Pawliczek, and R. I. Zhong (2023). Say on esg: The adoption of say-on-pay laws, esg contracting, and firm esg performance. *European Corporate Governance Institute–Finance Working Paper* (886).

Casas-Arce, P., S. M. Lourenço, and F. A. Martínez-Jerez (2017). The performance effect of feedback frequency and detail: Evidence from a field experiment in customer satisfaction. *Journal of Accounting Research 55*(5), 1051–1088.

Casas-Arce, P. and F. A. Martínez-Jerez (2022). Leader effects in competition among teams: Evidence from a field intervention. *The Accounting Review 97*(1), 99–122.

Cawley, B. D., L. M. Keeping, and P. E. Levy (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology 83*(4), 615.

Cederblom, D. (1982). The performance appraisal interview: A review, implications, and suggestions. *Academy of Management Review 7*(2), 219–227.

Chen, J., E. Demers, and B. Lev (2018). Oh what a beautiful morning! diurnal influences on executives and analysts: Evidence from conference calls. *Management Science 64*(12), 5899–5924.

Cheng, M. M., P. F. Luckett, and H. Mahama (2007). Effect of perceived conflict among multiple performance goals and goal difficulty on task performance. *Accounting & Finance 47*(2), 221–242.

Chrisman, J. J. (2019). Stewardship theory: Realism, relevance, and family firm governance. *Entrepreneurship Theory and Practice 43*(6), 1051–1066.

Christensen-Salem, A., L. F. Mesquita, M. Hashimoto, P. W. Hom, and L. R. Gomez-Mejia (2021). Family firms are indeed better places to work than non-family firms! socioemotional wealth and employees' perceived organizational caring. *Journal of Family Business Strategy 12*(1), 100412.

Christian, M. S., A. S. Garza, and J. E. Slaughter (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology 64*(1), 89–136.

Clark, A. E., C. D'ambrosio, and R. Zhu (2021). Job quality and workplace gender diversity in europe. *Journal of Economic Behavior & Organization 183*, 420–432.

Cohen, S., I. Kadach, G. Ormazabal, and S. Reichelstein (2023). Executive compensation tied to esg performance: International evidence. *Journal of Accounting Research 61*(3), 805–853.

Cohn, J. B., Z. Liu, and M. I. Wardlaw (2022). Count (and count-like) data in finance. *Journal of Financial Economics 146*(2), 529–551.

Corbetta, G. and C. Salvato (2004). Self–serving or self–actualizing? models of man and agency costs in different types of family firms: A commentary on "comparing the agency costs of family and non–family firms: Conceptual issues and exploratory evidence". *Entrepreneurship Theory and Practice 28*(4), 355–362.

Crawford Camiciottoli, B. (2018). Persuasion in earnings calls: A diachronic pragmalinguistic analysis. *International Journal of Business Communication 55*(3), 275–292.

Creemers, S., L. Peeters, J. L. Q. Castillo, M. Vancauteren, and W. Voordeckers (2023). Family firms and the labor productivity controversy: A distributional analysis of varying labor productivity gaps. *Journal of Family Business Strategy 14*(2), 100515.

Cruz, C., S. Firfiray, and L. R. Gomez-Mejia (2011). Socioemotional wealth and human resource management (hrm) in family-controlled firms. In *Research in Personnel and Human Resources Management*, pp. 159–217. Emerald Group Publishing Limited.

Cruz, C., M. Larraza-Kintana, L. Garcés-Galdeano, and P. Berrone (2014). Are family firms really more socially responsible? *Entrepreneurship Theory and Practice 38*(6), 1295–1316.

Cucculelli, M., L. Mannarino, V. Pupo, and F. Ricotta (2014). Owner-management, firm age, and productivity in italian family firms. *Journal of Small Business Management 52*(2), 325–343.

Cummings, L. L. and D. P. Schwab (1978). Designing appraisal systems for information yield. *California Management Review 20*(4), 18–25.

Dai, L., P. Gong, A. B. Jackson, and Z. R. Peng (2022). Controlling the narrative: Managers' topic-shifting behavior in conference calls. *Available at SSRN 4024646*.

Dal Maso, L., R. Basco, T. Bassetti, and N. Lattanzi (2020). Family ownership and environmental performance: The mediation effect of human resource practices. *Business Strategy and the Environment 29*(3), 1548–1562.

Davis, A. K., W. Ge, D. Matsumoto, and J. L. Zhang (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies 20*, 639–673.

Davis, J. H., M. R. Allen, and H. D. Hayes (2010). Is blood thicker than water? a study of stewardship perceptions in family business. *Entrepreneurship Theory and Practice 34*(6), 1093–1116.

Davis, J. H., F. D. Schoorman, and L. Donaldson (1997). Toward a stewardship theory of management. *Academy of Management Review*, 20–47.

De Amicis, C., S. Falconieri, and M. Tastan (2021). Sentiment analysis and gender differences in earnings conference calls. *Journal of Corporate Finance 71*, 101809.

De Franco, G., T. Shohfi, D. Xu, and Z. V. Zhu (2023). Fixed income conference calls. *Journal of Accounting and Economics 75*(1), 101518.

De Kok, J. M., L. M. Uhlaner, and A. R. Thurik (2006). Professional hrm practices in family owned-managed enterprises. *Journal of Small Business Management 44*(3), 441–460.

De Kok, T. (2025). Chatgpt for textual analysis? how to use generative llms in accounting research. *Management Science 71*(9), 7888–7906.

Dekker, H. C., T. Groot, and M. Schoute (2012). Determining performance targets. *Behavioral Research in Accounting 24*(2), 21–46.

Demartini, M. C. and D. Otley (2020). Beyond the system vs. package dualism in performance management systems design: A loose coupling approach. *Accounting, Organizations and Society 86*, 101072.

Derfuss, K. (2009). The relationship of budgetary participation and reliance on accounting performance measures with individual-level consequent variables: A meta-analysis. *European Accounting Review 18*(2), 203–239.

Derrien, F., P. Krueger, A. Landier, and T. Yao (2025). Esg news, future cash flows, and firm value. *The Journal of Finance 80*(6), 3499–3554.

Dube, S., H. S. G. Lee, and D. Wang (2026). Do consumers vote with their feet in response to negative esg news? evidence from foot traffic to retail locations. *Journal of Accounting Research*.

Duchin, R., J. Gao, and Q. Xu (2025). Sustainability or greenwashing: Evidence from the asset market for industrial pollution. *The Journal of Finance 80*(2), 699–754.

Duran, P., N. Kammerlander, M. Van Essen, and T. Zellweger (2016). Doing more with less: Innovation input and output in family firms. *Academy of Management Journal 59*(4), 1224–1264.

Dyer, W. G. and D. A. Whetten (2006). Family firms and social responsibility: Preliminary evidence from the s&p 500. *Entrepreneurship Theory and Practice 30*(6), 785–802.

Edmans, A. (2024). Rational sustainability. *Journal of Applied Corporate Finance 36*(2), 8–15.

Edmans, A., X. Gabaix, and D. Jenter (2017). Executive compensation: A survey of theory and evidence. In B. E. Hermalin and M. S. Weisbach (Eds.), *The Handbook of the Economics of Corporate Governance*, Volume 1 of *The Handbook of the Economics of Corporate Governance*, pp. 383–539. North-Holland.

Edmans, A., T. Gosling, and D. Jenter (2023). Ceo compensation: Evidence from the field. *Journal of Financial Economics 150*(3), 103718.

Efing, M., S. Ehmann, P. Kampkötter, and R. Moritz (2026). Esg incentives in executive compensation. *Available at SSRN 4974204*.

Ehmann, S. (2026). Communicating through crisis - nonanswers in earnings calls during the covid pandemic. *Working Paper*.

Ehmann, S., P. Kampkötter, P. Maier, and P. Yang (2024). Performance management and work engagement–new evidence using longitudinal data. *Management Accounting Research 64*, 100867.

Ehmann, S., P. Kampkötter, J. Wenzel, and S. Wolter (2025). In the hand of the family: Management practices and perceived job quality. *Journal of Economic Behavior & Organization 237*, 107135.

El Ghoul, S., O. Guedhami, H. Wang, and C. C. Kwok (2016). Family control and corporate social responsibility. *Journal of Banking & Finance 73*, 131–146.

Ellguth, P., S. Kohaut, and I. Möller (2014). The iab establishment panel—methodological essentials and data quality. *Journal for Labour Market Research 47*(1-2), 27–41.

Ellul, A., M. Pagano, and F. Schivardi (2018). Employment and wage insurance within firms: Worldwide evidence. *The Review of Financial Studies 31*(4), 1298–1340.

Englmaier, F., G. Muehlheusser, A. Roider, and N. Wallmeier (2022). Management and performance in the public sector: Evidence from german municipalities. *CESifo Working Paper*.

Falconieri, S. and C. De Amicis (2023). Managerial diversity and corporate communication in periods of crisis. *Review of Corporate Finance 3*(1–2), 213–244.

Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science 326*(5952), 535–538.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and M. C. Jensen (1983). Separation of ownership and control. *The Journal of Law and Economics 26*(2), 301–325.

Fang, D.-J., H.-H. Chang, S.-K. Lin, and C. Chen (2024). How do corporates respond to economic recessions? evidence from form 10-k, 10-q, and earnings call transcripts. *Available at SSRN 4961716*.

Feichter, C. (2023). The effect of supervisors' prior task performance on employees' targets. *The Accounting Review 98*(1), 191–214.

Feichter, C., I. Grabner, and F. Moers (2018). Target setting in multi-divisional firms: State of the art and avenues for future research. *Journal of Management Accounting Research 30*(3), 29–54.

Ferreira, A. and D. Otley (2009). The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research 20*(4), 263–282.

Ferris, K. R. (1977). A test of the expectancy theory of motivation in an accounting environment. *The Accounting Review 52*(3), 605–615.

Firfiray, S., C. Cruz, I. Neacsu, and L. R. Gomez-Mejia (2018). Is nepotism so bad for family firms? a socioemotional wealth approach. *Human Resource Management Review 28*(1), 83–97.

Fischer, G., F. Janik, D. Müller, and A. Schmucker (2009). The iab establishment panel–things users should know. *Journal of Contextual Economics–Schmollers Jahrbuch 129*(1), 133–148.

Flammer, C., B. Hong, and D. Minor (2019). Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: Effectiveness and implications for firm outcomes. *Strategic Management Journal 40*(7), 1097–1122.

Folger, R., M. A. Konovsky, and R. Cropanzano (1992). A due process metaphor for performance appraisal. *Research in Organizational Behavior 14*, 129–129.

Franco-Santos, M., L. Lucianetti, and M. Bourne (2012). Contemporary performance measurement systems: A review of their consequences and a framework for research. *Management Accounting Research 23*(2), 79–119.

Frankel, R., J. Jennings, and J. Lee (2022). Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science 68*(7), 5514–5532.

Franks, J., C. Mayer, P. Volpin, and H. F. Wagner (2012). The life cycle of family ownership: International evidence. *The Review of Financial Studies 25*(6), 1675–1712.

Frederiksen, A., F. Lange, and B. Kriechel (2017). Subjective performance evaluations and employee careers. *Journal of Economic Behavior & Organization 134*, 408–429.

Freeman, R. B. (1978). Job satisfaction as an economic variable. *American Economic Review 68*(2), 135–141.

Fritsch, F., Q. Zhang, and X. Zheng (2025). Responding to climate change crises: Firms' trade-offs. *Journal of Accounting Research 63*(5), 2137–2179.

Fuller, K., J. Netter, and M. Stegemoller (2002). What do returns to acquiring firms tell us? evidence from firms that make many acquisitions. *The Journal of Finance 57*(4), 1763–1793.

Gaertner, S. (1999). Structural determinants of job satisfaction and organizational commitment in turnover models. *Human Resource Management Review 9*(4), 479–493.

Gantchev, N., M. Giannetti, and M. Hober (2025). Beyond esg: Executive pay metrics and shareholder support. *FEB-RN Research Paper* (26).

Gantchev, N., M. Giannetti, and R. Li (2022). Does money talk? divestitures and corporate environmental and social policies. *Review of Finance 26*(6), 1469–1508.

Ganzach, Y. (1998). Intelligence and job satisfaction. *Academy of Management Journal 41*(5), 526–539.

García-Sánchez, I.-M., J. Martín-Moreno, S. A. Khan, and N. Hussain (2021). Socio-emotional wealth and corporate responses to environmental hostility: Are family firms more stakeholder oriented? *Business Strategy and the Environment 30*(2), 1003–1018.

Garel, A., A. Romec, Z. Sautner, and A. F. Wagner (2024). Do investors care about biodiversity? *Review of Finance 28*(4), 1151–1186.

Geelen, T., J. Hajda, and J. Starmans (2025). Sustainable organizations. *European Corporate Governance Institute–Finance Working Paper No. 994/2024*.

Gibbons, R. and K. J. Murphy (1990). Relative performance evaluation for chief executive officers. *ILR Review 43*(3), 30–S.

Gibbs, M., K. A. Merchant, W. A. V. d. Stede, and M. E. Vargus (2004). Determinants and effects of subjectivity in incentives. *The Accounting Review 79*(2), 409–436.

Gómez-Mejía, L. R., K. T. Haynes, M. Núñez-Nickel, K. J. Jacobson, and J. Moyano-Fuentes (2007). Socioemotional wealth and business risks in family-controlled firms: Evidence from spanish olive oil mills. *Administrative Science Quarterly 52*(1), 106–137.

Gordon, M. E. and L. P. Stewart (2009). Conversing about performance: Discursive resources for the appraisal interview. *Management Communication Quarterly 22*(3), 473–501.

Gow, I. D., D. F. Larcker, and A. A. Zakolyukina (2021). Non-answers during conference calls. *Journal of Accounting Research 59*(4), 1349–1384.

Grabner, I. and F. Moers (2013). Management control as a system or a package? conceptual and empirical issues. *Accounting, Organizations and Society 38*(6), 407–419.

Grabner, I., A. Renders, and L. Yang (2024). The complementarity between csr disclosures and the use of csr-based performance measures in ceo annual incentive contracts. *Available at SSRN 3670738*.

Greenberg, J. (1987). A taxonomy of organizational justice theories. *Academy of Management Review 12*(1), 9–22.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management 16*(2), 399–432.

Groen, B. A., M. J. Wouters, and C. P. Wilderom (2017). Employee participation, performance metrics, and job performance: A survey study based on self-determination theory. *Management Accounting Research 36*, 51–66.

Gruman, J. A. and A. M. Saks (2011). Performance management and employee engagement. *Human Resource Management Review 21*(2), 123–136.

Grund, C. and D. Sliwka (2009). The anatomy of performance appraisals in germany. *The International Journal of Human Resource Management 20*(10), 2049–2065.

Guenther, T. W. and A. Heinicke (2019). Relationships among types of use, levels of sophistication, and organizational outcomes of performance measurement systems: The crucial role of design choices. *Management Accounting Research 42*, 1–25.

Halbesleben, J. R. (2010). A meta-analysis of work engagement: Relationships with burnout, demands, resources, and consequences. *Work engagement: A Handbook of Essential Theory and Research 8*(1), 102–117.

Hale, T., N. Angrist, R. Goldszmidt, et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 government response tracker). *Nature Human Behaviour 5*(4), 529–538.

Hall, M. (2008). The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance. *Accounting, Organizations and Society 33*(2-3), 141–163.

Hamermesh, D. S. (1977). Economic aspects of job satisfaction. *Essays in Labor Market Analysis*, 53–72.

Harter, J. K., F. L. Schmidt, and T. L. Hayes (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology 87*(2), 268–279.

Hartmann, F. and S. Slapničar (2009). How formal performance evaluation affects trust between superior and subordinate managers. *Accounting, Organizations and Society 34*(6-7), 722–737.

Hartmann, F. and S. Slapničar (2012). The perceived fairness of performance evaluation: The role of uncertainty. *Management Accounting Research 23*(1), 17–33.

Hassan, T. A., S. Hollander, L. Van Lent, M. Schwedeler, and A. Tahoun (2023). Firm-level exposure to epidemic diseases: Covid-19, sars, and h1n1. *The Review of Financial Studies 36*(12), 4919–4964.

Hauff, S., J. Felfe, and K. Klug (2022). High-performance work practices, employee well-being, and supportive leadership: spillover mechanisms and boundary conditions between hrm and leadership behavior. *The International Journal of Human Resource Management 33*(10), 2109–2137.

Haylock, M. and P. Kampkötter (2019). Dataset: The linked personnel panel (lpp). *Data in Brief 27*, 104824.

Hazarika, S., A. Kashikar, L. Peng, A. Röell, and Y. Shen (2025). Esg-linked pay around the world—trends, determinants, and outcomes. *Working Paper*.

Hernández-Linares, R., M. C. López-Fernández, E. Memili, F. Mullins, and P. C. Patel (2023). High-performance work practices, socioemotional wealth preservation, and family firm labor productivity. *Business Research Quarterly 26*(3), 237–255.

Hoffmann, C., T. Wulf, and S. Stubner (2016). Understanding the performance consequences of family involvement in the top management team: The role of long-term orientation. *International Small Business Journal 34*(3), 345–368.

Holbrook, R. L. (2002). Contact points and flash points: conceptualizing the use of justice mechanisms in the performance appraisal interview. *Human Resource Management Review 12*(1), 101–123.

Hollander, S., M. Pronk, and E. Roelofsen (2010). Does silence speak? an empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research 48*(3), 531–563.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 74–91.

Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics 13*(2), 324–340.

Holzhacker, M., S. Kramer, M. Matějka, and N. Hoffmeister (2019). Relative target setting and cooperation. *Journal of Accounting Research 57*(1), 211–239.

Hong, B., Z. Li, and D. Minor (2016). Corporate governance and executive compensation for corporate social responsibility. *Journal of Business Ethics 136*, 199–213.

Howe, L. C., L. M. Giurge, A. F. Wagner, and J. I. Menges (2024). Ceos showing humanity: human care statements in conference calls and stock market performance during a crisis. *Academy of Management Discoveries 10*(4), 589–610.

Huang, C., Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet 395*(10223), 497–506.

Huang, M., P. Li, F. Meschke, and J. P. Guthrie (2015). Family firms, employee satisfaction, and corporate performance. *Journal of Corporate Finance 34*, 108–127.

Ikram, A., Z. F. Li, and D. Minor (2023). Csr-contingent executive compensation contracts. *Journal of Banking & Finance 151*, 105655.

Impink, J., M. Lubberink, B. Van Praag, and D. Veenman (2012). Did accelerated filing requirements and sox section 404 affect the timeliness of 10-k filings? *Review of Accounting Studies 17*, 227–253.

Inceoglu, I., G. Thomas, C. Chu, D. Plans, and A. Gerbasi (2018). Leadership behavior and employee well-being: An integrated review and a future research agenda. *The Leadership Quarterly 29*(1), 179–202.

Indjejikian, R. J., M. Matějka, and J. D. Schloetzer (2014). Target ratcheting and incentives: Theory, evidence, and new opportunities. *The Accounting Review 89*(4), 1259–1267.

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

Ioannou, I., S. X. Li, and G. Serafeim (2016). The effect of target difficulty on target completion: The case of reducing carbon emissions. *The Accounting Review 91*(5), 1467–1492.

Ivancevich, J. M. (1982). Subordinates' reactions to performance appraisal interviews: A test of feedback and goal-setting techniques. *Journal of Applied Psychology 67*(5), 581.

Jacobs, G., F. D. Belschak, and D. N. Den Hartog (2014). (Un)ethical behavior and performance appraisal: The role of affect, support, and organizational justice. *Journal of Business Ethics 121*, 63–76.

Jensen, M. C. and W. H. Meckling (1979). Rights and production functions: An application to labor-managed firms and codetermination. *Journal of Business*, 469–506.

Jory, S., T. Ngo, and J. Susnjara (2025). The effect of ceo overconfidence on corporate disclosures amid a pervasive shock: Evidence from the covid-19 pandemic. *Journal of Business Finance & Accounting 52*(3), 1431–1462.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal 33*(4), 692–724.

Kale, A. and D. Kale (2023). Do exogenous economic crises change investors' response to earnings announcements?: A detailed review using the data from covid-19 pandemic. *The Journal of Economic Asymmetries 28*, e00330.

Kampkötter, P., K. Marggraf, and J.-H. Zimmermann (2017). Determinants and effects of formal target agreements: An empirical investigation of german firms. *Review of Managerial Science 11*(1), 1–18.

Kampkötter, P., J. Mohrenweiser, D. Sliwka, S. Steffes, and S. Wolter (2016). Measuring the use of human resources practices and employee attitudes: The linked personnel panel. In *Evidence-based HRM: a Global Forum for Empirical Scholarship*, Volume 4, pp. 94–115. Emerald Group Publishing Limited.

Kampkötter, P. and D. Sliwka (2016). The complementary use of experiments and field data to evaluate management practices: The case of subjective performance evaluations. *Journal of Institutional and Theoretical Economics 172*(2), 364–389.

Kang, J.-K. and J. Kim (2020). Do family firms invest more than nonfamily firms in employee-friendly policies? *Management Science 66*(3), 1300–1324.

Kay, E. and H. H. Meyer (1965). Effects of threat in a performance appraisal interview. *Journal of Applied Psychology 49*(5), 311.

Khan, M., G. Serafeim, and A. Yoon (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review 91*(6), 1697–1724.

Kim, T.-Y. and K. Leung (2007). Forming and reacting to overall fairness: A cross-cultural comparison. *Organizational Behavior and Human Decision Processes 104*(1), 83–95.

Kolk, A. and P. Perego (2014). Sustainable bonuses: sign of corporate responsibility or window dressing? *Journal of Business Ethics 119*(1), 1–15.

Kominis, G. and C. R. Emmanuel (2007). The expectancy–valence theory revisited: Developing an extended model of managerial motivation. *Management Accounting Research 18*(1), 49–75.

Kotey, B. and C. Folker (2007). Employee training in smes: effect of size and firm type — family and nonfamily. *Journal of Small Business Management 45*(2), 214–238.

Kreidl, F. and H. Scholz (2021). Exploiting the dividend month premium: evidence from germany. *Journal of Asset Management 22*(4), 253–266.

Kölling, A. (2020). Employment in family managed firms: Less but safe? analysing labour demand of german family managed firms with a treatment model for panel data. *Bulletin of Economic Research 72*(1), 14–32.

La Porta, R., F. Lopez-de Silanes, and A. Shleifer (1999). Corporate ownership around the world. *The Journal of Finance 54*(2), 471–517.

Lambert, R. A. (2006). Agency theory and management accounting. *Handbooks of Management Accounting Research 1*, 247–268.

Latham, G. P. and E. A. Locke (2007). New developments in and directions for goal-setting research. *European Psychologist 12*(4), 290.

Lau, C. M. (2011). Nonfinancial and financial performance measures: How do they affect employee role clarity and performance? *Advances in Accounting 27*(2), 286–293.

Lau, C. M. and A. Moser (2008). Behavioral effects of nonfinancial performance measures: The role of procedural fairness. *Behavioral Research in Accounting 20*(2), 55–71.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review 90*(5), 1346–361.

Le Breton-Miller, I. and D. Miller (2009). Agency vs. stewardship in public family firms: A social embeddedness reconciliation. *Entrepreneurship Theory and Practice 33*(6), 1169–1191.

Lee, J. (2016). Can investors detect managers' lack of spontaneity? adherence to predetermined scripts during earnings conference calls. *The Accounting Review 91*(1), 229–250.

Lemos, R. and D. Scur (2019). The ties that bind: implicit contracts and management practices in family-run firms. *CEPR Discussion Paper No. DP13794*.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics 30*(1), 67–80.

Li, S. X. and T. Sandino (2018). Effects of an information sharing system on employee creativity, engagement, and performance. *Journal of Accounting Research 56*(2), 713–747.

Liu, M., Y. Cao, and J. Flake (2024). The credibility of non-disclosure: Evidence from real-time market response to non-answers in conference calls. *Available at SSRN 4955634*.

Liu, X. and Y. Zhang (2015). Effects of target timing and contract frame on individual performance. *European Accounting Review 24*(2), 329–345.

Liu, X. K. and R. A. Leitch (2013). Performance effects of setting targets and pay–performance relations before or after operations. *Management Accounting Research 24*(1), 64–79.

Locke, E. A. and G. P. Latham (1990). *A theory of goal setting & task performance.* Prentice-Hall, Inc.

Locke, E. A. and G. P. Latham (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist 57*(9), 705.

Locke, E. A. and G. P. Latham (2019). The development of goal setting theory: A half century retrospective. *Motivation Science 5*(2), 93.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance 66*(1), 35–65.

Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research 54*(4), 1187–1230.

Loughran, T. and B. McDonald (2024). Measuring firm complexity. *Journal of Financial and Quantitative Analysis 59*(6), 2487–2514.

LSEG (2024). Environmental, social and governance scores from lseg. https://www.lseg.com/content/dam/data-analytics/en$_u$s/*documents/methodology/lseg* − *esg* − *scores* − *methodology.pdf*, *LastAccessed* : 2024 − 03 − 03.

Luft, J. (2016). Cooperation and competition among employees: Experimental evidence on the role of management control systems. *Management Accounting Research 31*, 75–85.

Lunenburg, F. C. (2011). Goal-setting theory of motivation. *International Journal of Management, Business, and Administration 15*(1), 1–6.

Maas, K. (2018). Do corporate social performance targets in executive compensation contribute to corporate social performance? *Journal of Business Ethics 148*(3), 573–585.

Maas, V. S. and B. Shi (2023). The effects of target difficulty and relative ability on managers' delegation decisions. *Management Accounting Research 60*, 100851.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature 35*(1), 13–39.

Maeckle, K. (2025). Short-termist carbon emissions. Working paper, University of Mannheim.

Malmi, T. and D. A. Brown (2008). Management control systems as a package - opportunities, challenges and research directions. *Management Accounting Research 19*(4), 287–300.

Manthei, K., D. Sliwka, and T. Vogelsang (2023). Information, Incentives, and Attention: A Field Experiment on the Interaction of Management Controls. *The Accounting Review 98*(5), 455–479.

Martin, R. and T. Thomas (2022). Target setting with compensation discretion: How are ex ante targets affected when superiors have ex post discretion? *Accounting, Organizations and Society 97*, 101295.

Maslar, D. A., M. Serfling, and S. Shaikh (2021). Economic downturns and the informativeness of management earnings forecasts. *Journal of Accounting Research 59*(4), 1481–1520.

Matsumoto, D., M. Pronk, and E. Roelofsen (2011). What makes conference calls useful? the information content of managers' presentations and analysts' discussion sessions. *The Accounting Review 86*(4), 1383–1414.

Matzler, K., V. Veider, J. Hautz, and C. Stadler (2015). The impact of family ownership, management, and governance on innovation. *Journal of Product Innovation Management 32*(3), 319–333.

Mayew, W. J., M. Sethuraman, and M. Venkatachalam (2020). Individual analysts' stock recommendations, earnings forecasts, and the informativeness of conference call question and answer sessions. *The Accounting Review 95*(6), 311–337.

Mayew, W. J., N. Y. Sharp, and M. Venkatachalam (2013). Using earnings conference calls to identify analysts with superior private information. *Review of Accounting Studies 18*, 386–413.

Mayew, W. J. and M. Venkatachalam (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance 67*(1), 1–43.

Meinecke, A. L., N. Lehmann-Willenbrock, and S. Kauffeld (2017). What happens during annual appraisal interviews? how leader–follower interactions unfold and impact interview outcomes. *Journal of Applied Psychology 102*(7), 1054.

Mercer (2016). International compensation disclosure standard (icds), hkp/// group. Description of the methodology for board compensation reporting.

Messner, M. (2016). Does industry matter? how industry context shapes management accounting practice. *Management Accounting Research 31*, 103–111.

Michaely, R., T. Schmid, and M. Wang (2024). Implicit versus explicit contracting in executive compensation for environmental and social performance. *ECGI Working Paper No 1016/2024*.

Micheli, P. and L. Mari (2014). The theory and practice of performance measurement. *Management Accounting Research 25*(2), 147–156.

Milgrom, P. and J. Roberts (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics 19*(2-3), 179–208.

Miller, D. and I. Le Breton-Miller (2006). Family governance and firm performance: Agency, stewardship, and capabilities. *Family Business Review 19*(1), 73–87.

Miller, D., I. Le Breton-Miller, and B. Scholnick (2008). Stewardship vs. stagnation: An empirical comparison of small family and non-family businesses. *Journal of Management Studies 45*(1), 51–78.

Miller, D., A. Minichilli, and G. Corbetta (2013). Is family leadership always beneficial? *Strategic Management Journal 34*(5), 553–571.

Miroshnychenko, I., D. Miller, A. De Massis, and I. Le Breton-Miller (2025). Are family firms green? *Small Business Economics 64*(2), 279–306.

Moorman, R. H., T. C. Darnold, and M. Priesemuth (2013). Perceived leader integrity: Supporting the construct validity and utility of a multi-dimensional measure in two samples. *The Leadership Quarterly 24*(3), 427–444.

Mueller, H. M. and T. Philippon (2011). Family firms and labor relations. *American Economic Journal: Macroeconomics 3*(2), 218–245.

Murphy, K. J. (1999). Executive compensation. *Handbook of Labor Economics 3*, 2485–2563.

Murphy, K. J. (2000). Performance standards in incentive contracts. *Journal of Accounting and Economics 30*(3), 245–278.

Neckebrouck, J., W. Schulze, and T. Zellweger (2018). Are family firms good employers? *Academy of Management Journal 61*(2), 553–585.

Noh, J. and D. Zhou (2022). Executives' blaming external factors and market reactions: Evidence from earnings conference calls. *Journal of Banking & Finance 134*, 106358.

Otley, D. (1980). The contingency theory of management accounting: Achievement and prognosis. *Accounting, Organizations and Society 5*(4), 413–428.

Otley, D. (1999). Performance management: A framework for management control systems research. *Management Accounting Research 10*(4), 363–382.

Otley, D. (2016). The contingency theory of management accounting and control: 1980–2014. *Management Accounting Research 31*, 45–62.

Pagano, M., E. Di Porto, F. Schivardi, R. Saggio, and V. Pezone (2024). Careers and wages in family firms: Evidence from matched employer-employee data. *European Corporate Governance Institute–Finance Working Paper*.

Patterson, M. G., M. A. West, V. J. Shackleton, J. F. Dawson, R. Lawthom, S. Maitlis, D. L. Robinson, and A. M. Wallace (2005). Validating the organizational climate measure: links to managerial practices, productivity and innovation. *Journal of Organizational Behavior 26*(4), 379–408.

Peláez-León, J. D. and G. Sánchez-Marín (2022). Socioemotional wealth and human resource policies: effects on family firm performance. *International Journal of Entrepreneurial Behavior & Research 28*(1), 109–135.

Peláez-León, J. D. and G. Sánchez-Marín (2023). High-performance work systems in family firms: A mixed gamble approach. *Journal of Business Research 156*, 113532.

Pichler, S. (2012). The social context of performance appraisal and appraisal reactions: A meta-analysis. *Human Resource Management 51*(5), 709–732.

Pichler, S., G. Beenen, and S. Wood (2020). Feedback frequency and appraisal reactions: a meta-analytic test of moderators. *The International Journal of Human Resource Management 31*(17), 2238–2263.

Pitkänen, H. and K. Lukka (2011). Three dimensions of formal and informal feedback in management accounting. *Management Accounting Research 22*(2), 125–137.

Pittino, D., F. Visintin, T. Lenger, and D. Sternad (2016). Are high performance work practices really necessary in family smes? an analysis of the impact on employee retention. *Journal of Family Business Strategy 7*(2), 75–89.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*(1), 7–63.

Presslee, A., G. Richins, S. Saiy, and A. Webb (2023). Small sample field study: The effects of team-based recognition on employee engagement and effort. *Management Accounting Research 59*, 100829.

Price, S. M., J. S. Doran, D. R. Peterson, and B. A. Bliss (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance 36*(4), 992–1011.

PricewaterhouseCoopers (2021). Paying well by paying for good. www.pwc.co.uk/human-resourceservices/assets/pdfs/environmental-social-governance-exec-pay-report.pdf, Last Accessed: 2024-07-23.

Prottas, D. J. (2013). Relationships among employee perception of their manager's behavioral integrity, moral distress, and employee attitudes and well-being. *Journal of Business Ethics 113*, 51–60.

Qiu, F., N. Hu, P. Liang, and K. Dow (2022). Measuring management accounting practices using textual analysis. *Management Accounting Research 58*, 100818.

Rajan, R., P. Ramella, and L. Zingales (2023). What purpose do corporations purport? evidence from letters to shareholders. Technical report, NBER Working Paper No. 31054.

Rees, W. and T. Rodionova (2015). The influence of family ownership on corporate social responsibility: An international analysis of publicly listed companies. *Corporate Governance: An International Review 23*(3), 184–202.

Reid, R. S. and J. S. Adams (2001). Human resource management–a survey of practices within family and non-family firms. *Journal of European Industrial Training 25*(6), 310–320.

Rennekamp, K. M., M. Sethuraman, and B. A. Steenhoven (2022). Engagement in earnings conference calls. *Journal of Accounting and Economics 74*(1), 101498.

Rich, B. L., J. A. Lepine, and E. R. Crawford (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal 53*(3), 617–635.

Richter, D., M. Metzing, M. Weinhardt, and J. Schupp (2013). Soep scales manual. Technical report, SOEP Survey Papers.

Ruf, K., J. Mackeben, T. Haepp, S. Wolter, and P. Grunau (2020). Lpp - linked personnel panel 1819: Quality of work and economic success. longitudinal study in german establishments (data documentation on the fourth wave). *FDZ-Datenreport 11/2020 (en)*.

Sageder, M., C. Mitter, and B. Feldbauer-Durstmüller (2018). Image and reputation of family firms: a systematic literature review of the state of research. *Review of Managerial Science 12*, 335–377.

Salanova, M., S. Agut, and J. M. Peiró (2005). Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate. *Journal of Applied Psychology 90*(6), 1217–1227.

Sánchez-Marín, G., Á. L. Meroño-Cerdán, and A. J. Carrasco-Hernández (2019). Formalized hr practices and firm performance: an empirical comparison of family and non-family firms. *The International Journal of Human Resource Management 30*(7), 1084–1110.

Schäffer, U. (2018). The richness of management control research. *Journal of Management Control 29*(2), 91–92.

Schaufeli, W. B. and A. B. Bakker (2003). Utrecht work engagement scale: Preliminary manual. *Occupational Health Psychology Unit, Utrecht University*.

Schaufeli, W. B. and A. B. Bakker (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior 25*(3), 293–315.

Schaufeli, W. B., A. B. Bakker, and M. Salanova (2006a). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement 66*(4), 701–716.

Schaufeli, W. B., M. Salanova, V. González-Romá, and A. B. Bakker (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies 3*(1), 71–92.

Schaufeli, W. B., T. W. Taris, and A. B. Bakker (2006b). Dr. jekyll or mr. hyde: On the differences between work engagement and workaholism. *Research companion to working time and work addiction*, 193–217.

Shang, R., Z. Wang, and Y. Zu (2023). Manager narcissism, target difficulty, and employee dysfunctional behavior. *Contemporary Accounting Research 40*(3), 1795–1822.

Shleifer, A. and L. H. Summers (1988). Breach of trust in hostile takeovers. In *Corporate Takeovers: Causes and Consequences*, pp. 33–68. University of Chicago Press.

Sholihin, M., R. Pike, M. Mangena, and J. Li (2011). Goal-setting participation and goal commitment: Examining the mediating roles of procedural fairness and interpersonal trust in a uk financial services organisation. *The British Accounting Review 43*(2), 135–146.

Silverman, S. B. and K. N. Wexley (1984). Reaction of employees to performance appraisal interviews as a function of their participation in rating scale development. *Personnel Psychology 37*(4), 703–710.

Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: A new look at the interface between nonwork and work. *Journal of Applied Psychology 88*(3), 518–528.

Speckbacher, G. and P. Wentges (2012). The impact of family control on the use of performance measures in strategic target setting and incentive compensation: A research note. *Management Accounting Research 23*(1), 34–46.

Sprinkle, G. B. and M. G. Williamson (2006). Experimental research in managerial accounting. *Handbooks of Management Accounting Research 1*, 415–444.

Sraer, D. and D. Thesmar (2007). Performance and behavior of family firms: Evidence from the french stock market. *Journal of the European Economic Association 5*(4), 709–751.

Statista (2025). Growth of the global gross domestic product (gdp) from 1980 to 2024, with forecasts until 2030. https://www.statista.com/statistics/273951/growth-of-the-global-gross-domestic-product-gdp/, Last Accessed: 2025-07-03.

Subramony, M. (2009). A meta-analytic investigation of the relationship between hrm bundles and firm performance. *Human Resource Management 48*(5), 745–768.

Sumelius, J., I. Björkman, M. Ehrnrooth, K. Mäkelä, and A. Smale (2014). What determines employee perceptions of hrm process features? the case of performance appraisal in mnc subsidiaries. *Human Resource Management 53*(4), 569–592.

Suslava, K. (2021). Stiff business headwinds and uncharted economic waters: The use of euphemisms in earnings conference calls. *Management Science 67*(11), 7184–7213.

Szczygielski, J. J., A. Charteris, P. R. Bwanya, and J. Brzeszczyński (2022). The impact and role of covid-19 uncertainty: A global industry analysis. *International Review of Financial Analysis 80*, 101837.

Tett, R. P. and J. P. Meyer (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: path analyses based on meta-analytic findings. *Personnel Psychology 46*(2), 259–293.

Thurston, P. W. and L. McNall (2010). Justice perceptions of performance appraisal practices. *Journal of Managerial Psychology 25*(3), 201–228.

Tsang, A., K. T. Wang, S. Liu, and L. Yu (2021). Integrating corporate social responsibility criteria into executive compensation and firm innovation: International evidence. *Journal of Corporate Finance 70*, 102070.

Tsoutsoura, M. (2021). Family firms and management practices. *Oxford Review of Economic Policy 37*(2), 323–334.

Tuomela, T.-S. (2005). The interplay of different levers of control: A case study of introducing a new performance measurement system. *Management Accounting Research 16*(3), 293–320.

Unger, O., A. Szczesny, and M. Holderried (2020). Does performance pay increase productivity? evidence from a medical typing unit. *Management Accounting Research 47*, 100649.

U.S. Securities and Exchange Commission (2020). Sec extends conditional exemptions from reporting and proxy delivery requirements for public companies, funds, and investment advisers affected by coronavirus disease 2019 (covid-19). https://www.sec.gov/newsroom/press-releases/2020-73, Last Accessed: 2025-04-04.

Voußem, L., S. Kramer, and U. Schäffer (2016). Fairness perceptions of annual bonus payments: The effects of subjective performance measures and the achievement of bonus targets. *Management Accounting Research 30*, 32–46.

Vroom, V. (1964). *Work and Motivation*. Wiley and Sons, New York.

Wagner, D., J. H. Block, D. Miller, C. Schwens, and G. Xi (2015). A meta-analysis of the financial performance of family firms: Another attempt. *Journal of Family Business Strategy 6*(1), 3–13.

Wang, V. X. and B. B. Xing (2020). Battling uncertainty: Corporate disclosures of covid-19 in earnings conference calls and annual reports. *Available at SSRN 3586085*.

Wibbeke, L.-M. and M. Lachmann (2020). Psychology in management accounting and control research: An overview of the recent literature. *Journal of Management Control 31*(3), 275–328.

Widener, S. K. (2006). Associations between strategic resource importance and performance measure use: The impact on firm performance. *Management Accounting Research 17*(4), 433–457.

Widener, S. K. (2007). An empirical analysis of the levers of control framework. *Accounting, Organizations and Society 32*(7-8), 757–788.

Willis Towers Watson (2023). Global report on esg metrics in executive incentive plans. https://www.wtwco.com/en-ch/insights/2024/01/global-report-on-esg-metrics-in-incentive-plans-2023, Last Accessed: 2024-07-23.

Witter, Frank (2020). Volkswagen ag earnings call 17th march, 2020. Transcript accessed via Refinitiv.

Wood, S., S. Leoni, and D. Ladley (2023). Comparisons of the effects of individual and collective performance-related pay on performance: A review. *Human Resource Management Review 33*(4), 100982.

World Health Organization (2023). Statement on the fifteenth meeting of the international health regulations (2005) emergency committee regarding the coronavirus disease (covid-19) pandemic. https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic, Last Accessed: 2025-01-16.

Wu, P.-C. and S. Chaturvedi (2009). The role of procedural justice and power distance in the relationship between high performance work systems and employee attitudes: A multilevel perspective. *Journal of Management 35*(5), 1228–1247.

Xanthopoulou, D., A. B. Bakker, E. Demerouti, and W. B. Schaufeli (2009). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology 82*(1), 183–200.

Yang, B., X. Ren, and Y. Liu (2024). The spillover effects of managers' evasiveness: Evidence from earnings communication conferences. *Accounting & Finance 64*(3).

Yatsenko, D. (2022). Productivity effects of shared peer effort and relative performance information. *Management Accounting Research 56*, 100779.

Yitzhaky, L., B. Bahli, et al. (2021). Target setting and firm performance: A review. *Journal of Applied Business Research (JABR) 37*(3), 81–94.

Zhou, X. and Y. Huang (2023). Unraveling managerial tangents in firm disclosure: Concealing issues or being exposed? *arXiv preprint arXiv:2309.05555*.

Acknowledgments

This thesis would not have been possible without a combination of genuine curiosity, the occasional existential crisis, and, above all, an extraordinary amount of support from mentors, colleagues, friends, and family.

First and foremost, I am deeply grateful to my first supervisor, Patrick Kampkötter, whose guidance, patience, and encouragement supported me through both the moments of progress and the inevitable detours along the research journey. Thank you for always taking the time to discuss every (empirical) question that crossed my mind, from fixed-effects estimations and standard errors to coding challenges, and model robustness. I truly appreciate your willingness to take on the supervision of this thesis without hesitation. Thank you for making all of this possible in the first place. Secondly, I would also like to thank Martin Ruf for his support as my second supervisor and Kerstin Pull for her immediate willingness to serve as head of the examination committee.

I also want to thank all my wonderful co-authors, Sarah Diederich, Matthias Efing, Michael Haylock, Patrick Kampkötter, Patrick Maier, Raphael Moritz, Julian Nüssle, Julian Wenzel, Stefanie Wolter, and Philip Yang for their insights, creativity, and thoughtfulness, navigating both lively debates and unforeseen challenges with a hint of humor. I have learned so much by being able to work alongside you, no matter if it was a late night evaluation of theory alignment, trying to make sense of the 101st pay scheme we encountered, or going through coding loops and reviewers remarks. Thank you for the countless discussions, laughter, and moral support. It definitely made all the difference.

Thanks also to all the other amazing people I met here over the years. Thanks to Vicky Müller-Henneberg, Maxi Unsorg, and Kristina Uhl for the third-floor morning coffee rounds, Christmas cookie contests, and after-work runs. A particular thanks to you, Vicky. I started this journey together with you and without you, I wouldn't have mastered the inevitable chaos in the beginning. I would also like to thank Marit Holler for encouraging me to try so many new sports I otherwise wouldn't have, and for listening patiently along the way. Thanks also to Daniel Volkmann for all the wonderful homemade treats that kept me fueled, to Nadja Guter-

mann for filling my educational gaps on TV shows, to Karo Ströhlein for countless caffeine-assisted brainstorming sessions, and to Jan Riepe, whom I could always ask for advice. And many thanks to Dagmar Hegedüs for always keeping me well organized and guiding me through all bureaucratic challenges.

Finally, to everyone else who was part of this journey, I cannot thank you enough. Your encouragement and understanding kept me going, even on the days when all I produced were notes that barely made sense (or were readable) to anyone but me. I am deeply thankful to my parents for always believing in me. Thank you, mum, for patiently answering all proofreading questions and for offering your incredible ability to discuss almost anything at length whenever I needed it. Thank you, dad. I wish you could have been here to see this day, I know you would have been proud. Thank you, Ferdi, for keeping me somewhat sane during late-night p-value discussions. Thank you, Meike, for being there through all my (academic) wins and losses since day one of my undergraduate studies here in Tübingen. Thank you, to both you and Marc, for all the countless game nights and cups of tea throughout. Thank you, Aimee, for always being just a phone call away. Thanks to the whole AIAS team with its inspiring enthusiasm and motivation to work together and register as many students as possible as stem cell donors. Thanks to the Sportsout trainer crew for helping me get off my desk, breathe fresh air, and move toward some much-needed balance, both physically and mentally. And most of all, thank you, Fabi, for your endless patience and kindness, for listening to me and trying to make sense of my sometimes rather abrupt changes in thoughts, for intense academic discussions on walks and runs, no matter how much breath there was left, and for always having my back.