University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 389

Uninformative Performance Signals and Forced CEO Turnover

Raphael Flepp

July 2021

University of Zurich, Plattenstrasse 14, CH-8032 Zurich

http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Raphael Flepp

University of Zurich

Department of Business Administration

Plattenstrasse 14, CH-8032 Zurich, Switzerland

raphael.flepp@business.uzh.ch

Uninformative Performance Signals and Forced CEO Turnover

Raphael Flepp*

July 2021

Abstract

This paper provides evidence that corporate boards violate the informativeness principle in their forced CEO turnover decisions by failing to ignore uninformative performance outcome signals. I show that CEOs of firms with barely positive shareholder returns in the previous year are less likely to be dismissed than CEOs of firms with barely negative returns, even though this return outcome is conditionally uninformative. I observe a similar pattern for stock returns relative to the S&P 500 index return: a firm's board is less likely to dismiss its CEO if the firm barely outperformed the S&P 500 index than if the firm barely underperformed the S&P 500 index. Moreover, I demonstrate that the tendency of boards to consider uninformative absolute return outcomes has decreased over time, while their tendency to consider uninformative relative return outcomes has increased over time. This suggests that boards have shifted their focus toward relative returns while continuing to violate the informativeness principle.

Keywords: forced CEO turnover, board of directors, informativeness principle, outcome bias, regression discontinuity design

*Department of Business Administration, University of Zurich. Email: raphael.flepp@business.uzh.ch.

1 Introduction

One of the most important decisions made by corporate boards of directors is whether to retain or to dismiss CEOs (Jenter & Kanaan, 2015). While this decision is a challenging task in the context of uncertainty, boards should follow the informativeness principle developed by Holmström (1979) and include all valuable performance signals when evaluating a CEO. Consequently, boards should ignore uninformative signals from which nothing about the quality of a CEO can be inferred. If a board fails to follow the informativeness principle by including uninformative signals in its turnover decision, the CEO's exposure to risk increases and ex ante managerial effort incentives decrease (Holmström, 1982; Jayaraman, Milbourn, Peters, & Seo, 2020).

One type of (arguably) uninformative performance signal consists of exogenous performance shocks that are beyond the control of a CEO (Holmström, 1982). To completely filter out exogenous performance shocks, boards may use relative performance evaluation (RPE) to assess the quality of CEOs; indeed, numerous previous studies have investigated whether boards employ strong-form RPE (e.g., Bushman, Dai, & Wang, 2010; Jenter & Kanaan, 2015; Fee, Hadlock, Huang, & Pierce, 2018). However, the empirical evidence on whether peer performance affects the CEO retention decisions of boards is inconclusive. For example, while Jenter and Kanaan (2015) find that CEOs are more likely to be dismissed during recessions than economic booms, Fee et al. (2018) find that the relation between industry performance and CEO turnover is both weak and fragile. Moreover, there are several explanations for why peer performance could be informative about CEO quality in certain situations. For example, Eisfeldt and Kuhnen (2013) argue that industry shocks might change the skills required of a CEO. Thus, it remains unclear whether boards follow the informativeness principle and eliminate the uninformative components of peer performance shocks when evaluating CEOs.

In this paper, I investigate novel uninformative performance signals to test whether boards' CEO turnover decisions are consistent with the informativeness principle. In particular, I test whether boards fail to ignore uninformative outcomes. The tendency to consider uninformative outcomes outside the control of an agent in the context of decision-making is referred to as "outcome bias", and this bias has been documented in various experimental studies (e.g., Baron & Hershey, 1988; Rubin & Sheremeta, 2016; Brownback & Kuhn, 2019) and field studies using sports data (e.g., Lefgren, Platt, & Price, 2015; Gauriot & Page, 2019).

I use the information on whether a firm's (relative) stock return in the previous fiscal year was positive or negative, conditional on the (relative) stock return itself, as an uninformative outcome. While the size of a (relative) stock return is clearly an informative signal for CEO evaluation and plays an important role in predicting CEO dismissals (e.g., Parrino, 1997; Peters & Wagner, 2014), a barely positive or barely negative (relative) stock return can be viewed as random and thus uninformative. I use the total shareholder return (TSR) at the end of a fiscal year because this return is usually reported in firms' annual reports; thus, boards are well aware of whether their firms have increased or decreased in shareholder value during the last year. As a measure of the relative total shareholder return (rTSR), I calculate the return relative to the S&P 500 index return, which is an unambiguous and salient benchmark.[1] If boards are less likely to dismiss CEOs when their (relative) total shareholder returns are barely positive than when their (relative) total shareholder returns are barely negative, they reward (or punish) CEOs for good (or bad) luck, thus violating the informativeness principle.

Using the firms and CEOs included in the ExecuComp database between 1993 and 2018, data from the CEO dismissal database of Gentry, Harrison, Quigley, and Boivie (2021a), return information from CRSP, and accounting information from Compustat, I employ a regression discontinuity (RD) design to test whether the treatment of positive (relative) return outcomes has an effect on forced CEO turnover decisions. If CEOs are unable to perfectly select themselves into the treatment of positive (relative) return

[1] Boards certainly employ additional or different measures of relative returns. However, I do not observe either the peer group or the return aggregation rule that the examined boards employ in practice.

outcomes rather than negative (relative) return outcomes, the local causal effect of uninformative (relative) return outcome signals on the likelihood of forced CEO turnover will be revealed (Cattaneo, Idrobo, & Titiunik, 2020).

I follow Calonico, Cattaneo, and Titiunik (2014) and report non-parametric bias-corrected RD estimates with robust variance estimators based on local polynomial regressions. The RD estimate for the treatment of a positive TSR is significantly negative and shows that CEOs of firms with barely positive returns in the previous fiscal year are 1.91 percentage points less likely to be dismissed than CEOs of firms with barely negative returns. This corresponds to a reduction in the probability of forced CEO turnover of approximately 50% of the mean forced CEO turnover rate of 3.85% for non-treatment counterfactual CEOs just below the threshold of zero TSR. The RD estimate for firms that barely outperformed the S&P 500 index return is also negative and statistically significant. It suggests that a CEO is approximately 1.1 percentage points less likely to get dismissed if his or her firm barely outperformed the S&P 500 index. Compared to the mean forced CEO turnover rate of 2.56% for the non-treatment counterfactual CEOs who barely underperformed the S&P 500 index, this corresponds to a decrease of approximately 43%.

Due to an increase in the explicit use of rTSR in executive compensation contracts over time (Ma, Shin, & Wang, 2021), I expect that rTSR has also become more salient for CEO turnover decisions during more recent years. As a consequence, the effect of CEO dismissal being less likely after barely outperforming the S&P 500 index than after barely underperforming the S&P 500 index should become more pronounced over time. Indeed, I find that the RD estimate for barely outperforming the S&P 500 index is insignificant and close to zero between 1993 and 2006. However, between 2007 and 2018, this RD estimate becomes more pronounced, showing that CEOs are approximately 2.2 percentage points less likely to be dismissed after barely outperforming the S&P 500 index. In contrast, the RD estimate corresponding to the positive TSR outcome is more pronounced for the earlier time period, namely, from 1993 to 2006, while it is statistically insignificant between 2007 and 2018. Overall, it seems that while the focus of boards shifted to relative performance evaluation metrics, considerations of uninformative return outcomes in the context of turnover decisions neither disappeared nor weakened over time.

The main assumption underlying the RD design of this study is that the only change that occurs at the cutoff of zero (relative) returns is a shift in the treatment status. While this continuity assumption cannot be tested directly, several methods have been established in the literature that provide indirect evidence regarding the validity of an RD design (e.g., Imbens & Lemieux, 2008; Cattaneo et al., 2020). I do not find any evidence that CEOs or firms are able to manipulate the (relative) returns precisely to produce barely positive returns instead of barely negative returns. Furthermore, I test whether firms just below and above the cutoff of zero (relative) returns are similar in terms of other predetermined characteristics. Repeating the same RD methods, I do not find any evidence of discontinuities for various firm characteristics, such as firm size, market value, and firm performance, and CEO characteristics, such as CEO tenure, whether a CEO is of retirement age, and whether a CEO is the chairman of the board. Similarly, I do not find evidence of discontinuities at artificial cutoff values in the (relative) return functions. Finally, I show that the results remain mostly robust to the use of different bandwidth values and to the use of alternative parametric specifications of the RD design. Overall, these falsification and validation tests provide convincing evidence that the RD design is valid in this setting.

Even though my results are consistent with boards failing to ignore uninformative performance outcome signals and thus violating the informativeness principle, several alternative explanations could be given for such an empirical pattern. First, CEOs also observe (relative) return outcomes, which may impact their level of effort and performance in the subsequent year. For example, CEOs who barely outperform a benchmark might become more motivated and thus increase their performance during the following year, while CEOs that barely underperform a benchmark may not. Thus, dismissing fewer CEOs who barely outperform could be beneficial because these CEOs might actually perform better in the subsequent year. To test whether this explanation is the main driver of the results, I investigate whether the CEOs of firms with barely positive return outcomes perform significantly better in the subsequent year than CEOs of firms with barely negative return outcomes, conditional upon those CEOs remained in office. Using the RD approach outlined above, I do not find any evidence that a positive return outcome induces better firm performance during the subsequent year.

Another confounding explanation of the results could be that CEOs of firms with barely positive return outcomes are more likely to leave their firms voluntarily because they may feel more inclined to retire or because other firms may offer them more attractive job opportunities (Amore & Schwenen, 2020). If some of these CEOs who left voluntarily would have been fired in the subsequent year, this phenomenon could lead to the empirical pattern that CEOs are less likely to be fired after their firms produce barely positive return outcomes. However, the data do not indicate that voluntary CEO turnover is more likely to occur after a barely positive return.

This paper contributes to the literature in several ways. First, I show that boards consider uninformative performance signals in their forced CEO turnover decisions, which is inconsistent with the informativeness principle. Using an RD design, I present novel, causal empirical evidence showing that CEOs of firms with barely positive (relative) shareholder returns are less likely to get fired than CEOs of firms with barely negative (relative) returns, although CEOs cannot precisely influence return realization.

Second, this paper contributes to the debate on the extent to which the informativeness principle is followed in boards' CEO turnover decisions. While earlier studies have focused on boards' use of RPE to filter out exogenous shocks in accordance with the informativeness principle, the empirical results on this topic are mixed and subject to alternative explanations. Using an RD design in which firms and CEOs with barely positive and barely negative (relative) returns are very similar allows me to overcome several arguments regarding the potential informativeness of this signal. For example, while peer performance could be informative with respect to newly required CEO skills (Eisfeldt & Kuhnen, 2013) or strategic exposure to peer performance (Gopalan, Milbourn, & Song, 2010), the sign of the (relative) return conditional on the (relative) return is not informative. Finally, I demonstrate that outcome bias exists in the context of corporate boards, which has not been investigated so far. Thus, this paper extends the literature on outcome bias, which has primarily focused on laboratory studies and sports settings, to a more traditional field setting.

The results of this paper have important implications for firms. Basing CEO turnover decisions partly on uninformative signals related to factors that are not under control of a CEO is costly, as it exposes CEOs to unnecessary risks and may lead to inefficient effort incentives. The magnitude of the effects found in this study seems considerable and economically substantial. A barely positive (relative) return results in a probability of forced turnover that is approximately $40-50\%$ less than that of CEOs of firms with barely negative (relative) returns. Thus, firms that avoid this tendency could gain a competitive advantage. However, my results also show that boards' consideration of uninformative return outcomes has not disappeared over time, which suggests that even modern boards with sophisticated data analyis tools at their disposal may consider uninformative performance signals.

The remainder of this paper is structured as follows. Section 2 reviews the literature and develops the hypotheses. Section 3 describes the data and the empirical methods. Section 4 presents the main results, and Section 5 concludes the paper.

2 Theory and Hypotheses

2.1 Related Literature

The informativeness principle developed by Holmström (1979) states that all valuable information signals should be used by a principal to optimize contracts when an agent's actions are unobservable. Importantly, an information signal is valuable only if it provides additional information about the actions of the corresponding agent given any previous information signals. Consequently, uninformative signals should not be considered because they will only make a contract riskier (Holmström, 1979). When deciding whether to dismiss or retain a CEO, boards of directors should follow the informativeness principle to optimize ex ante managerial incentives and to reduce the CEO's risk exposure. In particular, CEOs should not be dismissed due to factors beyond their control and thus unrelated to their managerial effort (Jayaraman et al., 2020).

One type of (arguably) uninformative signal that has drawn much attention in the literature consists of common observable uncertainties such as industry or market shocks. Because such shocks are exogenous, they should be filtered out by boards using relative performance evaluation (RPE) when the quality of a CEO is being assessed (Holmström, 1982). Based on Holmström (1982), Antle and Smith (1986) propose a strong-form RPE test, in which CEOs are only held accountable for unsystematic firm performance that is unrelated to peer performance; thus, peer performance should be completely filtered out.[2] However, the empirical evidence on strong-form RPE in the context of CEO turnover decisions and the interpretations thereof are mixed.

Using data on the firms in the ExecuComp database from 1993 to 2009, Jenter and Kanaan (2015) find that CEOs are more likely to be dismissed after negative performance shocks to their peer groups while controlling for firm-specific performance. Thus, they conclude that boards allow exogenous shocks to firm performance to affect their CEO

[2] Under weak-form RPE, external shocks are expected to be filtered out only to some extent (Jenter & Kanaan, 2015; Albuquerque, 2009; Gibbons & Murphy, 1990).

turnover decisions. Similar results are provided by Bushman et al. (2010) and Kaplan and Minton (2012). In contrast, Barro and Barro (1990) show that boards completely filter out peer performance in the context of CEO turnover decisions, and Fee et al. (2018) find that the relation between industry performance and CEO turnover is both weak and fragile. Moreover, although a sample and method similar to those of Jenter and Kanaan (2015) are used, the results of Hazarika, Karpoff, and Nahata (2012) show that industry returns are not significantly related to forced CEO turnover. Furthermore, in Huang, Maharjan, and Thakor (2020), the coefficients of equal-weighted industry returns reveal no significant effect on forced CEO turnover using a large ExecuComp sample from 1993 to 2017.

Although the fact that CEOs are dismissed more often during recessions than during economic booms is consistent with inefficient board decisions, there are several alternative explanations for why this pattern may be optimal. First, industry shocks might change the skills required of CEOs (Eisfeldt & Kuhnen, 2013). In their model, Eisfeldt and Kuhnen (2013) propose that boards should consider industry conditions in their hiring and firing decisions because the current CEO of a company may not possess the skills necessary for the new state of the industry. Consistent with their theoretical predictions, Eisfeldt and Kuhnen (2013) find that CEOs are more likely to be replaced by successors from outside of the industry after industry skill-weight shocks. Second, Gopalan et al. (2010) argue that CEOs can influence their firms' exposure to sector performance through their choice of firm strategy. Thus, to provide incentives for CEOs to choose the optimal level of sector exposure, a firm's forced turnover should depend on sector performance if it offers strategic flexbility to its CEO. Consistent with their model's predictions, Gopalan et al. (2010) find that sector performance is only related to forced CEO turnover in firms that offer a high level of strategic flexibility, proxied by high market-to-book ratios and R&D expenditures.

Third, Albuquerque (2009) and Jayaraman et al. (2020) argue that the lack of strong-form RPE in the context of CEO dismissals is due to the limited number of peers available in practice and the noisy peer groups employed by previous studies. Jayaraman et al. (2020) postulate that the ideal peer group should consist of firms that produce similar products and have a common uncertainty parameter. Thus, they use the Text-based Network Industry Classifications (TNIC) developed by Hoberg and Phillips (2016) to construct a peer group instead of relying on industry classification codes such as the Standard Industry Classification (SIC). Using data of ExecuComp CEOs between 1996 and 2015, Jayaraman et al. (2020) show that more peer performance is filtered out of forced CEO turnover decisions when there is a greater number of peers. Importantly, in an environment with many peer firms, boards seem to completely filter out common uncertainties.

Finally, these previous studies face a joint hypothesis problem because the peer group used by a board is typically not observable (Dikolli, Hofmann, & Pfeiffer, 2013). Thus, peer groups are implicitly chosen under the assumption that they are the most likely peer groups to be used by boards.[3] However, it seems likely that empiricists aggregate peer performance differently than boards do, and Dikolli et al. (2013) show that chosing more, fewer or different peers than a board does creates a bias that can lead to a false rejection of the relative performance evaluation hypothesis. Thus, the joint hypothesis problem further complicates the interpetation of the previous empirical evidence on relative performance evaluation. Overall, the empirical evidence on whether boards follow the informativeness principle when evaluating the performance of CEOs is mixed at best, and the methods employed allow for several alternative explanations for this phenomenon other than boards behaving suboptimally.

[3] As Jayaraman et al. (2020) note, the SEC has required public firms to disclose their performance peers for executive compensation since 2006, and it allows for the explicit construction of peer groups. However, the SEC regulations allow a considerable scope in this regard, and only approximately 10-20% of the largest 750 firms disclose individual peer firms (Jayaraman et al., 2020). Moreover, even if the selected peers are known, the peer group aggregation weights remain unobservable (Dikolli et al., 2013).

Another type of uninformative signal is an outcome from which nothing about the actions of an agent can be inferred (Holmström, 1979; Gauriot & Page, 2019). If such outcomes are taken into account when evaluating this agent, outcome bias is present (Baron & Hershey, 1988). In their laboratory experiments, Baron and Hershey (1988) gave subjects different medical decisions with known outcome probabilities. For example, a heart surgery led to death or to a cure with fixed ex ante probabilities. These authors found a consistent outcome bias, as the subjects rated the decisions as better and the decision makers as more competent when an outcome was favorable than when it was unfavorable. Similar outcome biases have been documented in a variety of vignette-based experimental studies (e.g., Marshall & Mowen, 1993; Mazzocco, Alicke, & Davis, 2004; Gino, Shu, & Bazerman, 2010).

Furthermore, there is a considerable amount of experimental literature on outcome bias that employs real subject interactions with monetary consequences (e.g., Gurdal, Miller, & Rustichini, 2013; Rubin & Sheremeta, 2016; König-Kersting, Pollmann, Potters, & Trautmann, 2021). Investigating a principal-agent setting with random shocks, Rubin and Sheremeta (2016) find that the average punishment inflicted by principals is significantly greater after negative shocks than after positive shocks. Moreover, principals reward agents similarly in the case of increases in outcomes even if those agents' effort is observed. Thus, principals punish or reward agents partly based on outcomes that are outside the agents' control. In a similar principal-agent experiment, Brownback and Kuhn (2019) investigate the outcome bias that occurs when the effort of an agent is perfectly observed by a principal, which ensures that information about an outcome provides no additional signal value regarding effort. Even in this transparent environment, outcome bias remains strong, as principals' punishments are influenced by luck. By eliciting the beliefs of principals and third parties, Brownback and Kuhn (2019) discover that this outcome bias likely stems from biased beliefs about agents because lucky agents are thought to be harder workers than unlucky agents.

Most of the field evidence on outcome bias stems from sports settings. Using data from professional basketball, Lefgren et al. (2015) employ a regression discontinuity design and compare coaching decisions after narrow losses to coaching decisions after narrow wins. Even though narrow losses are uninformative in terms of team effectiveness and future success, coaches are more likely to revise their starting lineup strategies following a narrow loss than following a narrow win. Consistent with their model's predictions, the results of Lefgren et al. (2015) suggest that coaches overweight the likelihood of outcomes that have occurred, which influences their beliefs regarding the correct strategy in a biased way. Thus, outcome bias also affects the real-life decisions of experts in highstakes settings.

Gauriot and Page (2019) examine shots made by professional football (soccer) players that hit goal posts. If a post is hit, whether a goals is scored is essentially random. Thus, the outcomes of such shots are uninformative and should not play a role in player evaluations if evaluators follow the informativeness principle of Holmström (1979). Gauriot and Page (2019) find causal evidence that lucky successes of players are rewarded by team managers through the provision of more playing time during the following match and by third-party expert journalists through higher ratings. Similarly, Kausel, Ventura, and Rodríguez (2019) examine the outcomes of penalty shoot-outs, which seem unrelated to in-game performance. Even when the players who took part in the examined penalty shoot-outs are removed from the analysis, winning on penalties is associated with higher subjective performance ratings from reporters.

Additionally, in football clubs, decision-maker outcome bias may affect the quality of coach dismissal decisions because match outcomes in professional football are disproportionately influenced by random factors due to the low-scoring nature of the game (Brechot & Flepp, 2020). Indeed, Flepp and Franck (2021) show that only dismissals following actual poor performance on the pitch improve team performance, while dismissals after seemingly poor performance due to bad luck do not.

Given that there is generally less information available about agents' actions in firms than in sports settings, one would expect that outcome bias is even more prominent in firm environments than in sports environments (Gauriot & Page, 2019). However, no study has investigated how outcome bias may contribute to violations of the informativeness principle in the context of CEO turnover decisions. If boards fail to ignore uninformative outcomes conditional on informative performance signals, inefficient dismissal decisions will result. In the following section, I carefully derive salient but conditionally uninformative outcome signals to test whether such signals play a role in boards' evaluations of CEOs.

2.2 Hypotheses

Prior research on CEO turnover has shown that stock returns play an important role in predicting CEO turnover as measures of firm performance (e.g., Parrino, 1997). Prior evidence suggests that stock returns are a more important determinant of CEO turnover than accounting measures because they are largely unpredictable and less prone to manipulation (Fee et al., 2018; Barro & Barro, 1990). Furthermore, shareholder returns are usually reported in firms' annual reports and thus are salient for boards of directors. An obvious outcome related to annual shareholder returns is whether shareholder value has increased or been destroyed. If a firm's annual shareholder return is negative, shareholder value has been destroyed, while a positive return indicates an increase in shareholder value. Although this return outcome is informative about the performance of a CEO when analyzed overall, it is uninformative when it is just below or above the return threshold of zero. In other words, the outcome of a barely positive or barely negative annual shareholder return can be viewed as random and thus uninformative.

If boards follow the informativeness principle, they should ignore whether annual returns are barely positive or barely negative, and the sign of a return should not affect the likelihood of CEO dismissal. However, the preceding discussion has shown that outcome bias is persistent in various contexts even when effort is perfectly observable (e.g., Brownback & Kuhn, 2019). Thus, assuming that boards recognize whether annual returns are positive or negative and prefer a positive return outcome over a negative return outcome, I expect that boards also exhibit outcome bias. This motivates the first hypothesis:

H1. *Boards are less likely to dismiss CEOs after barely positive annual shareholder returns than after barely negative returns.*

However, most boards not only consider absolute shareholder returns but also employ relative shareholder return measures to evaluate CEOs in an attempt to filter out common market and industry factors. A firm's relative total shareholder return (rTSR) is defined as the difference between its total shareholder return (TSR) and the aggregate return of a group of peer firms or an index (Ma et al., 2021). Since 2006, SEC executive compensation rules have required firms to disclose their use of rTSR (Gong, Li, & Shin, 2011), and more than 40% of the largest 1,000 listed firms in the U.S. explicitly reported the use of rTSR in 2014 (Ma et al., 2021). Among these firms, most select a specific peer group or use of an index to derive rTSR (Bakke, Mahmudi, & Newton, 2020; Park & Vrettos, 2015). Moreover, other market participants, such as activist investors, also focus on rTSR to evaluate firms' management quality (Brav, Jiang, Partnoy, & Thomas, 2008; Ma et al., 2021). Even though data on the use of rTSR in CEO turnover decisions are not available, boards are likely to use a form of peer-group adjusted stock returns when evaluating CEOs. Indeed, several earlier papers have shown that the peer-adjusted stock returns of firms (e.g., the industry-adjusted firm returns) are an important predictor of CEO dismissals (e.g., Jenter & Kanaan, 2015; Jayaraman et al., 2020; Peters & Wagner, 2014).

Unfortunately, the use of rTSR based on specifically reported peer groups for executive compensation is not fruitful in the context of examining CEO turnover decisions for several reasons. First, we do not know whether boards use the same peer groups for executive compensation and CEO retention decisions. Second, the specific peer groups selected by firms are only observable if they report these groups, and as Jayaraman et al. (2020)

highlight, the relevant SEC regulation allows for significant flexibility in the extent of disclosure. Third, even if the specific peers chosen by every firm are known, how firms aggregate this peer perfromance remains unclear. Firms may use an equal-weighted, value-weighted, or other algorithm to aggregate the performance of their peer group. Fourth, each firm utilizes a different benchmark, which makes firms that barely underperform a benchmark less comparable to those that barely outperform a different benchmark. Finally, some firms seem to systematically choose underperforming peers (Bakke et al., 2020), making them more likely to outperform their benchmarks.

In testing whether boards fail to ignore uninformative outcome signals with respect to relative total shareholder returns, it is of utmost importance to aggregate peer performance precisely. Thus, I choose the S&P 500 index return as an unambiguous, salient benchmark. First, this return is unambiguous because its calculation follows a transparent set of rules. Second, it is salient because the S&P 500 index incorporates the 500 largest companies in the U.S. based on market value. Even if boards are likely to use additional relative performance measures to evaluate their CEOs, it is also likely that they are aware of the S&P 500 index return and whether their company was able to outperform this index during a given year. Again, conditional on the relative total shareholder return with respect to the S&P 500 index, information on whether a firm barely underperformed the index or barely outperformed the index is not informative. However, based on the reasoning above, I hypothesize the following:

H2. *Boards are less likely to dismiss CEOs after barely outperforming the S&P 500 index than after barely underperforming the S&P 500 index.*

Because the use of relative total shareholder returns in executive compensation has become more common over time (Bakke et al., 2020; Ma et al., 2021), it is possible that the use of rTSR in CEO turnover decisions has also become more widespread during recent years. As a consequence, I expect that the aforementioned decrease in the likelihood of CEO dismissal after firms barely outperform the S&P 500 index has become more pronounced over time. Additionally, boards' shift in focus toward relative measures might have decreased the salience of TSR, which would in turn weaken the tendency to consider the return outcome sign of TSR in subsequent years. These arguments lead to Hypotheses 3 and 4:

H3. *The effect of CEO dismissal being less likely after barely positive annual shareholder returns than after barely negative returns has decreased over time.*

H4. *The effect of CEO dismissal being less likely after barely outperforming the S&P 500 index than after barely underperforming the S&P 500 index has increased over time.*

Finding evidence consistent with Hypotheses 3 and 4 would suggest that even though the focus of boards has shifted from absolute shareholder returns to relative shareholder returns, board's consideration of uninformative performance signals has not disappeared over time.

3 Data and Methods

3.1 Data

The sample firms and CEOs in my data set for the period between 1993 and 2018 are taken from the ExecuComp database, which contains S&P 1500 firms. I merge this sample with the CEO dismissal database[4] constructed by Gentry et al. (2021a) and drop all non-CEO observations. The dismissal database categorizes all the CEO turnovers in ExecuComp into groups corresponding to various forms of voluntary and involuntary departures. I categorize an instance of CEO turnover as forced if the dismissal is related to job performance ($departure\_code = 3$) and exclude all fiscal years with other turnover reasons,

[4] These data are open source and can be downloaded from https://doi.org/10.5281/zenodo.4543893 (Gentry, Harrison, Quigley, & Boivie, 2021b).

such as voluntary turnover due to retirement or new job opportunities and involuntary turnover due to illness or personal issues, from the main analysis.[5]

I add annualized stock returns and S&P 500 index returns at the end of each fiscal year, which are derived from CRSP, to these data. The annualized stock return of a firm is equal to its total shareholder return (*TSR*) including reinvested dividends. I define the difference between the TSR of a firm and the S&P 500 index return as its relative total shareholder return (*rTSR_S&P500*). Finally, I add accounting information regarding the total assets (item *AT*) and market value (items $CSHO \times PRCC\_F$) of each firm at the end of each fiscal year, which I take from Compustat.

Using information from ExecuComp, I derive several CEO characteristics. Following Peters and Wagner (2014) and Jenter and Kanaan (2015), I construct a dummy variable indicating whether a CEO is older than 59 years old ($CEO\ age \geq 60$) based on the item *age*. Furthermore, using the item *becameceo*, I calculate each CEO's tenure in years at the end of each fiscal year.[6] Finally, I construct a dummy variable indicating whether each CEO is the chairman of the board based on the item *titleann*.

Following Jenter and Kanaan (2015), Fee et al. (2018), and Parrino (1997), I only include CEOs who have been in office for at least 24 months at the beginning of a given fiscal year to ensure that the previous performance measures are fully attributable to the current CEO.[7] Table 1 reports the descriptive statistics of the sample. In this sample, forced CEO turnover due to performance occurs in approximately 2.86% of all firm-year observations. This rate is comparable to those found by previous studies that reported forced CEO turnover rates between 2.0% and 2.8% (Peters & Wagner, 2014; Jenter & Kanaan, 2015; Guo & Masulis, 2015; Jayaraman et al., 2020).

[5] Except when a CEO departure is coded as an ExecuComp error (*departure_code* = 9) in the CEO dismissal database constructed by Gentry et al. (2021a).

[6] If the item *becameceo* is missing or inaccurate, I infer the approximate tenure of the corresponding CEO based on the number of fiscal years he or she is observable in the ExecuComp database.

[7] If I relax this restriction to a minimum CEO tenure of 12 months at the beginning of the fiscal year, the magnitude of the effects found slightly decreases. However, the main interpretation of the results remains unchanged.

Table 1
Descriptive Statistics

Panel A: Forced CEO Turnover				
	Mean	Median	Std. Dev.	N
Forced CEO Turnover	0.0286	0	0.1666	29,801
Panel B: Firm Characteristics				
	Mean	Median	Std. Dev.	N
TSR_{t-1}	0.2060	0.1268	0.7030	27,956
$rTSR\_S\&P500_{t-1}$	0.1202	0.0372	0.6805	27,956
$Total\ assets_{t-1}$ (\$m)	14,315	1,681	92,000	29,727
$Market\ value_{t-1}$ (\$m)	7,500	1,502	26,000	29,253
Panel C: CEO Characteristics				
CEO age $\geq$ *60*	0.3247	0	0.4683	29,801
CEO tenure	10	7.75	7.29	29,801
CEO is chairman	0.6206	1	0.4853	29,801

Notes: The sample consists of the firm-year observations from the ExecuComp database between 1993 and 2018 excluding those with turnovers that are not forced due to performance and those involving CEOs who have been in office for less than 24 months.

3.2 Regression Discontinuity Design

The RD design is a non-experimental research design that allows for a credible analysis of causal effects (Cattaneo et al., 2020). In this paper, the running variable is either the *TSR* or the *rTSR_S&P500* at the end of the previous fiscal year, the treatment variable is an indicator of whether the (relative) return at the end of the previous fiscal year was positive (Pos_TSR_{t-1} or $Pos\_rTSR\_S\&P500_{t-1}$), the cutoff is set to a (relative) return of zero and the dependent variable Y is an indicator of forced CEO turnover. Following Calonico et al. (2014) and Cattaneo et al. (2020), I use a non-parametric local linear approach to estimate the sharp RD point estimate τ.[8] The estimating equation using TSR_{t-1} as the running variable is given by

$$Y_{ij} = \beta + \tau \cdot Pos\_TSR_{t-1,ij} + \gamma \cdot TSR_{t-1,ij} + \delta \cdot Pos\_TSR_{t-1,ij} \cdot TSR_{t-1,ij} + \epsilon_{ij} \quad (1)$$

where i denotes the firm and j refers to the fiscal year. Analogously, the estimating equation using $rTSR\_S\&P500_{t-1}$ as the running variable is given by

$$\begin{aligned} Y_{ij} = \quad & \beta + \tau \cdot Pos\_rTSR\_S\&P500_{t-1,ij} + \gamma \cdot rTSR\_S\&P500_{t-1,ij} \\ & + \delta \cdot Pos\_rTSR\_S\&P500_{t-1,ij} \cdot rTSR\_S\&P500_{t-1,ij} + \epsilon_{ij} \end{aligned} \quad (2)$$

Following Calonico et al. (2014) and Cattaneo et al. (2020), I employ a triangular kernel function to determine the weight of the observations within the bandwidth and choose the bandwidth that optimizes the mean squared error (MSE) of the estimate. The choice of bandwidth is crucial in RD designs, and a data-driven bandwidth selector prevents intransparent, ad hoc choices. Because larger bandwidths tend to increase the bias of an estimator while smaller bandwidths tend to increase the variance, a bandwidth that minimizes the MSE optimizes this bias-variance trade-off (Cattaneo et al., 2020).

However, because an MSE-optimal bandwidth is not small enough to remove the bias term, conventional statistical inference methods that ignore this bias term are invalid (Cattaneo et al., 2020). Thus, I employ the bias-corrected RD approach developed by Calonico et al. (2014), based on which the estimated bias term is removed from the RD point estimator τ in Equations (1) and (2) and robust confidence intervals are used for inference.[9]

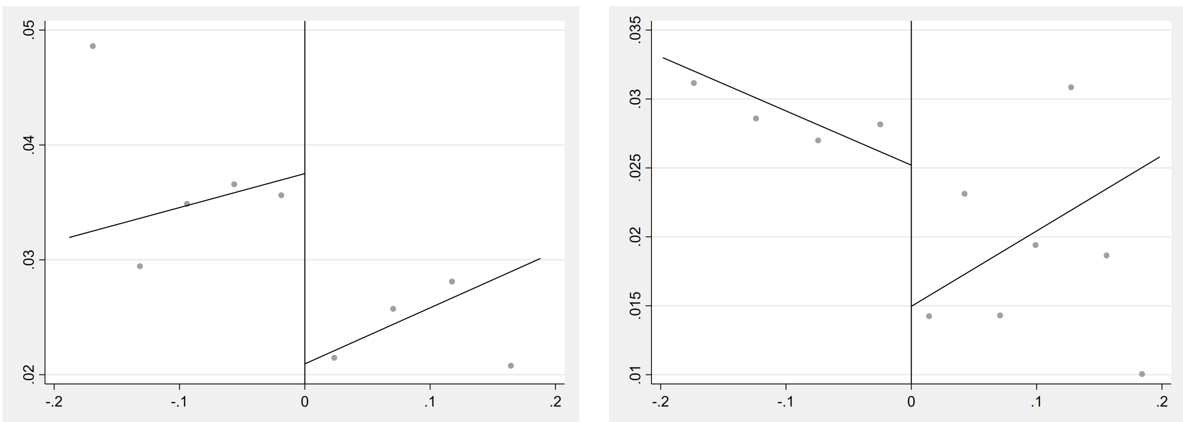
[8] Because an RD estimator is a boundary point, global parametric polynomial regressions can lead to misleading RD point estimators (Cattaneo et al., 2020). However, I also conduct a parametric approach as a sensitivity test in Section 4.2.5.

[9] See Calonico et al. (2014) for a detailed discussion of this approach.

4 Results

4.1 Main Results

I start by providing graphical evidence on whether a barely positive *TSR* or *rTSR_S&P500* in the previous year decreases the likelihood of forced CEO turnover compared to a barely negative *TSR* or *rTSR_S&P500* in the previous year. In Figure 1, each dot represents the mean forced CEO turnover probability within a bin. The number of evenly spaced bins is chosen using the integrated mean squared error (IMSE)-optimal approach within the MSE-optimal bandwidth (Cattaneo et al., 2020; Calonico et al., 2014).[10] The solid line represents the predicted forced CEO turnover probability estimated using a local linear fit. Figure 1a indicates that there is a discontinuous drop in the CEO turnover probability at a TSR_{t-1} of zero. Such a drop is also visible in Figure 1b at a $rTSR\_S\&P500_{t-1}$ of zero, which provides the first suggestive evidence that boards take the uninformative performance signal of a barely positive return into account in their forced CEO turnover decisions.



(a) TSR_{t-1} (b) $rTSR\_S\&P500_{t-1}$

Figure 1
IMSE-optimal RD plot with evenly spaced bins and a local linear fit within the MSE-optimal bandwidth.

[10] Minimizing the IMSE of the local means estimator allows for a data-driven choice of the number of bins. Using the IMSE-optimal number of bins balances the trade-off between greater variability within a bin, which exists if the number of bins is large and a greater bias, which exists if the number of bins is small (Cattaneo et al., 2020).

To investigate this more formally, I follow Calonico et al. (2014) and estimate the non-parametric bias-corrected RD effect with a robust variance estimator. Table 2 displays the main results. As shown in Column (1), the RD estimate of having a barely positive TSR_{t-1} rather than a barely negative TSR_{t-1} is negative and statistically significant at the 5% level. More precisely, this point estimate suggests that CEOs of firms with a barely positive TSR in the previous fiscal year are 1.91 percentage points less likely to be fired than CEOs of firms with a barely negative TSR in the previous fiscal year. This corresponds to a decrease of approximately 50 percent in the control mean of 3.85%.[11] Although not required in RD designs, Column (2) of Table 2 shows that the RD estimate remains similar when covariates for firm characteristics (*Ln(Total assets)*$_{t-1}$ and *Ln(Market value)*$_{t-1}$) and CEO characteristics (*CEO age* $\geq$ *60*, *CEO tenure*, and *CEO is chairman*), as well as year and industry dummies are included.[12] These results are consistent with Hypothesis 1, which states that boards are less likely to dismiss CEOs after barely positive annual shareholder returns than after barely negative annual shareholder returns.

The results regarding *rTSR_S&P500*$_{t-1}$ with a cutoff of zero outperformance are displayed in Columns (3) and (4) of Table 2. The point estimates suggest that CEOs of firms that barely outperformed the S&P 500 index during the previous fiscal year are approximately 1.10 percentage points less likely to be fired than CEOs of firms that barely underperformed the S&P 500 index. Compared to the control mean of 2.56%, this corresponds to a decrease of approximately 43 percent.[13] Thus, these findings are supportive of Hypothesis 2. Overall, both stock return measures suggest that there is a negative discontinuity of approximately $1 - 2$ percentage points in the probability of forced CEO turnover at (relative) returns of zero, which corresponds to a decrease of $40 - 50\%$ in the control mean turnover probability.

[11] Following Ludwig and Miller (2007), I define the control mean as the bias-corrected local polynomial estimate of the likelihood of forced CEO turnover just below the threshold of zero, which represents the non-treatment counterfactual.

[12] The industries are based on the Fama and French (1997) classification of 48 industries.

[13] As one would expect, the control mean of forced CEO turnover barely below a TSR_{t-1} of zero is higher than that barely below a *rTSR_S&P500*$_{t-1}$ of zero because firms that perform at a level similar to that of the S&P 500 index have TSRs that are typically greater than zero.

Table 2
Main Results

	Forced CEO Turnover			
	(1)	(2)	(3)	(4)
Bias-corrected RD estimate	-0.0191**	-0.0172**		
c($TSR_{t-1} = 0$)	(0.0081)	(0.0077)		
Bias-corrected RD estimate			-0.0118**	-0.0106*
c($rTSR\_S\&P500_{t-1} = 0$)			(0.0059)	(0.0058)
Covariates	No	Yes	No	Yes
Observations	27,956	27,949	27,956	27,949
Eff. Obs. left of cutoff	4,682	4,844	6,511	6,747
Eff. Obs. right of cutoff	6,499	6,805	6,756	7,029
Bandwidth	0.188	0.198	0.198	0.208
Bandwidth for bias estimate	0.339	0.356	0.373	0.391

Notes: This table reports the RD results of local linear regressions with triangular kernel weights and MSE-optimal bandwidths. The cutoff is equalt to a (relative) return of zero. The RD estimates are bias-corrected following Calonico et al. (2014), with local quadratic regressions and MSE-optimal bandwidths for the bias estimators. The estimates in Columns (2) and (4) are covariate-adjusted (Calonico, Cattaneo, Farrell, & Titiunik, 2019) with the following covariates: *Ln(Total assets)*$_{t-1}$, *Ln(Market value)*$_{t-1}$, *CEO age* $\geq$ *60*, *CEO tenure*, and *CEO is chairman*, as well as year and industry dummies. The robust standard errors are reported in parentheses. In all the models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

To test Hypotheses 3 and 4, I split the sample in half and repeat the RD methods mentioned above. Panel A of Table 3 shows the results corresponding to the time period from 1993 to 2006. While the RD estimates of a positive TSR_{t-1} return outcome are more pronounced during this period, the RD estimates of a positive $rTSR\_S\&P500_{t-1}$ return outcome are close to zero and statistically insignificant. This is consistent with the explanation that during this period, boards focused on absolute returns, while relative returns were less salient.

Panel B of Table 3 shows the results for the period from 2007 to 2018. In this later period, the RD estimates of TSR_{t-1} are still negative but less pronounced and statistically insignificant. However, the RD estimates are not significantly different from those for the

Table 3
Time Period Subsample Results

Panel A: 1993 to 2006

	Forced CEO Turnover			
	(1)	(2)	(3)	(4)
Bias-corrected RD estimate	-0.0311**	-0.0290**		
$c(TSR_{t-1} = 0)$	(0.0132)	(0.0125)		
Bias-corrected RD estimate			-0.0001	0.0012
$c(rTSR\_S\&P500_{t-1} = 0)$			(0.0083)	(0.0082)
Covariates	No	Yes	No	Yes
Observations	13,319	13,314	13,319	13,314
Eff. Obs. left of cutoff	1,952	1,997	3,442	3,461
Eff. Obs. right of cutoff	2,517	2,604	3,723	3,754
Bandwidth	0.161	0.166	0.265	0.268
Bandwidth for bias estimate	0.282	0.300	0.445	0.453

Panel B: 2007 to 2018

	Forced CEO Turnover			
	(1)	(2)	(3)	(4)
Bias-corrected RD estimate	-0.0101	-0.0098		
$c(TSR_{t-1} = 0)$	(0.0108)	(0.0106)		
Bias-corrected RD estimate			-0.0220**	-0.0218**
$c(rTSR\_S\&P500_{t-1} = 0)$			(0.0088)	(0.0087)
Covariates	No	Yes	No	Yes
Observations	14,637	14,635	14,637	14,635
Eff. Obs. left of cutoff	2,578	2,492	2,874	2,805
Eff. Obs. right of cutoff	3,796	3,622	2,910	2,819
Bandwidth	0.200	0.191	0.144	0.140
Bandwidth for bias estimate	0.323	0.325	0.293	0.293

Notes: This table reports the RD results of local linear regressions with triangular kernel weights and MSE-optimal bandwidths for the periods from 1993 to 2006 (Panel A) and 2007 to 2018 (Panel B). The cutoff is equal to a (relative) return of zero. The RD estimates are bias-corrected following Calonico et al. (2014), with local quadratic regressions and MSE-optimal bandwidths for the bias estimators. The estimates in Columns (2) and (4) are covariate-adjusted (Calonico et al., 2019) with the following covariates: $Ln(Total\ assets)_{t-1}$, $Ln(Market\ value)_{t-1}$, $CEO\ age \geq 60$, *CEO tenure*, and *CEO is chairman*, as well as year and industry dummies. The robust standard errors are reported in parentheses. In all the models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

period between 1993 and 2006.[14] In contrast, the RD estimates of $rTSR\_S\&P500_{t-1}$ become more pronounced during this later period, and they are significantly different from those in Panel A of Table 3.[15] Thus, I find weak evidence for Hypothesis 3 but convincing evidence for Hypothesis 4. Overall, it seems that the focus of boards shifted to relative performance evaluation metrics over time, but their consideration of uninformative performance signals in their decision making did not disappear or weaken over the years.

4.2 Validation and Falsification Checks

While the continuity assumption underlying the RD design, namely, that the only change that occurs at the cutoff is a shift in the treatment status, cannot be tested directly, there are several validation methods that provide indirect evidence about the validity of an RD design (De la Cuesta & Imai, 2016; Cattaneo et al., 2020). I conduct five validation and falsification checks that are often employed in the context of RD designs (e.g., Imbens & Lemieux, 2008; Ludwig & Miller, 2007). First, I investigate whether CEOs or firms are able to precisely manipulate (relative) stock returns to make them barely positive rather than barely negative. Second, I test whether firms that are near the cutoff, as well as their CEOs, have similar observable characteristics. Third, I test whether discontinuities exist at the artificial cutoff values used in the return functions. Fourth, I test the model's sensitivity to the bandwidth choice, and finally, I test its sensitivity to alternative specifications.

4.2.1 Manipulation Tests

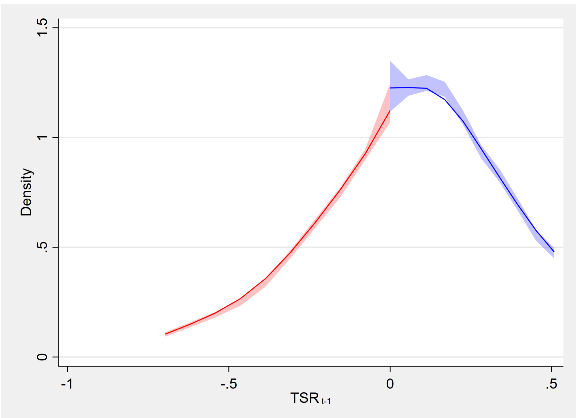
An major concern in the context of RD designs is that selection into the treatment group might exist. If CEOs or firms manage to precisely manipulate the stock returns such that the returns are barely positive rather than barely negative, the assignment into the

[14] Z-test of RD estimates without covariates: z-score = -1.23, p-value for one-tailed test = 0.11; z-test of RD estimates with covariates: z-score = -1.18, p-value for one-tailed test = 0.12.

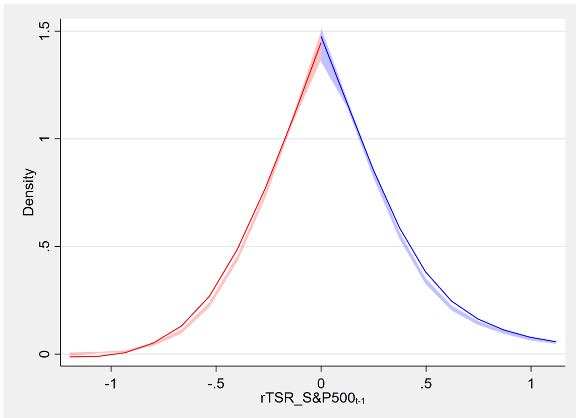
[15] Z-test of RD estimates without covariates: z-score = -1.81, p-value for one-tailed test = 0.04; z-test of RD estimates with covariates: z-score = 1.92, p-value for one-tailed test = 0.03.

treatment group would be non-random (Flammer & Bansal, 2017). However, I do not expect that such a precise manipulation of either stock returns or relative performance against the S&P 500 is likely. First, CEOs do not have direct control over stock prices and second, in the case of relative returns, CEOs do not have any control over S&P 500 index returns.

Nevertheless, I formally test the assumption of a smooth return function around the cutoff, employing a density test first proposed by McCrary (2008). Specifically, I follow the manipulation testing procedures developed by Cattaneo, Jansson, and Ma (2018), using local polynomial density estimators and testing the null hypothesis that the density of each return function is continuous around the cutoff of zero. Figure 2 displays the estimated densities.



(a) Smoothness around zero TSR_{t-1}



(b) Smoothness around zero $rTSR\_S\&P500_{t-1}$

Figure 2
Density of the returns around the cutoff of zero using a local quadratic approximation and robust bias-corrected test statistics.

Even though the estimated density shown in Figure 2a seems to be slightly higher just above a TSR_{t-1} of zero than just below this cutoff, this difference is not statistically significant (T = 1.00; p-value = 0.32). Regarding the results corresponding to $rTSR\_S\&P500_{t-1}$, which are displayed in Figure 2b, there is barely a difference visible in the estimated density and the difference is not statistically significant (T = 0.05; p-value = 0.96). Thus, this failure to reject the null hypothesis of the absence of a difference in these densities offers evidence supporting the validity of the RD design (Cattaneo et al., 2020).

4.2.2 Tests of Discontinuities for Predetermined Covariates

If CEOs or firms are unable to precisely influence (relative) total shareholder returns at the end of a fiscal year, the firms just above and just below the cutoff of zero (relative) returns should be similar in relation to all variables not affected by the treatment (Cattaneo et al., 2020). Thus, an important RD falsification test is to examine whether the firms with barely positive (relative) returns are similar to those with barely negative (relative) returns in terms of observable characteristics. If selection into the treatment is feasible, an imbalance of predetermined covariates would be expected (Ludwig & Miller, 2007).

Table 4 shows the results of discontinuity tests for various predetermined covariates. As before, I estimate the optimal bandwidth for each dependent variable separately and use robust bias-corrected methods for valid inference (Calonico et al., 2014). Panel A of Table 4 shows that the sizes and market values of the firms close to either side of the cutoff are similar for $t-1$ and that the RD estimates remain insignificant. Moreover, Columns (5) and (6) of Panel A show that the firms are similar in terms of firm performance, which is measured as the total shareholder return of each firm during $t-2$, as the RD estimates are close to zero and insignificant.

Panel B of Table 4 shows the results regarding differences in CEO characteristics, i.e., whether a firm's CEO is of retirement age, CEO tenure and whether a firm's CEO is the chairman of its board, among the firms around the cutoff. Again, all the RD estimates are insignificant, implying that the CEOs of the firms around the cutoff are similar with respect to these characteristics. Overall, these tests show that various predetermined firm and CEO characteristics are not discontinuous at the cutoff, further supporting the validity of the RD design.

Table 4

Tests of Discontinuities for Predetermined Covariates

Panel A: Firm Characteristics

	Ln(Total assets)$_{t-1}$		*Ln(Market value)*$_{t-1}$		TSR_{t-2}	
	(1)	(2)	(3)	(4)	(5)	(6)
Bias-corrected RD estimate	0.0171		0.0684		0.0040	
$c(TSR_{t-1} = 0)$	(0.0724)		(0.0663)		(0.0054)	
Bias-corrected RD estimate		-0.0723		0.0048		-0.0008
$c(rTSR\_S\&P500_{t-1} = 0)$		(0.0746)		(0.0685)		(0.0081)
Observations	27,952	27,952	27,949	27,949	27,896	27,896
Eff. Obs. left of cutoff	5,206	6,499	5,276	6,542	5,353	6,744
Eff. Obs. right of cutoff	7,397	6,744	7,489	6,821	7,616	7,024
Bandwidth	0.216	0.198	0.219	0.200	0.223	0.208
Bandwidth for bias estimate	0.432	0.350	0.371	0.317	0.373	0.358

Panel B: CEO Characteristics

	CEO age $\geq$*60*		*CEO tenure*		*CEO is chairman*	
	(1)	(2)	(3)	(4)	(5)	(6)
Bias-corrected RD estimate	-0.0064		-0.0512		0.0298	
$c(TSR_{t-1} = 0)$	(0.0194)		(0.3382)		(0.0240)	
Bias-corrected RD estimate		-0.0084		0.3720		0.0171
$c(rTSR\_S\&P500_{t-1} = 0)$		(0.0197)		(0.2965)		(0.0183)
Eff. Obs. left of cutoff	5,984	6,997	5,135	6,917	4,321	7,896
Eff. Obs. right of cutoff	8,731	7,278	7,290	7,192	5,820	8,198
Bandwidth	0.261	0.218	0.212	0.214	0.168	0.256
Bandwidth for bias estimate	0.421	0.353	0.329	0.371	0.272	0.436

Notes: This table reports the RD results of local linear regressions with triangular kernel weights and MSE-optimal bandwidths. The cutoff is equal to a (relative) return of zero. The RD estimates are bias-corrected following Calonico et al. (2014), with local quadratic regressions and MSE-optimal bandwidths for the bias estimators. The robust standard errors are reported in parentheses. There are 27,956 observations in Panel B. In all the models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

4.2.3 Tests of Discontinuities at Artificial Cutoff Values

In this section, I conduct a falsification analysis to test whether there are treatment effects at artificial cutoff values where no discontinuities are expected. I follow Imbens and Lemieux (2008) and investigate discontinuities at the median of the subsample with negative (relative) returns and at the median of the subsample with positive (relative) returns during the previous year. This ensures that only observations with the same treatment statuses are used for each artificial cutoff analysis (Cattaneo et al., 2020).

Table 5

Tests of Discontinuities at Artificial Cutoff Values

	Forced CEO Turnover			
	(1)	(2)	(3)	(4)
Robust bias-corrected RD estimate	0.0193			
c(TSR_{t-1} = -0.195)	(0.0172)			
Robust bias-corrected RD estimate		0.0007		
c(TSR_{t-1} = 0.278)		(0.0088)		
Robust bias-corrected RD estimate			0.0017	
c($rTSR\_S\&P500_{t-1}$ = -0.188)			(0.0161)	
Robust bias-corrected RD estimate				0.0114
c($rTSR\_S\&P500_{t-1}$ = 0.235)				(0.0126)
Observations	9,610	18,346	12,554	15,402
Eff. Obs. left of artificial cutoff	1,492	2,736	1,443	1,725
Eff. Obs. right of artificial cutoff	1,810	2,223	1,716	1,458
Bandwidth	0.087	0.092	0.060	0.065
Bandwidth for bias estimate	0.134	0.141	0.096	0.092

Notes: This table reports the RD results of local linear regressions with triangular kernel weights and MSE-optimal bandwidths. The RD estimates are bias-corrected following Calonico et al. (2014), with local quadratic regressions and MSE-optimal bandwidths for the bias estimators. Columns (1) and (3) only include observations below the real cutoff of zero. Colums (2) and (4) only include observations above the real cutoff of 0. The artificial cutoffs are equal to the medians of the two subsamples. The robust standard errors are reported in parentheses. In all the models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

Table 5 shows the results of these falsification tests using the estimation methods used for the analysis with the real cutoff of zero. All the estimates are statistically insignificant, suggesting that the likelihood of forced CEO turnover does not change discontinuously at these artificial cutoffs.

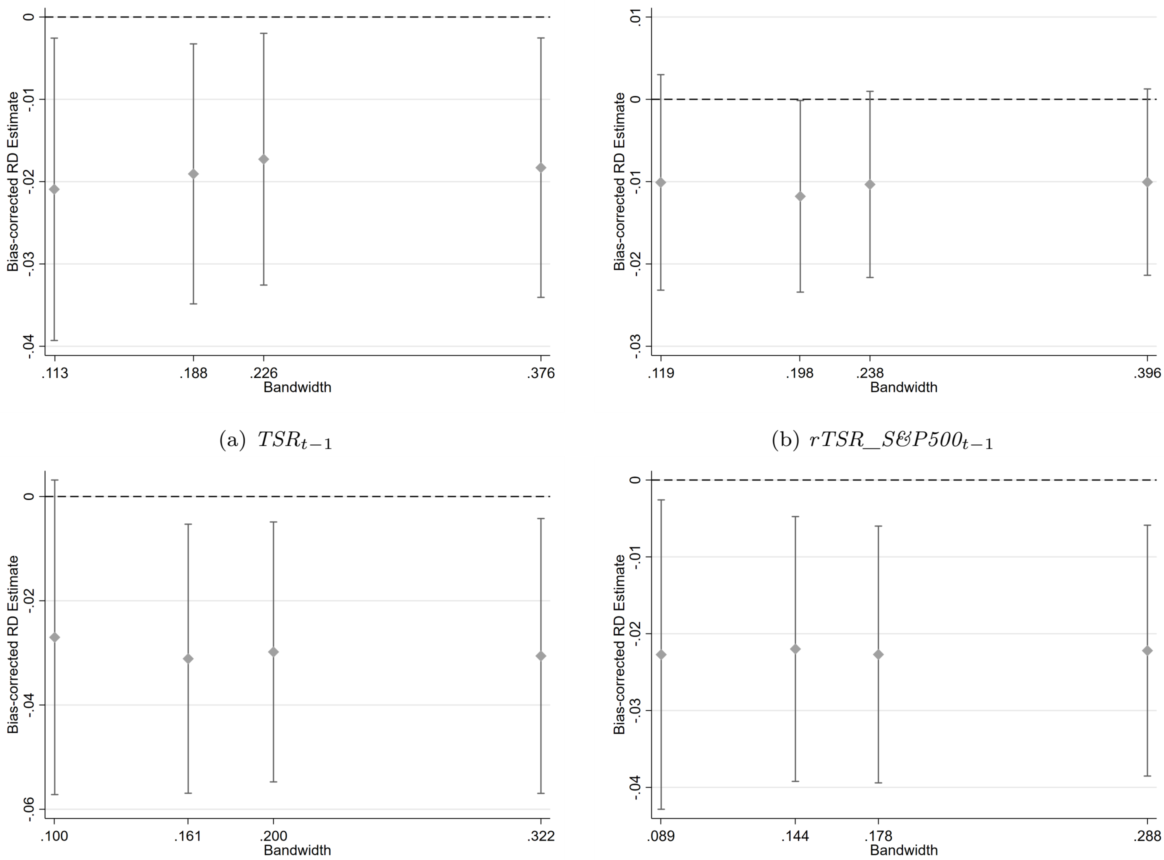
4.2.4 Sensitivity to Bandwidth Choice

Following Imbens and Kalyanaraman (2012), I investigate the sensitivity of the results using different bandwidths. A large bandwidth mechanically leads to a biased estimate, and a small bandwidth mechanically leads to increased confidence intervals due to the large variance of the estimate. Thus, investigating the sensitivity of the results to various bandwidth choices is meaningful only close to the MSE-optimal bandwidth and close to the bandwidth for which the coverage error (CER) of the robust bias-corrected confidence interval is minimized, i.e., the CER-optimal bandwidth (Cattaneo et al., 2020).

Following Cattaneo et al. (2020), I explore the sensitivity of the results to the CER-optimal bandwidth, the MSE-optimal bandwidth, twice the CER-optimal bandwidth, and twice the MSE-optimal bandwidth.[16] Figure 3 presents the bias-corrected RD estimates and robust 95% confidence intervals for each bandwidth without the covariates.[17] Figure 3a shows the estimates corresponding to a barely positive TSR_{t-1} for the whole sample period. As expected, the confidence intervals become longer for smaller bandwidths. However, the point estimates remain relatively stable between -1.7 and -2.1 percentage points and are statistically significant. Figure 3b displays the RD point estimates of a barely positive $rTSR\_S\&P500_{t-1}$. Even though the estimates remain stable at approximately -1.0 percentage points, the effect is only statistically significant at the 5% level for the MSE-optimal bandwidth choice of 0.198. For the larger bandwidths, the effect is significant at the 10% level, but it is insignificant for the CER-optimal bandwidth of 0.119.

Considering that the effect of a barely positive TSR_{t-1} is more pronounced in the sample period from 1993 to 2006 while the effect of a barely positive $rTSR\_S\&P500_{t-1}$ is

[16] The bandwidth for the bias estimate is fixed as the MSE-optimal bandwidth choice.
[17] The results remain similar if the covariates are included.



(a) TSR_{t-1}

(b) $rTSR\_S\&P500_{t-1}$

(c) TSR_{t-1} 1993 to 2006

(d) $rTSR\_S\&P500_{t-1}$ 2007 to 2018

Figure 3
Sensitivity to bandwidth selection of bias-corrected RD estimates with robust 95% confidence intervals.

more pronounced in the sample period from 2007 to 2018, I also evaluate the sensitivity of these results to bandwidth selection. Figure 3c and Figure 3d show that the point estimates remain similar and statistically significant for all but one of the bandwidths employed.

Overall, this analysis shows that the negative effect of a barely positive TSR_{t-1} is robust to different bandwidth choices, while the evidence for the effect of a barely positive $rTSR\_S\&P500_{t-1}$ is somewhat weaker for bandwidths other than the MSE-optimal choice. However, this is not surprising given that this effect is mainly driven by the later sample period, namely, 2007 to 2018. Analyzing this period separately reveals robust negative RD estimates for a barely positive $rTSR\_S\&P500_{t-1}$.

4.2.5 Sensitivity to Specification

Instead of a non-parametric regression discontinuity method, several empirical studies employ a parametric approach (e.g., Berger & Pope, 2011; Cunat, Gine, & Guadalupe, 2012; Flammer & Bansal, 2017). Even though Cattaneo et al. (2020) recommend against using global parametric RD methods, I test whether the main results are robust to various parametric specifications. In Column (1) of Table 6, only a dummy variable indicating whether the total shareholder return in the previous fiscal year is positive (Pos_TSR_{t-1}) and TSR_{t-1} itself are included in an OLS regression. In this specification, the point estimate is approximately -2.4 percentage points and statistically significant at the 1% level. Column (2) of Table 6 shows the results for the same specification using $rTSR\_S\&P500_{t-1}$. Again, the coefficient of a positive relative return with respect to the S&P 500 index ($Pos\_rTSR\_S\&P500_{t-1}$) is significantly negative and of a similar magnitude.

In Columns (3) and (4) of Table 6, I include the quadratic trend of the return variables and a full set of interaction terms, allowing for different slopes above and below the cutoff of zero. While the estimated coefficients decrease in magnitude in this specification, they remain negative and statistically significant. Thus, the main results are robust to using a global parametric approach and the estimated coefficients are of similar magnitudes, varying between -1.0 and -2.4 percentage points.

As an additional specification test, I analyze whether the non-parametric results are sensitive to the choice of the polynomial degree. By default, the bias-corrected RD estimate is calculated with a local linear regression for the point estimate and a local quadratic regression for the bias estimate (Calonico et al., 2014). The main results do not change substantially if second- or third-order polynomials are used for the point estimate in combination with a one-order higher polynomial for bias correction.

Table 6
Parametric Regression Discontinuity Design Specifications

	Forced CEO Turnover			
	(1)	(2)	(3)	(4)
Pos_TSR_{t-1}	-0.0240***		-0.0114**	
	(0.0024)		(0.0045)	
$Pos\_rTSR\_S\&P500_{t-1}$		-0.0221***		-0.0103***
		(0.0021)		(0.0039)
TSR_{t-1} / $rTSR\_S\&P500_{t-1}$	×	×	×	×
TSR^2_{t-1} / $rTSR\_S\&P500^2_{t-1}$			×	×
Interactions			×	×
Observations	27,956	27,956	27,956	27,956
R^2	0.01	0.01	0.01	0.01

Notes: This table reports the results of the use of a global parametric approach. The heteroskedasticity-robust standard errors are reported in parentheses. In all the models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

4.3 Alternative Explanations

I find robust empirical evidence that firms with barely positive stock returns and barely positive outperformances of the S&P 500 index are significantly less likely to dismiss their CEOs than those with barely negative return outcomes. It is difficult to explain these findings as the result of inferior CEO quality and performance or different firm characteristics around the cutoff, because such firms seem to be equal in terms of size, market value and stock performance and their CEOs seem to be equal in terms of tenure, whether they are of retirement age and whether they are chairmen of their firms' boards (see Section 4.2.2). Additionally, my results do not suggest that a precise selection into the treatment group of positive (relative) returns is possible. Thus, I attribute my findings to the outcome bias of boards of directors when they evaluate the performance of CEOs and decide whether to retain or dismiss them.

However, several alternative explanations could lead to similar empirical patterns. First, because CEOs also observe return outcomes, their observations of these outcomes could impact their effort and performance in the subsequent year. For example, if a firm's stock return is barely positive, or the firm barely outperforms a benchmark, its CEO might become more motivated and performs better than he or she would have if the (relative) return had been barely negative. Thus, the boards' turnover decisions might not be biased and could be in accordance with the informativeness principle. If this explanation is the main driver of my results, one would expect to see increased firm performance among firms with barely positive return outcomes over those with barely negative return outcomes in the subsample of CEOs that remained in office.[18]

Columns (1) and (2) of Table 7 show the RD estimates of a barely positive TSR_{t-1} and a barely positive $rTSR\_S\&P500_{t-1}$ for subsequent firm performance measured as TSR using the subsample of firm-years without CEO turnover in the focal year. Neither of these RD estimates is statistically significant, suggesting that the CEOs of firms with barely positive (relative) return outcomes do not perform systematically better in the subsequent year than the CEOs of firms with barely negative (relative) return outcomes.

A different alternative explanation for the examined phenomenon might be that CEOs of firms that perform just above the cutoff are more likely to leave their firms voluntarily. Thus, some CEOs who would have been fired in the subsequent year leave voluntarily, resulting in an empirical pattern that indicates that CEOs who perform just below the cutoff are more likely to be dismissed. If this explanation holds, one would expect to observe more voluntary turnover when the return outcome in the previous year was barely positive. However, the results in Columns (3) and (4) of Table 7 show that voluntary CEO turnover is not more likely after barely positive return outcomes than after barely negative return outcomes.[19]

Finally, one might argue that shareholders know when a CEO is of lower quality and thus that the stock returns of such firms are lower. While this might be true in general,

[18] If a firm's CEO changed during the focal year, the firm's performance cannot be unambiguously attributed to the departing CEO.

[19] I define CEO turnover as voluntary if the departure of a CEO is categorized as either voluntary due to retirement (*departure_code* = 5) or voluntary due to a new opportunity (*departure_code* = 6) in the dismissal database constructed by Gentry et al. (2021a).

Table 7
Alternative Explanations

	TSR		*Voluntary CEO Turnover*	
	(1)	(2)	(3)	(4)
Bias-corrected RD estimate	0.0269		-0.0146	
c($TSR_{t-1} = 0$)	(0.0197)		(0.0123)	
Bias-corrected RD estimate		0.0109		0.0140
c($rTSR\_S\&P500_{t-1} = 0$)		(0.0181)		(0.0123)
Observations	27,167	27,167	30,028	30,028
Eff. Obs. left of cutoff	4,552	5,704	5,915	6,650
Eff. Obs. right of cutoff	6,427	5,932	8,366	6,877
Bandwidth	0.191	0.173	0.229	0.184
Bandwidth for bias estimate	0.319	0.285	0.382	0.328

Notes: This table reports the RD results of local linear regressions with triangular kernel weights and MSE-optimal bandwidths. The RD estimates are bias-corrected following (Calonico et al., 2014), with local quadratic regressions and MSE-optimal bandwidths for the bias estimators. The cutoff is equal to a (relative) return of zero. The robust standard errors are reported in parentheses. Columns (1) and (2) are based on the subsample of firm-years without CEO turnover. In all the models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

shareholders are unable to precisely control stock returns such that only firms with lower-quality CEOs face (relative) total shareholder returns just below zero at the end of a fiscal year. This is confirmed through the manipulation tests detailed in Section 4.2.1, which do not indicate a discontinuity in the density of the (relative) returns around the cutoff. Moreover, it is unclear why shareholders would time the returns just at the end of a fiscal year. Thus, a precise sorting of firms around the cutoff based on CEO quality by shareholders seems unlikely.

5 Conclusion

In this paper, I examine whether boards follow the informativeness principle and filter out uninformative performance signals in their forced CEO turnover decisions. I find that a firm with a barely positive annual stock return in the previous fiscal year is approximately 50% less likely to exhibit forced CEO turnover than one with a barely negative stock return, although the outcome of a barely positive or barely negative return can be viewed as random. I observe a similar pattern in relation to firms' stock returns relative to the S&P 500 index return: a board is less likely to dismiss a CEO if the firm barely outperformed the S&P 500 index than if the firm barely underperformed the S&P 500 index. In exploring the development of this effect over time, I demonstrate that the tendency of boards to consider uninformative absolute return outcomes has decreased over time, while their tendency to consider uninformative relative return outcomes has increased over time. This suggests that boards have shifted their focus toward relative returns but nevertheless have continued to violate the informativeness principle.

Although the RD design of this study facilitates a credible analysis of the relevant causal effects, the estimates represent a local average treatment effect for firms with (relative) returns close to the cutoff of zero. Thus, the estimates do not necessarily apply to firms with (relative) returns very different from zero. However, analyzing the firms' absolute and relative returns separately facilitates the estimation of two local average treatment effects that are located at different points of the return distribution: the cutoff of zero absolute returns is located in the second firm performance quintile, while the cutoff of zero relative returns is located in the third firm performance quintile. Thus, even though the firms generating returns of approximately zero might be substantially different from those generating relative returns of approximately zero, my results suggest that these firms' boards consider uninformative (relative) return outcomes in a similar manner.

Another limitation is that I do not identify the performance metrics on which individual firms most prominently focus in their their performance evaluations of CEOs. If boards use only performance metrics other than the (relative) returns employed in this paper, they would remain unaware of the associated return outcomes, and no discontinuity would be expected in the probability of forced turnover. As I find a discontinuity at zero absolute returns and at zero relative returns with respect to the S&P 500 index, boards must also care about these performance metrics to some extent. However, examining the most salient performance evaluation metrics considered by individual boards may reveal even more pronounced outcome bias, which could be explored in future research.

References

Albuquerque, A. (2009). Peer firms in relative performance evaluation. *Journal of Accounting and Economics*, *48*(1), 69-89.

Amore, M. D., & Schwenen, S. (2020). The value of luck in the labor market for CEOs. *Working Paper*.

Antle, R., & Smith, A. (1986). An empirical investigation of the relative performance evaluation of corporate executives. *Journal of Accounting Research*, *24*(1), 1–39.

Bakke, T.-E., Mahmudi, H., & Newton, A. (2020). Performance peer groups in CEO compensation contracts. *Financial Management*, *49*(4), 997-1027.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, *54*(4), 569.

Barro, J. R., & Barro, R. J. (1990). Pay, performance, and turnover of bank CEOs. *Journal of Labor Economics*, *8*(4), 448-481.

Berger, J., & Pope, D. (2011). Can losing lead to winning? *Management Science*, *57*(5), 817-827.

Brav, A., Jiang, W., Partnoy, F., & Thomas, R. (2008). Hedge fund activism, corporate governance, and firm performance. *The Journal of Finance*, *63*(4), 1729-1775.

Brechot, M., & Flepp, R. (2020). Dealing with randomness in match outcomes: How to rethink performance evaluation in european club football using expected goals. *Journal of Sports Economics*, *21*(4), 335-362.

Brownback, A., & Kuhn, M. A. (2019). Understanding outcome bias. *Games and Economic Behavior*, *117*, 342-360.

Bushman, R., Dai, Z., & Wang, X. (2010). Risk and CEO turnover. *Journal of Financial Economics*, *96*(3), 381-398.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, *101*(3), 442–451.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295-2326.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2020). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), 234-261.

Cunat, V., Gine, M., & Guadalupe, M. (2012). The vote is cast: The effect of corporate governance on shareholder value. *The Journal of Finance*, *67*(5), 1943-1977.

De la Cuesta, B., & Imai, K. (2016). Misunderstandings about the regression discontinuity design in the study of close elections. *Annual Review of Political Science*, *19*, 375–396.

Dikolli, S. S., Hofmann, C., & Pfeiffer, T. (2013). Relative performance evaluation and peer-performance summarization errors. *Review of accounting studies*, *18*(1), 34–65.

Eisfeldt, A. L., & Kuhnen, C. M. (2013). CEO turnover in a competitive assignment framework. *Journal of Financial Economics*, *109*(2), 351-372.

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

Fee, C. E., Hadlock, C. J., Huang, J., & Pierce, J. R. (2018). Robust models of CEO turnover: New evidence on relative performance evaluation. *Review of Corporate Finance Studies*, *7*(1), 70–100.

Flammer, C., & Bansal, P. (2017). Does a long-term orientation create value? evidence from a regression discontinuity. *Strategic Management Journal*, *38*(9), 1827-1847.

Flepp, R., & Franck, E. (2021). The performance effects of wise and unwise managerial dismissals. *Economic Inquiry*, *59*(1), 186-198.

Gauriot, R., & Page, L. (2019). Fooled by performance randomness: overrewarding luck. *Review of Economics and Statistics*, *101*(4), 658–666.

Gentry, R. J., Harrison, J., Quigley, T., & Boivie, S. (2021b, February). *Open sourced database for ceo dismissal 1992-2018.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.4543893`

Gentry, R. J., Harrison, J. S., Quigley, T. J., & Boivie, S. (2021a). A database of CEO turnover and dismissal in S&P 1500 firms, 2000 - 2018. *Strategic Management Journal*, *42*(5), 968-991.

Gibbons, R., & Murphy, K. J. (1990). Relative performance evaluation for chief executive officers. *Industrial and Labor Relations Review*, *43*(3), 30S–51S.

Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless+harmless=blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organizational Behavior and Human Decision Processes*, *111*(2), 93-101.

Gong, G., Li, L. Y., & Shin, J. Y. (2011). Relative performance evaluation and related peer groups in executive compensation contracts. *The Accounting Review*, *86*(3), 1007–1043.

Gopalan, R., Milbourn, T., & Song, F. (2010, 01). Strategic Flexibility and the Optimality of Pay for Sector Performance. *The Review of Financial Studies*, *23*(5), 2060-2098.

Guo, L., & Masulis, R. W. (2015). Board structure and monitoring: New evidence from ceo turnovers. *The Review of Financial Studies*, *28*(10), 2770–2811.

Gurdal, M. Y., Miller, J. B., & Rustichini, A. (2013). Why blame? *Journal of Political Economy*, *121*(6), 1205-1247.

Hazarika, S., Karpoff, J. M., & Nahata, R. (2012). Internal corporate governance, CEO turnover, and earnings management. *Journal of Financial Economics*, *104*(1), 44-69.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423-1465.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, *10*(1), 74–91.

Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324–340.

Huang, S., Maharjan, J., & Thakor, A. V. (2020). Disagreement-induced CEO turnover. *Journal of Financial Intermediation*, *43*, 100819.

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, *79*(3), 933–959.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635.

Jayaraman, S., Milbourn, T. T., Peters, F. S., & Seo, H. (2020). Product market peers and relative performance evaluation. *The Accounting Review, forthcoming*.

Jenter, D., & Kanaan, F. (2015). CEO turnover and relative performance evaluation. *Journal of Finance*, *70*(5), 2155–2184.

Kaplan, S. N., & Minton, B. A. (2012). How has CEO turnover changed? *International Review of Finance*, *12*(1), 57–87.

Kausel, E. E., Ventura, S., & Rodríguez, A. (2019). Outcome bias in subjective ratings of performance: Evidence from the (football) field. *Journal of Economic Psychology*, *75*, 102132.

König-Kersting, C., Pollmann, M., Potters, J., & Trautmann, S. T. (2021). Good decision vs. good results: Outcome bias in the evaluation of financial agents. *Theory and Decision*, *90*(1), 31–61.

Lefgren, L. J., Platt, B., & Price, J. (2015). Sticking with what (barely) worked: a test of outcome bias. *Management Science*, *61*(5), 1121–1136.

Ludwig, J., & Miller, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, *122*(1), 159-208.

Ma, P., Shin, J.-E., & Wang, C. C. (2021). rtsr: Properties, determinants, and consequences of benchmark choice. *Working Paper*.

Marshall, G. W., & Mowen, J. C. (1993). An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling & Sales Management*, *13*(3), 31-47.

Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, *26*(2-3), 131-146.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

Park, H., & Vrettos, D. (2015). The moderating effect of relative performance evaluation on the risk incentive properties of executives' equity portfolios. *Journal of Accounting Research*, *53*(5), 1055-1108.

Parrino, R. (1997). CEO turnover and outside succession a cross-sectional analysis. *Journal of Financial Economics*, *46*(2), 165–197.

Peters, F. S., & Wagner, A. F. (2014). The executive turnover risk premium. *The Journal of Finance*, *69*(4), 1529-1563.

Rubin, J., & Sheremeta, R. (2016). Principal-agent settings with random shocks. *Management Science*, *62*(4), 985-999.