HARVARD | BUSINESS | SCHOOL

Relative Performance Benchmarks: Do Boards Follow the Informativeness Principle?

Paul Ma
Jee Eun Shin
Charles C.Y. Wang

Working Paper 17-039

Relative Performance Benchmarks: Do Boards Follow the Informativeness Principle?

Paul Ma
University of Minnesota

Jee Eun Shin
University of Toronto

Charles C.Y. Wang
Harvard Business School

Working Paper 17-039

Relative Performance Benchmarks: Do Boards Follow the Informativeness Principle?*

Paul Ma
University of Minnesota
Carlson School of Management

Jee-Eun Shin
University of Toronto

Charles C.Y. Wang
Harvard Business School

October 4th, 2018

Abstract

We examine whether and to what extent managers are evaluated, in their relative-performance contracts, on the basis of systematic performance. Focusing on relative total shareholder returns (rTSR), the predominant metric specified in these contracts and used by market participants to evaluate managers, we document that 60% of firms—those that choose specific peers—do a remarkable job of capturing the systematic component of returns in adherence to the informativeness principle. However, 40% of firms—those that choose index-based benchmarks—retain substantial systematic noise in their rTSR metrics, which they could have substantially corrected by using their self-chosen compensation-benchmarking peers. The selection of noisy benchmarks is economically important, is associated with lower ROA, and can be explained by compensation consultants' tendencies and firms' governance weaknesses.

JEL: G30, J33, M12, M52

Keywords: Executive compensation; relative performance evaluation; relative TSR; common shock filtration; systematic risk; search-based peers; board of directors; corporate governance; compensation consultants

---

*This paper was previously circulated under the title "Relative Performance Benchmarks: Do Boards Get it Right?". The authors can be contacted at paulma@umn.edu, jee-eun.shin@rotman.utoronto.ca, and charles.cy.wang@hbs.edu. We have benefited from advice and suggestions from Hengjie Ai, Santiago Bazdresch, Alan Benson, Brian Cadman (discussant), Gerardo Perez Cavazos, Shane Dikolli (discussant), Vyacheslav Fos (discussant), Frank Gigler, Guojin Gong, Thomas Hemmer, Sudarshan Jayaraman, George John, Chandra Kanodia, Heedong Kim (discussant), David Larcker, Eva Liang, Melissa Martin, Kevin Murphy, Jonathan Nam, Gaizka Ormazabal (discussant), Xiaoxia Peng, Ruidi Shang, Matt Shaffer, John Shoven, Pervin Shroff, Martin Szydlowski, Chao Tang, Akhmed Umyarov, Felipe Varas, Martin Wu, Eric Yeung (discussant), Ivy Zhang and conference and seminar participants at FARS, Georgia State University, Harvard Business School, HKUST Accounting Symposium, SMU Accounting Symposium, UBC Winter Finance Conference, University of Florida, the Utah Winter Accounting Conference, and the American Accounting Association Annual Meeting. We give thanks to Srikant Datar and Paula Price from Harvard Business School, Yulan Shen (Head of Executive Compensation) of Biogen, Barry Sullivan (Managing Director) of Semler Brossy, Terry Adamson (Partner), Jon Burg (Partner), and Daniel Kapinos (Associate Partner) of Aon Hewitt, and to Nikhil Lele (Principal) and Trent Tishkowski (Principal) of Ernst and Young for clarifying institutional details and for helpful feedback. We also thank Kyle Thomas and Raaj Zutshi for outstanding research assistance. Paul Ma is grateful for research support from the University of Minnesota Accounting Research Center and the Dean's Small Research and Travel Grant. All errors remain our own.

1 Introduction

Measurement of performance is a critical element in managerial evaluation and in the design of incentives. A classic insight of principal-agent theory specifies that the principal—in practice, the board—should evaluate managers' performance using performance metrics that are informative about effort or talent ("the informativeness principle") (Holmström, 1979; Shavell, 1979). In particular, when a performance metric contains common noise that is beyond the CEO's control, a metric of relative performance that filters out such noise can be desirable in two ways. First, designing explicit incentive contracts around such relative performance metrics should increase the efficacy of risk sharing and help elicit costly unobservable effort from risk-averse managers. Second, greater precision in measurement of non-systematic performance can increase the agent's *implicit* incentive (e.g., due to career concerns) to exert additional effort (Holmström, 1999).

In keeping with this insight, a large empirical literature has devoted itself to understanding whether and to what degree corporate managers are evaluated and rewarded on the basis of the systematic and non-systematic components of firm performance (e.g., Antle and Smith, 1986; Janakiraman, Lambert and Larcker, 1992). However, a significant challenge to this literature has been the limited visibility to researchers of the actual relative performance metrics chosen by firms—including the metric itself as well as the peers used to isolate the non-systematic component of it (as discussed in, e.g., Albuquerque, 2009)—due to limited disclosure of such information in public proxy statements.[1] As a result, the interpretation of the prior literature's findings has been limited and mixed.

However, the recent shifts in the corporate governance landscape and SEC disclosure reforms provide new opportunities to examine this classic and outstanding question. Over the last ten years, relative performance measures have come to dominate the corporate governance landscape, calling for a fresh examination of the phenomenon. In particular, relative total shareholder returns (rTSR)—that is, a firm's own TSR relative to that of an index or a group of peer firms—appears to

[1] As a result, much of the empirical literature has been forced to rely on indirect tests—tests of whether variation in executive compensation is explained by variation in firm performance relative to some measure of "systematic" performance selected by the researcher (e.g., industry or competitive peers)—that have produced mixed findings. For example, Albuquerque (2009) summarizes the evidence for and against relative performance evaluation in CEO compensation and turnover, and argues that the mixed findings in the literature are likely the result of the joint hypothesis problem: that is, researchers are testing both the theory and the choice of an empirical metric to measure the systematic component of firm performance.

have become the single most widely used performance metric by which market participants judge companies and their executives. For example, since 2006 the SEC has required firms to disclose rTSR in their annual reports to shareholders.[2] The growing preference for rTSR as a performance metric is also evident in the trend towards linking rTSR to performance-based executive contracts: the proportion of the largest 1,500 firms in the market with explicit relative performance incentives has increased from 20% in 2006 to 48% in 2014; among them, the use of rTSR has increased from 70% to 87% over the same time period (see Figure 1). Finally, rTSR plays an economically significant role in executive compensation: our back-of-the-envelope estimates suggest that meeting rTSR targets increases the CEO's incentive-plan-based compensation by an average of 40%.

Given the predominance of rTSR, an important question is to what extent firms' chosen rTSR metrics evaluate managers on the systematic or non-systematic component of firm performance. The SEC's 2006 disclosure reforms—which requires firms to disclose the details of their performance-based incentives, including relative performance metrics and the peer benchmarks used by the firms—provide researchers with new data to examine such a question. In conjunction with the rising importance of relative performance metrics in corporate governance, this reform therefore provides an exciting and new opportunity to distill new insights into the evaluation of managerial performance, by allowing for *direct* assessment of the properties of firms' chosen relative performance metrics.

In this paper, we exploit the disclosure reform to examine the relative performance evaluation metrics—in particular, relative shareholder returns (rTSR)—that firms choose to evaluate managers in their performance-based contracts. We assess the extent to which firms' chosen peer-return benchmarks yield rTSR measures that evaluate managers on the basis of the systematic and non-systematic components of firm performance. To guide our analysis, we exploit the linear factor structure of firm performance that is implicit in the definition of relative performance metrics like rTSR and explicit in the standard theoretical framework (e.g., $p = 1 \times c + e$ where c is the systematic

[2]The New York Stock Exchange's Listing Company Manual (Section 303A.05) recommends that compensation committees consider a firm's rTSR in determining long-run executive incentives. The influential proxy advisory firm Institutional Shareholder Services (ISS) relies on an analysis of the relationship between a firm's rTSR and its executive's pay relative to peers to judge whether an executive's compensation is justified by performance and to formulate its say-on-pay recommendations. Activist investors often focus on poor rTSR as evidence of poor management quality or poor performance (Brav, Jiang, Partnoy and Thomas, 2008). Finally, as part of its implementation of the 2010 Dodd-Frank Act, the SEC recently proposed Rule No. 34-74835 requiring firms to disclose comparisons of executive compensation to that of peers in terms of rTSR in annual proxy statements.

or common component of performance p and e is the non-systematic or idiosyncratic component), under which the common component of firm performance should exhibit two properties. First, the structure of the relative performance metric ($p - 1 \times c$) implies that a firm's performance should have a slope ("benchmark-return-beta") of 1 with respect to the common component. Second, the common component should capture all of the contemporaneous variation in firm performance that is systematic (i.e., non-idiosyncratic to the firm). We examine to what extent firms' self-chosen rTSR peer benchmarks satisfy these theoretical properties and *if not*, the sources of economic frictions that could explain the difference.

Our analyses produce four main findings. First, with regards to the first property, we find that all firms on average choose rTSR benchmarks that exhibit benchmark-return-betas of 1, consistent with the benchmarks measuring the systematic component of firm returns. However, with regards to the second property, these benchmarks on average produce rTSR metrics that continue to retain some degree of systematic noise in that alternative benchmarks generate larger time-series R^2s from contemporaneous return regression than the chosen RP benchmarks. Specifically, we compare firms' chosen rTSR benchmarks to three normative peer benchmarks: the S&P500, firms' self-chosen compensation-benchmarking peers, and search-based peers (SBPs).[3] Based on these comparisons, we document that firms' disclosed benchmarks (1) outperform the S&P500 index by 47% in terms of time-series R^2s from contemporaneous-return regressions; (2) perform equally well relative to compensation-benchmarking peers; and (3) underperform SBPs by about 7%.

Our second set of findings reveals a significant difference between the two predominant approaches to selecting rTSR benchmarks: (a) based on a customized set of peer firms ("specific peers") and (b) based on an industry or market index. We find that the underperformance of firms' chosen benchmarks is concentrated in the 40% of firms that choose index-based benchmarks: compared to the index-based benchmarks, SBPs explain 16% more of the time-series variation in firms' monthly stock returns (i.e., time-series regression R^2); and, surprisingly, firms' chosen compensation-benchmarking peers also perform 8% better than the index-based benchmarks. These findings raise questions about the appropriateness of choosing indexes—effectively, benchmarking against a large number of peers—in lieu of a narrower but possibly more relevant peer set (e.g., firms'

[3] Lee, Ma and Wang (2015) and Lee, Ma and Wang (2016) show that SBPs are superior to other state-of-the-art peer identification schemes at explaining variation in firms' stock returns, valuation multiples, and fundamental performance characteristics.

self-chosen compensation-benchmarking peers that are already available). On the other hand, the approximately 60% of firms that use specific peers fare well relative to the normative benchmarks. For example, SBPs only outperform firms' chosen specific peers by 2%; moreover, firms' chosen compensation-benchmarking peers underperform the specific peers chosen for rTSR purposes by over 4%. Viewing SBPs as a normative upper bound in capturing systematic performance, these results indicate that the majority of boards are remarkably efficient at isolating the non-systematic component of performance through their chosen rTSR benchmarks.

Our third set of findings shows that the R^2 differences in firms that choose index-based benchmarks are economically meaningful. We provide an economic interpretation of the R^2 differences in the context of a baseline principal-agent framework absent any frictions in peer selection. For any set of performance benchmarks, this framework enables us to translate our R^2 results to the variance in measurement errors (for the systematic component of performance) resulting from the chosen peers (up to a scalar constant). More importantly, this framework allows us to calibrate the performance consequences of counter-factual rTSR benchmarks. Our results suggest that, in the absence of frictions against selecting a precise set of peers, measurement errors due to firms' choices of performance benchmarks reduce managerial effort and result in an on-average performance penalty of 60 to 153 basis points in annual stock returns, under plausible values of risk aversion. Again, these effects are driven by the subset of firms that use index-based benchmarks, for which we find an on-average penalty of 106-277 bps in annual stock returns.

Our final set of findings stem from an exploratory analysis to better understand why certain firms' boards may have chosen apparently poorer rTSR benchmarks. Generalizing from the baseline principal-agent framework, boards may in theory select rTSR benchmarks that exhibit lower R^2 due to various sources of economic frictions: for example, when the firm has greater idiosyncratic volatility, or when board-level governance is weak. We also examine a number of plausible alternative theories offered by the prior literature that may explain why certain firms would select index-based benchmarks: (1) firms' own actions influencing peer performance (Janakiraman, Lambert and Larcker, 1992; Aggarwal and Samwick, 1999a); (2) managers' ability to self-insure against the systematic factor (Garvey and Milbourn, 2003); (3) the provision of tournament incentives (Lazear and Rosen, 1981; Hvide, 2002); (4) boards trading off ex-ante vs. ex-post efficiency due to the perception that indexes are less gameable (Godfrey and Bourchier, 2015; Walker, 2016); (5) the selection benchmarks on the basis of aspiration (Scharfstein and Stein, 1990; Hayes and Schaefer, 2009; Hemmer, 2015; Francis, Hasan, Mani and Ye, 2016); or (6) firms facing differential implementation costs between specific peers and index-based benchmarks.

Our exploratory empirical analysis suggests that the observed selection of index-based benchmarks is probably a result of governance-related frictions and compensation consultants' tendencies. We argue that the other alternative theories are unlikely to drive our main findings. Instead, the selection of index-based benchmarks is systematically associated with proxies for governance weaknesses—such as excess compensation, larger board size, and heavier director workload—and compensation-consultant-fixed effects. Our analysis also suggests that weaker boards' failure to carefully scrutinize the default recommendations of a compensation consultant result in managerial evaluation based on systematic performance. Finally, in reduced-form regression analysis we find that the choice of an index-based benchmark is cross-sectionally associated with lower future ROA, consistent with the calibration exercise; we find similar results using annual stock returns. These findings are consistent with the selection of an index being a symptom of corporate governance weaknesses that is correlated with future performance and that is not subsumed by existing measurements of corporate governance quality used in the literature. An implication is that the properties of performance metrics chosen by boards to evaluate CEOs is an indicator of board monitoring quality.

Our results make several contributions to the literature. First, we provide new evidence on the properties of rTSR measures chosen by boards to evaluate executives in relative-performance contracts. Our evidence is important in the context of the rising importance of rTSR for evaluating managerial performance.

Second, we contribute direct evidence (i.e., by directly observing and analyzing the specific performance measures chosen by firms) to the literature that examines the extent to which managers are evaluated on the basis of systematic versus non-systematic components of performance (Antle and Smith, 1986; Albuquerque, 2009). Our findings that over half of firms do a remarkable job of measuring and filtering for systematic risk in TSR, consistent with the informativeness principle, is particularly surprising in light of the prevailing view that the executive pay-setting and evaluation process is broken and compromised by powerful executives (Bebchuk, Cremers and Peyer, 2011; Morse, Nanda and Seru, 2011). It is perhaps also surprising that, in spite of the significant theoretical advances over the last four decades, e.g., why directors may not wish to fully capture and filter out the systematic component of performance in evaluating CEOs (e.g., Janakiraman, Lambert and Larcker, 1992; Aggarwal and Samwick, 1999a; DeMarzo, Fishman, He and Wang, 2012; Eisfeldt and Kuhnen, 2013), the choice of rTSR measures for the majority of firms remains consistent with the predictions of the classical model in Holmström (1979).

Third, we provide novel evidence on how compensation consultants impact the managerial evaluation and pay-setting process. Our results show how compensation consultants' preferences interacting with board-level governance weaknesses can explain the observed evaluation of managers on the basis of systematic performance in many firms. Finally, our findings suggest that scrutinizing the quality of performance metrics chosen by boards for monitoring the performance of executives can be a useful way to "monitor the monitor."

Together, the empirical findings of our work adds to the body of research that has emerged since the 2006 mandate on compensation benchmarking practices. Examining only firms with specific peers as relative performance benchmarks, Gong, Li and Shin (2011) suggests that firms' chosen relative-performance benchmarks perform better than randomly chosen benchmarks in eliminating common shocks, but the analysis does not investigate the overall extent to which firms' benchmarking practices adhere to the informativeness principle and its implications.[4] Complementing our results, Bizjak, Kalpathy, Li and Young (2016) focuses on the effect of the choice of performance benchmarks on CEO compensation levels, and finds that boards' selection of performance peers does not affect the level of pay in an economically significant manner.

The rest of the paper is organized as follows. Section 2 lays out data and descriptive statistics illustrating the rise of explicit grant-based relative-performance benchmarking; it also provides empirical evidence on the properties of firms' chosen rTSR benchmarks. Section 3 maps our empirical test to the principal-agent framework in order to recover primitives that allow us to calibrate the implications of relative-performance benchmark properties for firm performance. Section 4 explores alternative theories about why certain firms choose index-based benchmarks. Section 5 investigates the reduced form cross-sectional association between index selection and future ROA. Section 6 concludes.

[4] A related set of papers tests for whether a firm's choice in tying incentives to relative performance, on the extensive margin, is consistent with its costs and benefits as predicted in Gibbons and Murphy (1990). Carter *et al.* (2009) studies a sample of UK firms and finds that the propensity to tie incentives to relative performance—identified via explicit disclosures—is not associated with the degree of a firm's exposure to systematic shocks. Gong *et al.* (2011) finds the opposite result in the US.

2 Data and Descriptive Evidence of Benchmarking Behavior

This section provides empirical evidence on the quality of firms' chosen relative-performance benchmarks, in terms of the extent to which they follow the informativeness principle and eliminate common performance shocks. Our analyses focus on the sample of firms that explicitly tie executive compensation to rTSR; the quality or informativeness of rTSR is expected to be of greater importance to such firms.

2.1 Data Description

Our data come from ISS Incentive Lab, which collected details on compensation contracts and incentive-plan-based awards of named executive officers, at the individual-grant level, from firms' proxy statements. Incentive Lab covers every U.S. firm ever ranked in the top 750 in terms of market capitalization in any year since 2004. Due to backward- and forward-filling, the raw Incentive Lab data (2004-2014) encompasses the entire S&P500, most of the S&P Midcap 400, and a small proportion of the S&P Small-Cap 600. Thus, roughly speaking, each annual cross-section encompasses the largest 1,000 firms listed on the U.S. stock market in terms of market capitalization. Our analysis focuses on the sample from 2006 onward; mandatory disclosure of compensation details began in 2006, and coverage of firms is more comprehensive after that year.

For each grant, ISS Incentive Lab collected information on the form of the payout (cash, stock options, or stock units); conditions for payout (tenure [Time], fulfillment of absolute performance criteria [Abs], relative performance criteria [Rel], or a combination of the two [Abs/Rel]); and specific accounting- or stock-based performance metrics associated with performance-based grants. Finally, ISS Incentive Lab collected information on the specific peer firms or indexes selected for purposes of awarding grants based on relative performance.

Table 1, Panel A, provides summary statistics on 34,321 CEO grants awarded by 1,547 unique firms in the 2006-2014 period. During this period, on average, companies awarded 3.2 CEO grants per year. The proportion of incentive awards paid out in cash is stable within the sample period at roughly 35% of all CEO grants; in the same time period, stock-based payouts increased from 36% to 49% while option-based payouts declined from 29% to 15%. Notably, the proportion of CEO grants that included a relative performance component (Abs/Rel or Rel) more than doubled, from

8% in 2006 to 17% in 2014.[5]

Table 1, Panel B, suggests that, at the firm level, usage of relative performance incentives has more than doubled since 2006. Relative to the total number of companies in our sample, the proportion of firms with explicit relative performance (RP) incentives increased from 20% in 2006 to 48% in 2014 (see the solid line in Figure 1). Moreover, Panel C suggests that the use of rTSR has been increasingly prevalent at such firms: whereas 70% of the companies that provide RP incentives used rTSR in 2006, 87% did so by 2014 (see the dashed line in Figure 1). Jointly, the summary statistics presented in Table 1 and Figure 1 illustrate the increasing pervasiveness of explicit RP-based incentives and the prominence of rTSR in such incentive plans.

To further assess the economic magnitude of CEOs' RP-based incentives, Table 2 provides back-of-the-envelope estimates of the relative importance of meeting RP targets. We estimate how much incremental incentive-plan-based compensation the CEO would earn by meeting RP-based targets, assuming that all other incentives are earned. Column 3 estimates expected total plan-based compensation when all incentives are earned, including meeting all RP-based targets.[6] Columns 4 and 5 estimate the allocated expected compensation stemming from meeting RP-based targets and from meeting rTSR-based targets respectively. Overall, RP-based incentives comprise a significant proportion of the total expected plan-based compensation, with rTSR accounting for the vast majority of RP-based incentives (73% on average as reported in column 6).

We also estimate the improvement in incentive-plan-based compensation expected from meeting RP-based and rTSR-based targets (the "Incentive Ratio" reported in columns 7 and 8). Column 7 suggests that, relative to not meeting RP-based targets, meeting them increases CEOs' plan-based compensation by an average of 58%, assuming all other incentives are earned. Column 8 suggests that, assuming that all non-rTSR-based RP targets are met and that all other incentives are earned, meeting rTSR-based targets increases CEOs' plan-based compensation by an average of 40%.[7]

[5] This increase in the explicit use of relative performance grants is consistent with descriptive evidence from the prior literature. For example, our summary statistics are comparable to those of Bettis, Bizjak, Coles and Young (2014), which also uses data from ISS Incentive Lab spanning the time period 1998-2012 (e.g., see their Table 1).

[6] Expected compensation is calculated using values reported in the Grants of Plan-Based Awards Table by adding the dollar values of Estimated Future Payouts Under Non-Equity Incentive Plan Awards based on target performance and the Grant Date Fair Value of Stock and Option Awards reported in the proxy statements.

[7] The incentive ratio in column 7 (8) is calculated as the expected plan based compensation from meeting RP-based targets, assuming that all other incentives are earned as reported in column 3, divided by the counterfactual expected compensation excluding RP-based allocations (rTSR-based allocations). For example, the RP-based incentive ratio of 1.50 in 2014 implies that on average, CEOs who achieve their RP-based targets can earn 50% more than the counterfactual in which they do not earn their RP-based threshold performance payouts. When an incentive grant

Our back-of-the-envelope estimates are consistent with existing and growing evidence of the importance of performance-based—and in particular RP-based—incentives for CEOs. For example, Bettis *et al.* (2014) shows that the RP-related components of compensation at RP-grant-issuing firms between 1998 to 2012 consistently determined more than 30% of the realized total compensation amount. Similarly, De Angelis and Grinstein (2016) shows that, for a hand-collected sample of S&P500 firms in 2007, about one-third of firms explicitly mentioned that their performance-based awards were RP-based, and that firms with RP contracts attributed about half of the estimated total performance award value to RP. The paper also documents that about 75% of the performance metrics associated with RP are market measures; this finding is consistent with the notion that stock-price-based measures prevail for relative performance purposes.

Table 3 summarizes the different types of benchmarks used to measure relative performance. The sample of RP grants is identical to Table 1, Panel B. Specifically, we consider four benchmark categories: a specific peer set, the S&P500 index, the S&P 1500 index, and other indexes (typically industry-based). Columns 4-7 report the percentages of RP grants that use each type of benchmark in a given fiscal year. Column 8 reports the percentage of RP grants whose benchmark cannot be identified. Because each grant can be associated with multiple types of benchmarks, the sum of the values across columns 4-8 can exceed one. Finally, column 9 reports the average number of peer firms used by firms that opt for a specific peer set.

Overall, we observe that around half of all relative-performance grants use specific peers as a benchmark, and that the average number of peers is 15-18. At firms that choose an index benchmark, the most popular choice is the S&P500. In fiscal year 2014, for example, 48% of RP grants to CEOs identify specific peer firms as the relative benchmark; 21% use the S&P500 or 1500 indexes, 19% use another index (e.g., narrower or industry-specific indexes), and 15% do not specify a peer benchmark. The distribution of relative benchmark types remained stable over the eight-year period from 2006 to 2014. Among the firms that chose an index, the distribution of index choices also remained stable; in 2014, for example, 40% chose the S&P500, 12.5% chose the S&P1500, and the remaining 47.5% chose other indexes.

involves multiple performance criteria, we attribute the total expected payout from meeting all targets equally to each metric.

2.2 Assessing Properties of RP Benchmarks

Given the rising importance of rTSR as a metric for judging and incentivizing managerial performance, our paper seeks to assess the properties of boards' performance-measurement choices. We examine the extent to which boards' choices of relative-performance benchmarks follow the informativeness principle (Holmström, 1979; Holmström and Milgrom, 1987): that is, how well do firms' choices of RP benchmarks capture the systematic component of their stock returns?

To examine whether chosen RP benchmarks satisfy the expected properties of the common component in firm performance, we exploit the linear factor structure of firm performance that is implicit in the definition of relative performance metrics like rTSR and explicit in the standard theoretical framework (e.g., $p = 1 \times c + e$ where c is the systematic or common component of performance p and e is the non-systematic or idiosyncratic component), under which the common component of firm performance should exhibit two properties.[8] First, the structure of the relative performance metric ($p - 1 \times c$) implies that the firm's performance should have a slope ("benchmark-return-beta") of 1 with respect to the common component.[9] Second, the common component should capture all of the contemporaneous variation in firm performance that is systematic (i.e., non-idiosyncratic to the firm).

To examine whether firms' chosen RP benchmarks exhibit these two properties, we estimate the following time-series returns regression for each firm:

$$R_{it} = \alpha_i + \beta_i R_{p_{it}} + \epsilon_{it} \tag{1}$$

where R_{it} is firm i's monthly cum-dividend returns in period t and $R_{p_{it}}$ is the returns of that firm's benchmark peers. Our empirical analysis focuses on those firms that tie their CEOs' performance-based incentives to rTSR; the quality of the RP metric should be especially important to them. We therefore restrict our attention to the subsample of firms covered by ISS Incentive Lab that (1) issued rTSR-based grants to their CEOs (that is, the sample described in Table 1, Panel C),

[8]That the factor structure is linear is without loss of generality. A unique linear structure with respect to a set of factors is guaranteed by the projection theorem. Note also that a linear factor structure is consistent with the relative performance metrics observed in practice, like rTSR, which are expressed as the difference between two performance measures. That the factor structure has a single common component is also without loss of generality: with multiple factors, the common component is simply $c = b'f$, where b' is a vector of factor sensitivities and f is a vector of factor returns.

[9]Together with the definition of rTSR, this prediction is also implied in Equation 3 of Janakiraman *et al.* (1992).

(2) disclose the peers or indexes used in determining performance payouts, and that (3) intersect with available alternative benchmark peers introduced by Lee *et al.* (2015). In total, our sample consists of 356 unique firm-benchmark-type (i.e., index vs. specific peers) observations between fiscal years 2006 and 2013; this sample represents 330 unique firms, due to the inclusion of 26 firms that switched benchmark types during the sample period. Returns data are obtained from CRSP monthly files, and firms with fewer than ten months of valid monthly returns in total are excluded from the sample. Detailed construction of our final sample is described in Appendix Table A.II.

In estimating Eqn. (1), we use the median of the peer set's returns for firms that select a set of specific RP peer firms. Although the choice of the order statistic from the peer-return distribution can be arbitrary, median is the most popular performance target in relative-performance contracts (Reda and Tonello, 2015; Bennett *et al.*, 2017).[10] For firms that select an index as the relative benchmark, we use the corresponding index returns. For the RP benchmarks disclosed in the proxy statement for a given fiscal year, we use returns from the following fiscal year. For example, if firm i reports its fiscal-year-end date as December 2000, we obtain monthly stock-return data for the calendar window January 2001 to December 2001 for it and for its performance peers, disclosed in that proxy statement, to calculate $R_{p_{it}}$. Our choice reflects how the selected peers are *used* in RP contracts and thus how they relate to realized firm performance *ex post*. However, the results of our paper do not change if we use the prior-year stock returns, due to a low level of turnover in firms' chosen RP peers.

To examine the first property of the common component of firm performance, we analyze the slopes from estimating Eqn. 1, summarized in Table 4. Interestingly, we find a cross-sectional average slope coefficient of 1.03 across all firms, which is statistically no different from the normative benchmark of 1. Moreover, we find that the average slope is close to, and statistically no different from, 1 for both prevailing approaches to selecting RP benchmarks: (a) based on a customized set of peer firms ("specific peers") and (b) based on an industry or market index. Thus, it appears that firms' chosen benchmarks, whether index or specific peers, exhibit a return-regression slope of 1, consistent with these benchmarks capturing the systematic component of firm performance.

To examine the second property of the common component of firm performance, we analyze

[10] The optimal aggregation rule from an informativeness perspective is discussed in Holmström (1982) and Dikolli, Hofmann and Pfeiffer (2012).

the R^2s from estimating Eqn. 1, summarized in Table 5. If firms' chosen RP benchmarks capture the common component of their stock returns, no other contemporaneous variables should exhibit incremental explanatory power. Equivalently, no other contemporaneous variables or alternative peer formulations should generate R^2 (from estimating Eqn. 1) higher than can be obtained by using the chosen RP benchmark returns.

To test this implication, we choose three alternative RP benchmarks: the S&P500 index, firms' compensation-benchmarking peers and the search-based peer firms (SBPs) of Lee *et al.* (2015). We choose the S&P500 index as a normative benchmark since it is salient and a "free option" to firms. We examine compensation-benchmarking peers because they represent a set of peers already chosen by firms' boards of directors—in this case to set their CEOs' compensation levels—and thus a readily available alternative. Finally, we utilize SBPs, which represent firms' economic benchmarks as collectively perceived by investors and inferred from co-search patterns on the SEC's Electronic Data-Gathering, Analysis, and Retrieval (EDGAR) website, because the findings of Lee *et al.* (2015, 2016) suggest that SBPs prevail over other state-of-the-art methods for identifying economically related firms for purposes of explaining co-movement of stock returns, valuation multiples, growth rates, R&D expenditures, leverage, and profitability ratios.[11]

Table 5, Panel A, compares the R^2s generated by firms' selected benchmarks to those obtained from both the S&P500 index and SBPs.[12] To facilitate comparisons, all regressions are conducted using the same underlying set of base firms. In columns 1, 2, and 4, the first row shows that, across all 356 unique firm-benchmark observations, using the S&P500 as the benchmark yields an average R^2 of 32.8%, which is significantly (at the 1% level) lower than the 48.3% generated by firms' chosen RP benchmarks, consistent with the latter's capture of the common component of performance. However, in columns 1, 3, and 5, the first row shows that SBPs produce R^2s (of 51.8%) that are significantly higher (at the 1% level) than those produced by firms' chosen RP benchmarks, inconsistent with the latter's capture of the common component of performance.

[11] Among S&P500 firms, for example, an equal-weighted portfolio of top-10 SBPs explains 63% more of the variation in base-firm monthly stock returns than a randomly selected set of 10 peers from the same 6-digit Global Industry Classification System industry. A search-traffic-weighted portfolio of top-10 SBPs, weighted by the relative intensity of co-searches between two firms (a measure of perceived similarity), explains 85% more of the variation in base-firm monthly returns.

[12] We weight peer returns by the relative magnitude of EDGAR co-search fractions, interpreted as a measure of similarity or relevance between firms. Lee *et al.* (2015) and Lee *et al.* (2016) show that this weighting scheme performs best at explaining contemporaneous variations in base-firm returns. To avoid look-ahead bias, we follow Lee *et al.* (2015) in always identifying SBPs using search traffic from the prior calendar year.

We also examine how these results vary for the subset of firms that choose specific-peer benchmarks and the subset that choose index-based RP benchmarks. Prior literature has suggested that a narrower set of peer firms is generally more capable of measuring the common factor in performance than are broad indexes (Lewellen and Metrick, 2010). Table 5, Panel A, rows 2 and 3, compares the R^2s generated by firms' selected benchmarks for the subsets of firms that use specific peers (N=201) and an index (N=155) respectively. In both cases, we find that firms' selected RP benchmarks outperform the S&P500, although the index-based benchmarks outperform by a smaller amount—a 21.6% proportional improvement—than do the specific peers, whose average R^2 outperforms those produced by the S&P500 proportionally by 67.1%.[13] However, we find that the overall underperformance of firms' benchmarks relative to SBPs is concentrated among the set of firms that use index-based benchmarks, among which the average time-series R^2 is 40.0%. By comparison, the average R^2s produced by SBPs (46.4%) represent not only a statistically significant improvement but also an economically significant one (a 16% proportional improvement). On the other hand, specific RP peers produce an average R^2 of 54.8%, which is statistically no different from the average R^2 of 56.0% generated from firms' SBPs.

In untabulated tests, we assess the robustness of the above findings in a few ways. First, we re-frame the analysis by comparing the variance in firms' chosen rTSR metrics and the variance in their realized TSR. By comparing solutions to the contracting problem based on relative performance (where we do not perfectly observe the common shock) to that of a contract that rewards based on absolute firm performance, one can deduce that incentivizing based on relative performance is desirable when $\sigma_c^2 > \sigma_b^2$.[14] The intuition is that a relative performance metric ($p - \hat{c} = a + \omega_b + \epsilon$) is preferred when it is a less noisy reflection of managerial effort or talent than the absolute performance metric ($p = a + c + \epsilon$). Conversely, an absolute performance metric should be relied on if peer benchmarking introduces more noise into the inference of managerial effort or talent. Consistent with the R^2 results and with firms' intention to remove systematic noise, we find that, on average, firms with rTSR-based incentives have 40% lower variance (statistically significant at the 1% level) in realized relative performance ($Var(rTSR)$) than their realized absolute-performance counterparts ($Var(TSR)$). This is true both for firms that choose index-based benchmarks (which exhibit 30%

[13] Though index-based benchmarking firms include those that use the S&P500, the outperformance stems from narrower industry indexes that are also part of the group.

[14] Wu (2016) refers it as the "boundary condition" for relative performance evaluation.

lower variance on average, significant at the 1% level) and for firms that choose specific peers as benchmarks (which exhibit 48% lower variance on average, again significant at the 1% level).

Second, we examine alternative rules for aggregating peer performance, namely the mean and the 75^{th} percentile of peer portfolio returns. Since these variations do not affect index returns, these robustness tests focus only on firms that use specific RP peers. As in Panel A, the mean (75^{th} percentile) of chosen peers' returns yields an average time-series R^2 of 54.7% (51.7%) in return regressions across the set of specific-peer-benchmarking firms. The underperformance of 4.4% relative to SBPs is statistically significant at the 1% level for the 75^{th} percentile of specific peer performance, but not significant for the mean of the portfolio of specific peers. Third, we examine how results differ by using an alternative normative peer benchmark—the peers most commonly co-covered by sell-side analysts ("ACPs" of Lee *et al.*, 2016)—and find results very similar to those using SBPs.

Table 5, Panel B, compares the R^2s generated by the selected benchmarks of the 322 firms whose compensation-benchmarking peers are reported in ISS Incentive Labs. Using this sample, and the subsample of firms that choose specific peers and index-based benchmarks, the relative comparisons to SBPs yield virtually identical results. However, the comparison to firms' compensation-benchmarking peers reveals surprising results. In particular, although the R^2s produced by firms' selected RP benchmarks are statistically no different from the R^2s produced by firms' compensation-benchmarking peers, the subsamples of specific peers and index-based benchmarks reveal very different patterns. Whereas specific peers produce R^2s that are on average 4.7% higher than those produced by compensation-benchmarking peers, a difference that is statistically significant at the 1% level, index-based benchmarks produce R^2s that are on average 7.5% lower than those produced by compensation-benchmarking peers, a difference that is statistically significant at the 1% level.

Overall, we draw two conclusions from the empirical findings in Tables 4 and 5. First, the majority of firms—those choosing specific peers—select RP benchmarks that perform remarkably well. Consistent with capturing the common component of firms' returns, these specific-peer benchmarks exhibit a slope of 1 and exhibit R^2s of magnitudes similar to those of SBPs. This finding may be particularly surprising because it suggests that, although there have been significant advances in contract theory over the last four decades that build on the classic and simple model of Holmström (1979), the choice of rTSR measures for the majority of firms remains consistent with the informativeness principle and the predictions of Holmström (1979).

However, nearly half of the firms—those choosing indexes—do not select RP benchmarks in accordance with the informativeness principle. Although these index-based benchmarks also exhibit a slope of 1, they fail to capture a significant proportion of variation in performance that can be explained by the common component of returns. This finding is particularly puzzling in light of the availability of compensation-benchmarking peers that can serve as RP benchmarks: they not only capture well the common component of firms' returns, as shown in Panel B of Table 5, but untabulated tests also show that they exhibit a slope of 1 from estimating Eqn. 1. We explore possible explanations for firms' choice of index-based benchmarks in Section 5.

3 Interpreting R^2 Differences

The results on R^2s presented in the last section suggest that while the majority of firms adhere to the informativeness principle in selecting their RP benchmarks, over 40% of firms' chosen peers fail to capture the common component of firms' returns for relative performance assessments. Although these findings on R^2s are suggestive, they are difficult to interpret economically. In this section, we interpret these R^2 results in the context of a classic principle-agent model. Our intent is simply to provide a baseline framework for assessing the firm-performance ramifications of firms' choice of benchmark precision. In Section 5, we will explore potential alternative explanations—i.e., the sources of economic frictions—that explain firms' choice of index-based benchmarks.

3.1 Basic Setup

Like Margiotta and Miller (2000), Milbourn (2003), Gayle and Miller (2009), and Coles, Lemmon and Meschke (2012), the starting point of our model follows Holmström and Milgrom (1987) and Gibbons and Murphy (1990). We assume a risk-neutral principal (the board) and a risk-averse agent (the CEO), and assume further that firm performance follows a factor structure consisting of (i) unobserved managerial effort $[a]$, (ii) a common shock that is beyond the manager's control $[c \sim_{iid} N(0, \sigma_c^2)]$, and (iii) a firm-specific idiosyncratic shock $[\epsilon \sim_{iid} N(0, \sigma^2)]$

$$p = a + c + \epsilon. \tag{2}$$

The empirical counterpart to p in the RPE literature often includes stock returns or accounting performance measures (Antle and Smith, 1986; Barro and Barro, 1990; Bertrand and Mullainathan, 2001; Ittner and Larcker, 2002; Aggarwal and Samwick, 1999b; Mengistae and Colin Xu, 2004; Albuquerque, 2009, 2013; Lewellen, 2015).

Under linear contracts of the form $w = \alpha + \beta[p - c]$, exponential utility, and a quadratic cost of managerial effort, the manager's problem is given by

$$\max_{a} e^{-\eta\left(w - \frac{\kappa}{2}a^2\right)}, \tag{3}$$

where η is the manager's CARA risk aversion. The manager's optimal effort choice (and expected firm performance) is given by

$$a^* = \frac{\beta}{\kappa}, \tag{4}$$

which is the performance sensitivity of the linear contract scaled by the cost of effort parameter κ.

In this framework, the risk-neutral board's problem is given by

$$\max_{a,\alpha,\beta} \mathbb{E}(p - w), \text{s.t.} \tag{5}$$

$$\mathbb{E}[-e^{-\eta[w - \frac{\kappa}{2}a^2]}] \geq u(\underline{w}), \text{ and} \tag{5-PC}$$

$$a \in \text{argmax } \mathbb{E}[-e^{-\eta[w - \frac{\kappa}{2}a^2]}], \tag{5-IC}$$

and the optimal relative performance contract is given by

$$w^* = \alpha^* + \beta^*(p - c). \tag{6}$$

The first component of the optimal contract (α^*) is the manager's expected compensation when the firm meets its peers' performance, which depends on the manager's exogenously determined outside option. The second component, $\beta^* = \frac{1}{1+\eta\kappa\sigma^2}$, represents the pay-performance sensitivity portion of the contract. The contract form can also be interpreted within the stated objectives of rTSR in Table A.I. Without loss of generality, under Eqn. 2, relative returns $(p - c)$ is equivalent to the excess returns of the firm $(a + e)$, i.e., "shareholder alignment".

Finally, given the optimal contract chosen by the board, the manager's effort can be rewritten as

$$a^* = \mathbb{E}[p] = \frac{1}{\kappa + \eta\kappa^2\sigma^2}. \tag{7}$$

Thus, to motivate optimal effort, the principal designs a contract that rewards managers for effort by perfectly eliminating the common shocks in firm performance since they are beyond the manager's control.The key comparative static for our purposes is the negative effect of the variance in idiosyncratic shocks on performance through managerial effort: $\frac{\partial a^*}{\partial \sigma^2} < 0$ from Eqn. 7. The intuition is that, all else equal, higher σ^2 means that a greater proportion of the firm's performance is unpredictable—or explained by idiosyncratic shocks—even after filtering out the common component of performance. Thus, the increased risk in the compensation, coupled with the participation constraint, causes the optimal contract to be less sensitive to performance and thereby reducing the manager's incentives to exert effort. We also obtain the same comparative static under a standard career concerns framework whereby agents have greater incentive to exert effort when the variance of idiosyncratic shocks is lower (Holmström, 1999). Our results are consistent with either interpretation.

3.2 Imperfect Common Shock Filtration

We now depart from the baseline case by introducing imperfect filtering and assuming that the principal (the board) observes the common shock with error,

$$\hat{c} = c + \omega_b, \tag{8}$$

where $\omega_b \sim_{iid} N(0, \sigma_b^2)$.[15] Here, lower σ_b^2 represents the greater ability of performance peers or benchmarks to eliminate common shocks, and perfect common shock filtering reduces to the special case where $\sigma_b^2 = 0$.[16]

[15]We assume that ω_b has mean zero and is iid which is without loss of generality in the manager's optimization choice of effort. A non-zero mean would enter into the level of the manager's expected compensation. Both Sinkular and Kennedy (2016) and Bennett, Bettis, Gopalan and Milbourn (2017) find that firms beat the target payout approximately half the time, which suggests that $\mathbb{E}[\omega_b]$ is close to zero.

[16]Note that this formulation produces the same analytical prediction as the original Holmström and Milgrom (1987) and Aggarwal and Samwick (1999b) framework where a second signal of firm performance (performance peers/benchmarks) exists with the two signals sharing a correlation ρ. One can think of choosing better peers/benchmarks in the two models as either a decrease in σ_b^2 or an increase in the absolute value of ρ.

Under this framework, again assuming linear contracts, exponential utility, and quadratic cost of effort, the manager's optimal effort and expected equilibrium firm performance is given by

$$a^* = \mathbb{E}(p^*) = \frac{1}{\kappa + \eta\kappa^2(\sigma^2 + \sigma_b^2)}. \tag{9}$$

Notably, poorer measurement of the common shock (higher σ_b^2) reduces the equilibrium effort level and expected firm performance.

The intuition is that measurement errors introduce additional noise into the manager's compensation, and in particular into the performance metric $(p - \hat{c})$ from which the manager's effort level is inferred $(a + \omega_b + \epsilon)$. Thus, as in the baseline case above, the incremental volatility stemming from poorer measurement of the common shock causes compensation for the manager to become riskier which then induces the optimal per-for-performance sensitivity (β) to be lower and thereby forcing the manager to choose a lower level of effort.

The R^2 results in Table 5 can be interpreted in the framework of this model, since there is a one-to-one mapping with the variance in measurement errors. In particular, the time-series return regression $p_t = \delta_1 + \delta_2\hat{c}_t + \epsilon_t$ yields an R^2—the squared correlation coefficient $(\rho^2_{p,\hat{c}})$—that can be expressed as a function of the primitives of the model:

$$\hat{R^2} = \rho^2_{p,\hat{c}} = \frac{\sigma_c^2\sigma_c^2}{(\sigma_c^2 + \sigma^2)(\sigma_c^2 + \sigma_b^2)}. \tag{10}$$

For a given firm—i.e., fixing σ^2 and σ_c^2—lower σ_b^2 corresponds to higher R^2. Therefore, the results of Table 5 imply that SBPs produce lower measurement error variances in the common performance factor relative to firms' chosen performance benchmarks.

In fact, the assessment of peer benchmarking adequacy reported in Table 5 can be recast in terms of measurement error variances—ultimately the economic object of interest in our analysis.[17] Under the model assumptions, the following data moments—the variances in prediction errors from peer benchmarks—can identify the measurement error variances up to a scalar constant:

$$Var(p - \hat{c}_{peer}) \quad = \quad \sigma^2_{b,peer} + \sigma^2. \tag{11}$$

[17] Dikolli *et al.* (2012) also model measurement error in the common performance shock in order to understand how various forms of the error structure can bias the standard tests for (implicit) RPE.

Although we cannot identify the magnitude of the measurement error variances, their differences between one set of peer benchmarks and another can be identified—a refinement over the R^2 measures, which, as shown in Eqn. 10, contain σ_c^2, but other factors as well. Notably, these sample moments allow us to obtain a lower bound estimate on the proportional improvement between two benchmark candidates.[18]

3.3 Estimating Measurement Error Variances and Performance Implications

Table 6, panel A, rows 1-3 present simple method of moments parameter estimates of Eqn. 11 for the S&P500 (row 1), firms' chosen performance benchmarks (row 2), and SBPs (row 3), where p represents the monthly stock returns of the base firm. $\hat{c}_{sp500}$, $\hat{c}_{sbp}$, and $\hat{c}_{firm}$ are monthly stock returns of the S&P500 index, the (traffic-weighted-average) returns of firms' SBPs, and the median returns of firms' chosen performance benchmarks, respectively. Note that in this exercise we are implicitly constraining the coefficient on c in Eqn. 2 to equal 1, consistent with how the RP benchmarks are used in practice.

In column 1, the estimated $\sigma^2_{b,firm} + \sigma^2$ across the whole sample equals 44.489, whereas $\sigma^2_{b,sbp} + \sigma^2$ equals 40.742 for a statistically significant difference (at the 1% level) of 3.747. On a relative basis, firms' chosen performance benchmarks produce *at least* 9.2% greater variance in measurement errors. Columns 2 and 3 of the same panel recover estimates for the subset of firms that selected specific peers and indexes, respectively. Similar to our findings in Table 5, SBPs' out-performance of firms' chosen performance benchmarks is concentrated in the set of firms that choose indexes. Index-based benchmarks' measurement error variances are at least 16.4% greater than SBPs; for firms using specific peers, variances are at least 4.4% greater, both statistically significant at the 1% level. In summary, our findings on the ultimate construct of interest—the performance of firms' chosen RP benchmarks in terms of measurement error variances—are empirically and theoretically consistent with the earlier R^2 results of Table 5.[19]

Given the greater measurement error variance implicit in the benchmarks chosen by firms, we

[18] It is easily shown that $\frac{\sigma^2_{b,firm}+c}{\sigma^2_{b,sbp}+c} > 1 \implies \frac{\sigma^2_{b,firm}}{\sigma^2_{b,sbp}} > \frac{\sigma^2_{b,firm}+c}{\sigma^2_{b,sbp}+c}$.

[19] An interesting question is whether idiosyncratic firm performance isolated through relative performance is driven by discount rate or cash flow news. Both cases can support the effort story, depending on one's assumptions about the nature of the manager's effort. In either case, Vuolteenaho (2002) shows that, at the firm level, the majority of the variation in returns is driven by cash flow news.

proceed to quantify the economic implications benchmark selection in terms of managerial effort and expected firm performance, which can be estimated using the sample analogue of Eqn. 9. In particular, given the manager's risk aversion (η) and effort-cost parameter (κ), the impact of poorer benchmarks on expected performance is given by

$$\mathbb{E}\left[p\left(\sigma_{sbp}^2\right) - p\left(\sigma_{firm}^2\right)\right] = \frac{1}{\kappa + \eta\kappa^2(\sigma_{b,sbp}^2 + \sigma^2)} - \frac{1}{\kappa + \eta\kappa^2(\sigma_{b,firm}^2 + \sigma^2)}. \quad (12)$$

This computation requires identification of the risk-aversion (η) and the effort-cost (κ) parameters; however, with three unknowns (κ, η, $\sigma^2 + \sigma_{b,firm}^2$) and two equations (9 and 11), the model is underidentified. The under-identification of the risk aversion parameter is common in estimating these models (see, e.g., Gayle and Miller, 2015); our calibration, thus, borrows accepted ranges of the risk aversion parameter η from the prior literature. Following Haubrich (1994), we consider the range of η between 0.125 and 1.125. Consistent with the model, we also restrict $\kappa > 0$, since Eqn. 9 has two roots. Table 6, Panel A, row 4 reports method of moments estimates of κ, under the assumption that $\eta = 0.625$, the midpoint of the range considered in Haubrich (1994).

The implications of SBPs' outperformance of firms' chosen RP benchmarks are obtained by applying the method of moments parameter estimates and the assumed risk-aversion parameter to Eqn. 12. Across all firms in the sample, at the midpoint of risk aversion $\eta = 0.625$, we estimate the counterfactual performance gained (lost) under SBPs (S&P500) to be 91.73 (281.83) basis points in annual returns as reported in Table 6, Panel C. In other words, the on-average under-performance of firms' selected benchmarks—in terms of explaining the common component of firm performance—implies an economically significant performance effect. These performance implications are again driven by the set of firms that select index-based benchmarks. Interestingly, we find that firms that selected specific peers on average averted a loss of 385 basis points by not selecting the S&P500.

In un-tabulated robustness results, we also estimate the performance consequences (in annual returns) of counterfactually switching to SBPs using the lower ($\eta = 0.125$) and upper bound ($\eta = 1.125$) of the manager's risk-aversion parameter. We find an effect size of 60 basis points corresponding to the lower bound and 153 basis points corresponding to the upper bound for all firms and a range of 106 to 277 for index-based benchmarking firms. Relative to our sample median annual returns, this represents an economically significant 4.4 to 11.3% proportional decline overall, and 7.8 to 20.4% for the sub-sample of index-based benchmarking firms. Moreover, we examine in un-tabulated tests the robustness of the results in Table 6—which is based on a pooled sample estimation to provide sufficient power to estimate κ—to estimation on a firm-by-firm basis. We find similar but slightly larger magnitudes in our estimate of error variances than in the pooled estimates, suggesting that pooling and averaging across firm-specific estimates yield similar conclusions.

4 An Exploratory Analysis of Drivers of Benchmark Choice

The calibration estimates in the prior section suggest that, in the absence of frictions generated by selecting precise peer benchmarks, the on-average underperformance of firms' selected RP benchmarks—particularly index-based benchmarks—implies performance penalties that are economically large. Our hypothesis is that these economic magnitudes could be rationalized, at least in part, by certain economic frictions. In this section, we explore potential economic explanations that govern firms' choice of relative performance benchmarks.

4.1 Empirically Examining Alternative Explanations

There are a number of reasons why a significant proportion of firms may have chosen less precise relative performance benchmarks. One possibility is that more precise benchmarks are desirable, but certain economic frictions led to the choice of less precise benchmarks. One such set of frictions could be governance-related: for example, a low quality board might be less likely to exert effort to identify a precise set of peers and thus more likely to select a (readily-available) index. Another friction could be a high degree of volatility in firm performance, rendering the effort to select precise RP benchmarks (or more generally the filtration of common shocks) less advantageous.[20]

Alternatively, it may be possible that in some situations it is in fact more desirable to have less precise relative performance benchmarks. Two prominent theories offer such an explanation. Janakiraman *et al.* (1992) and Aggarwal and Samwick (1999a), for example, suggests that precise relative performance benchmarks would lead managers working in oligopolistic industries to sabotage their industry competitors rather than improve their own performance. As a result, it may be optimal to reward CEOs for the common shock in order to encourage firms to soften product market

[20] Appendix A derives these predictions analytically by endogenizing the choice of benchmark precision in the board's problem.

competition. Thus, one prediction of such a theory is that firms with greater market power are less likely to select direct competitors and more likely to adopt broad indexes in order to eliminate market-level volatility from their performance. Another theory, offered by Garvey and Milbourn (2003), is that managers with greater ability to self-insure against the common factor benefit less from more precise benchmarks. If so, it is possible that selection of index-based benchmarks on the part of certain firms may reflect lower benefits from risk-sharing motives.

To explore whether these theories may explain our findings, we investigate the empirical drivers of index-benchmark selection in Table 7. Our main dependent variable of interest is *Index*—an indicator variable equal to 1 if the firm-year involves an index-based benchmark, and 0 otherwise. We examine a number of explanatory variables relating to CEO, board, firm, and industry characteristics: we include three CEO characteristics—*CEO Total Pay*, *CEO Tenure*, and *CEO Age*; four measures of board characteristics—*% Busy Directors*, *Board Size*, *Director Workload*, and *% Age 65+ Directors*; and three firm characteristics—*Log Market Cap*, *Return Volatility*, and *Book-to-Market*.[21] We also include a census-based Herfindahl-Hirschman Index measure of SIC-based industry concentration (*Census-based HHI Index*) as an industry characteristic of interest.[22] The specifics of variable construction are explained in Table A.III.

Column 1, Table 7 reports the marginal effects from a probit regression of an index selection indicator on these characteristics; time-fixed effects are also included. We find that, all else equal, firms with higher *CEO Total Pay*, larger *Board Size*, and greater *Director Workload* are associated with a higher likelihood of index selection. To further examine the relation between CEO compensation and the choice of relative performance benchmarks, we decompose *CEO Total Pay* into *CEO Expected Pay* and *CEO Excess Pay*, following Core *et al.* (2008). In particular, *CEO Expected Pay* is the pay that is predicted by a cross-sectional regression model trained on a set of standard economic determinants of executive compensation; and *CEO Excess Pay* is the difference between *CEO Total Pay* and the estimated *CEO Expected Pay*.

In column 2, we find that the positive association between *CEO Total Pay* and *Index* is driven

[21] Our board characteristics are motivated based on prior literature and conversations with practitioners. For example, Fich and Shivdasani (2006) suggest that "busy" boards or over-tasked board members reflect weak board monitoring quality. Jensen (1993), Yermack (1996), and Cheng (2008) argue that larger board size is suggestive of less effective board monitoring. Finally, in interviews with compensation consultants, it was suggested that index benchmarks are sometimes easier to explain to older directors.

[22] Following Ali, Klasa and Yeung (2008) and Keil (2017), we avoid the selection issue within Compustat by using a census-based HHI index obtained from Jan Keil's website: `https://sites.google.com/site/drjankeil/data`.

by *CEO Excess Pay*. In this specification, we continue to find that larger *Board Size*, and greater *Director Workload* are associated with a higher likelihood of index selection. Columns 3 and 4 repeat the specifications of columns 1 and 2, but adds industry-fixed effects. Again, we observe a positive association between *CEO Total Pay* and the likelihood of index selection (in column 3), and that this relationship is primarily driven by *CEO Excess Pay* (in column 4). In both specifications, we again find that larger *Board Size*, and greater *Director Workload* are associated with a higher likelihood of index selection. Taken together, our empirical results are consistent with the possibility that weak board-level governance explains the choice of less precise RP benchmarks.

On the other hand, we do not find evidence that the choice of indexes could be due to firms having greater performance volatility: in all regression specifications, *Return Volatility* is not statistically significant at the 10% level. To the extent that firms with greater fundamental performance volatility may have greater *Book-to-Market*, we may also expect a positive and significant coefficient on *Book-to-Market*. However, in all of our specifications, the coefficient on *Book-to-Market* is negative in sign and statistically insignificant at the 10% level.

Nor do we find support for the choice of index-based benchmarks being driven by oligopolistic industries: *Census-based HHI Index* is not statistically significant at the 10% level in any of the specifications. To the extent that oligopolistic power could also be captured by firm size, we might also expect a positive and significant coefficient on *Log Market Cap*. However, across all specifications the coefficient is generally negative and statistically insignificant.

Finally, we do not find support for the choice of an index being driven by managers who have greater ability to self-insure. Our tests show that the coefficient on *CEO Age*, a common proxy for the ability to self-insure (Garvey and Milbourn, 2003), is not statistically significant at the 10% level in any of the specifications.

4.2 Exploring the Role of Compensation Consultants

To further explore possible explanations for index benchmark selection, we also examine the role of compensation consultants. Prior literature has documented that compensation consultants play an important role in the design of CEO compensation packages (e.g. Cai *et al.*, 2016). In addition, our interviews with compensation consultants and compensation experts suggest that different compensation consulting firms have different preferences for the choice of index versus specific relative performance peer benchmarks.[23]

Table 8 begins by estimating the main specifications of Table 7, but includes compensation-consultant-fixed effects. We make three observations about columns 1 and 2. First, we find that the coefficients on *CEO Total Pay*, *Board Size*, and *Director Workload* remain statistically significant at the 5% level. Second, the positive association between *CEO Total Pay* and *Index* continues to be driven by *CEO Excess Pay*, which remains significant at the 1% level. Thus, the governance-related frictions captured by these variables do not appear to be confounded by compensation-consultant-fixed effects.

Finally, the inclusion of compensation-consultant-fixed effects increases the R^2 of the regression specification by over 40%. The results of F-tests in columns 1 and 2 show that these fixed effects are jointly significant (at the 1% level), consistent with compensation consultants exhibiting systematic tendencies for recommending indexes or specific peers, even after for controlling for firm-level covariates. These empirical results corroborate the qualitative evidence from our interviews, which point to different compensation consultant "styles" with respect to the type of relative-performance benchmarks.

We further exploit the systematic preferences of compensation consultants to examine whether board-level governance weaknesses could explain firms' choice of indexes. Our hypothesis is that, given the preferences of certain compensation consultants, a board that does not carefully scrutinize the consultant's recommendations will ultimately evaluate executives based on the systematic component of performance to some degree.

Under this hypothesis, we expect to see the association between board-level governance weaknesses and the choice of an index to be concentrated at those firms whose compensation consulting firms have a penchant for index benchmarks. The rationale is that poor-monitoring-quality boards will follow the default recommendation of the hired consultant. Therefore, firms with low-quality boards that hire index-preferring consultants will on average choose indexes, and firms with high-quality boards that hire index-preferring consultants will, after careful scrutiny, on average choose a more relevant set of specific peers. This pattern would result in a negative association between monitoring

[23] We conducted interviews with four compensation consultants and two compensation experts involved in determining CEO compensation packages at their respective corporations. While these interviewees all acknowledged that a primary reason for using rTSR in performance contracts is to remove market or industry-level noise from performance, they differed in their preferences for index versus specific relative performance peer benchmarks. Certain consultants have built capabilities to better identify ideal specific-peer benchmarks; others choose indexes by default.

quality and the choice of an index. On the other hand, we would not expect this negative association to be present among firms that hire consultants with a preference for specific peers, because both high-quality- and low-quality-monitoring firms would choose specific peers, albeit for different reasons. Under this hypothesis, moreover, we should also expect a greater propensity to default to the compensation consultant's preferences among firms that hire consultants with a preference for specific peers than among firms that hire consultants with a preference for indexes.

To examine these predictions empirically, we partition the sample by the index preference of the compensation consulting firms into index-preferring (i.e., those compensation consultants whose fixed effect in column 1 is above the median of all compensation consultants) or specific-peer-preferring (i.e., those compensation consultants whose fixed effect in column 1 is below or equal to the median of all compensation consultants).[24] We then estimate the specification of column 1 using the subsample of firms that hired index-preferring consultants (column 3) and a subset of companies that hired specific-peer-preferring consultants (column 4). In addition, we summarize the mean of the dependent variable for each subsample to examine whether the propensity to default to compensation consultant preferences differs based on the consultant type.

Consistent with compensation consultant preferences and firm-level governance weaknesses explaining the choice of index-based peers, we find that the significance in *CEO Excess Pay* and *Director Workload* are concentrated in the set of firms whose compensation consultants prefer index-based benchmarks (column 3). Furthermore, the coefficients on both variables are not only statistically but also economically insignificant when estimated on the subsample of firms whose compensation consultants prefer specific peers (column 4), consistent with pooling. While *Board Size* and *% Age 65+ Directors* lose significance in both subsamples, the size of the coefficients are much larger in column 3.

Furthermore, we report in the last row of Table 8 that, consistent with our conjecture, the propensity to default to compensation consultant's preferences is nearly three times as high in the set of firms who hire specific-peer-preferring compensation consultants. Whereas more than 56.8% of these firm-year observations select specific peers, only about 20.7% of the firm-year observations matched to index-preferring compensation consultants end up choosing index-based benchmarks.

If some boards scrutinize executive compensation design choices more carefully over time, we

[24] The rank ordering of the compensation consultant fixed effects are unchanged between columns 1 and 2 in Table 8.

may also observe systematic patterns in the changes in relative performance benchmarks. In untabulated results, we find that among the 26 firms that switched benchmark types during our sample period (from index to specific peers or vice-versa), 19 (or 73%) switched from index to specific peer benchmarks. Based on a binomial test, we reject the null of equal likelihood against the alternative that benchmark-type switchers are more likely to have initially chosen an index (p-value of 0.014). Moreover, consistent with our main cross-sectional results, we find that index-based benchmarks performed worse in measurement error variances than SBPs prior to the switch to specific peers, and with nearly equal performance to SBPs after the switch.

Overall, the results of Table 8 are consistent with Cai, Kini and Williams (2016), who finds significant compensation consultant style effects in executive compensation levels and pay mix only within the set of firms with weak governance mechanisms. We build on this evidence by documenting significant compensation-consultant effects in the determination and design of relative performance metrics. Moreover, our evidence is consistent with better-governed firms undoing compensation-consultant preferences for index-based benchmarks.

4.3 Discussion of Alternative Explanations

We discuss the plausibility of several alternative theories for why certain firms may choose index-based benchmarks. These theories are difficult to directly test empirically, but we discuss their plausibility conceptually.

4.3.1 Tournament Incentives

One explanation may be that performance peers are selected in an effort to provide tournament incentives for the firm's manager (Hvide, 2002). Note first that peer selection for such purposes should also, in theory, be consistent with the informativeness principle (Lazear and Rosen, 1981). As such, tournament incentives would not appear to rationalize firms' choice of indexes.

Moreover, the large number of firms included in the index-based benchmarks we observe (e.g. S&P500, S&P1500, Russell 3000) makes it unlikely that the choice of these benchmarks is characterized by some "optimal tournament" involving hundreds or thousands of heterogeneous firms. In particular, Lazear and Rosen (1981) show that absent handicaps (which we do not observe in the performance contracts), heterogeneity among firms, which should be increasing in the peer set size, decreases the effectiveness of tournaments. Furthermore, the tournament mechanism requires all agents to be aware that they are competing in the same tournament, i.e., to be each other's mutually chosen peers. As noted in De Angelis and Grinstein (2016) and Shin (2016), a significant fraction of compensation and performance peers are one-sided (i.e., not mutually designated).

4.3.2 Gameability

An alternative explanation may be that index-based benchmarks may reflect ex-ante rather than ex-post efficiency concerns. For example, in our interviews, one compensation consulting firm offered the possible rationalization that "[w]ith the influence that proxy advisor like the ISS carry, many companies are concerned that a custom group of peers will be misconstrued by proxy advisors as cherrypicking." The presumption is that compensation packages tied to index benchmarks are more difficult to manipulate ex-post. However, we note that most performance contracts in our sample provide payouts that are linearly interpolated from the threshold targets. Thus there is less incentive to manipulate the performance metric due to the absence of discontinuous payoffs.[25] Supporting this view, Bennett *et al.* (2017) finds no evidence of asymmetry in firms' propensity to beat or miss a relative performance target. Moreover, cherrypicking requires that either the CEO or the board can forecast the returns of peer firms or indexes, since benchmarks are formulated ex-ante prior to the realization of performance.[26] But there is no evidence that CEOs or boards can systematically do so: for example, Bizjak *et al.* (2016) finds that boards' selection of performance peers does not affect the level of pay in an economically significant manner.

It is possible that firms are concerned that the selection of specific peers in relative-performance contracts may provide the *appearance* of gaming. For example, if firms with characteristics associated with poor governance are also more sensitive to the external perception of poor governance, they may prefer index benchmarks. Our empirical evidence is not consistent with this alternative as the primary driver of firm behavior, however. Under such a hypothesis, one would expect

[25] One compensation consultant, during our interview, suggested that such linear interpolations are the norm. We verified this by collecting a random sample of 20 contracts from our main sample (across both specific peer and index benchmarking firms) and found two standard contract forms, both of which are effectively linear in rTSR and payout in shares. Three of the 20 contracts are explicitly linear in rTSR. The remaining 17 are linear in the firm's TSR as a percentile of benchmark peers' stock-return distribution. (In untabulated simulations, we find that linearity in the percentiles of peers' TSR distribution also implies linearity in rTSR.) We infer linearity from firms' proxy-statement disclosures of "linear interpolation" in performance between pre-specified target levels.

[26] Morse *et al.* (2011) shows that, prior to 2006, ex-post cherrypicking of performance metrics was prevalent because contracts were not disclosed ex-ante.

governance characteristics to be associated with the choice of index-based benchmarks *regardless* of the compensation consultant's systemic tendencies. In contrast, our findings of a positive association, both statistically and economically, between governance weaknesses and the selection of index-based benchmark is found only in the subset of firms hiring index-preferring compensation consultants.

4.3.3 Aspiration

Another possibility is that the choice of performance peers is aspirational (Francis, Hasan, Mani and Ye, 2016; Hemmer, 2015). Under such a theory, the selection of peer benchmarks would push managers at the firm to generate performance commensurate with or superior to that of well-performing firms.[27]

This theory does not appear to explain the role of performance peers in relative performance contracts. For one, the narrative analysis of Table A.I suggests that the primary objective of rTSR (in the eyes of compensation consultants) is to provide a measure of firm-specific performance or shareholder value improvement that removes the effect of common shocks, rather than as a measure that motivates performance via aspiration. Consistent with this, our empirical findings show that relative performance peers (i) generate return-regression slope of 1 (Table 4) and (ii) significantly reduce the variance in firms' TSR (Section 2.2).

More specifically, the aspiration theory does not explain why some firms may prefer index-based benchmarks in lieu of specific peers. Put differently, it's not clear why firms would choose an index due to aspirational performance reasons.

4.3.4 Implementation Costs

It is also possible that implementation costs of computing rTSR are lower conditional on selecting an index rather than specific peers. One compensation consultant remarked, however, that the typical relative-performance incentive contract requires tracking of the individual performance of firms within an index, not just the overall performance of the index itself. Thus, from an implementation perspective, computing rTSR using index-based benchmarks is, in fact, costlier than using a narrower set of specific peers.

[27] Other models, such as that of Hayes and Schaefer (2009), argue that when there is asymmetric information about the match surplus between the manager and the firm, and when boards care about short-run perceptions of firm value, boards may inflate the CEO's wage as a costly signal of the match surplus (Scharfstein and Stein, 1990).

Overall, our exploratory empirical analyses suggest that compensation-consultant tendencies and governance-related frictions best explain empirical patterns in benchmark inadequacy and in the choice of index-based benchmarks. Boards of directors that do not carefully scrutinize compensation consultants' recommendations may result in evaluations of executives based, to a significant degree, on the systematic component of performance.

5 Reduced-Form Performance Implications in the Cross-Section

We conclude our exploratory empirical analysis by examining the association between realized operating performance and firms' choice of performance peers. Table 9 reports reduced-form regressions of firm performance, in terms of ROA, on the choice of index-based benchmarks, after controlling for *Log Market Cap* and *Book-to-Market*.

Column 1 of the table reports that firms choosing index-based performance peers are associated, on average, with 87 basis points lower ROA, which is statistically significant at the 5% level. Relative to the sample average ROA, this represents a 18% decline.

In columns 2-7, we incrementally add additional control variables to the specification of column 1: time-fixed effects (column 2); industry-fixed effects (column 3); the set of CEO characteristics (column 4), board and firm characteristics (column 5), and industry characteristics (column 6) used in Tables 7 and 8; and compensation-consultant-fixed effects (column 7). Given the lack of a sharp identification strategy, our approach here is motivated by Altonji, Elder and Taber (2005) and Oster (2017), which suggests that regression coefficient and R^2 movements after inclusion of controls provides information about the extent to which the main parameter of interest could be confounded by correlated omitted variables.

As additional control variables are included in Table 9, the estimated coefficient on *Index* remains stable while the adjusted R^2s increase significantly. Although our analysis is exploratory in nature, these patterns alleviate concerns that the negative association between *Index* and future ROA is driven by correlated omitted variables.[28]

As in the prior analyses of Table 8, we test for the joint significance of compensation consultant fixed effects and fail to reject the null of no compensation-consultant effects at the 10% level in column

[28] In untabulated tests, we also find similar results using annual stock returns.

7 of Table 9. This alleviates the potential concern that unobservables correlated with performance drive firms' selection of compensation consultants. Thus, the joint significance of compensation consultant fixed effects in Table 8 probably reflects compensation consultants' systematic tendencies.

Together with the findings of Tables 7 and 8, these findings are consistent with the selection of an index being a symptom of corporate governance weaknesses that is not subsumed by existing measurements used in the literature. These weaknesses are material and associated with lower future performance. An implication of our finding is that the properties of performance metrics chosen by boards to evaluate managers could be a useful way to "monitor the monitor."

6 Conclusion

Market participants have increasingly looked to relative performance metrics such as rTSR to evaluate the performance of firms and managers. Such attention has coincided with a growing trend toward tying executive performance-based compensation contracts to rTSR. Practitioners say that the growing popularity in rTSR is motivated by the desire to filter out the systematic component of performance, consistent with the theoretical motivation from Holmström (1979).

This paper tests the extent to which boards' choices of rTSR measures evaluate managers on the basis of the idiosyncratic component of TSR, following the informativeness principle. We find that over half of firms that tie CEO performance-based contracts to rTSR—those that choose specific peers as benchmarks—do a remarkable job of measuring and filtering for systematic risk in TSR, consistent with the informativeness principle. This finding may be particularly surprising in light of the prevailing view that the executive pay-setting and evaluation process is broken and compromised by powerful executives (Bebchuk *et al.*, 2011; Morse *et al.*, 2011). It is perhaps also surprising that, despite significant advances in contract theory over the last four decades, the choice of rTSR measures for the majority of firms remains consistent with the informativeness principle and the predictions of the simple model in Holmström (1979).

However, we find that nearly half of the firms—those that choose index-based RP benchmarks—use rTSR measures that contain a significant amount of systematic risk and thus do not abide by the informativeness principle. These results provide direct and novel evidence—made possible by the SEC's 2006 compensation disclosure reform—on the longstanding question of whether and to what extent boards evaluate managers on the basis of systematic or non-systematic performance (Antle and Smith, 1986; Janakiraman *et al.*, 1992).

We explore the potential frictions that could rationalize firms' choice of indexes as relative performance benchmarks. Our analysis suggests that this choice can be explained by compensation consultant's systematic tendencies and with board-level monitoring weaknesses. Consistent with the choice of an index being a symptom of corporate governance weaknesses not readily captured by common measurements of corporate governance quality, we find that the choice of an index is associated with lower future ROA.

Our analyses suggest a simple fix for firms that choose index-based benchmarks: their chosen rTSR measures could be significantly improved by using their compensation-benchmarking peers as measures of systematic risk. Indeed, we find that the vast majority of firms that altered their performance incentive plan structure during our sample period switched from index-based to specific peers, resulting in improvements to their measurements of systematic risk.

These findings provide new evidence on the role that compensation consultants play in the managerial evaluation process: we show that compensation consultant tendencies play an important role in the choice of relative performance benchmarks and thus contribute to the evaluation of managers on the basis of systematic risk. Our results also suggest that the properties of performance metrics chosen by boards to evaluate executives can serve as novel indicators of board-level governance quality.

Although our findings are restricted to a sample of firms that tie CEOs' performance-based contracts to rTSR, our analysis may also apply to firms whose executives have implicit rTSR-based incentives. For example, if shareholders, board members, or the executive labor market evaluate managers' competence at least partly on the basis of rTSR, managers' reputational, career, or prestige concerns could be tied to such relative-performance metrics. We believe that distinguishing the role of such implicit incentives from formal incentives for understanding the growing use of relative-performance metrics represents a challenging but promising avenue for future research.

References

Aggarwal, R. K. and Samwick, A. A. (1999a). Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence. *Journal of Finance*, **54** (6), 1999–2043. 4, 6, 21

— and — (1999b). The other side of the trade-off: The impact of risk on executive compensation. *Journal of Political Economy*, **107** (1), 65–105. 16, 17

Albuquerque, A. (2009). Peer firms in relative performance evaluation. *Journal of Accounting and Economics*, **48** (1), 69–89. 1, 5, 16

Albuquerque, A. M. (2013). Do growth-option firms use less relative performance evaluation? *The Accounting Review*, **89** (1), 27–60. 16

Ali, A., Klasa, S. and Yeung, E. (2008). The limitations of industry concentration measures constructed with compustat data: Implications for finance research. *Review of Financial Studies*, **22** (10), 3839–3871. 22

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, **113** (1), 151–184. 29

Antle, R. and Smith, A. (1986). An empirical investigation of the relative performance evaluation of corporate executives. *Journal of Accounting Research*, **24** (1), 1–39. 1, 5, 16, 31

Barro, J. R. and Barro, R. J. (1990). Pay, performance, and turnover of bank CEOs. *Journal of Labor Economics*, **8** (4), 448–481. 16

Bebchuk, L. A., Cremers, K. M. and Peyer, U. C. (2011). The CEO pay slice. *Journal of Financial Economics*, **102** (1), 199–221. 5, 30

Bennett, B., Bettis, J. C., Gopalan, R. and Milbourn, T. (2017). Compensation goals and firm performance. *Journal of Financial Economics*, **124** (2), 307–330. 11, 17, 27

Bertrand, M. and Mullainathan, S. (2001). Are CEOs rewarded for luck? the ones without principals are. *Quarterly Journal of Economics*, **116** (3), 901–932. 16

Bettis, J. C., Bizjak, J. M., Coles, J. L. and Young, B. (2014). The presence, value, and incentive properties of relative performance evaluation in executive compensation contracts. *SSRN Working Paper 2392861*. 8, 9

Bizjak, J. M., Kalpathy, S. L., Li, Z. F. and Young, B. (2016). The role of peer firm selection in explicit relative performance awards. *SSRN Working Paper 2833309*. 6, 27

Brav, A., Jiang, W., Partnoy, F. and Thomas, R. (2008). Hedge fund activism, corporate governance, and firm performance. *Journal of Finance*, **63** (4), 1729–1775. 2

Burney, B. (2016). Relative tsr prevalence and design of s&p 500 companies. *Exequity White Paper*. 37

Cai, C., Kini, O. and Williams, R. (2016). Do compensation consultants have distinct styles? *SSRN Working Paper 2724072*. 23, 26

Carter, M. E., Ittner, C. D. and Zechman, S. L. (2009). Explicit relative performance evaluation in performance-vested equity grants. *Review of Accounting Studies*, **14** (2-3), 269–306. 6

Cheng, S. (2008). Board size and the variability of corporate performance. *Journal of Financial Economics*, **87** (1), 157–176. 22

Coles, J. L., Lemmon, M. L. and Meschke, J. F. (2012). Structural models and endogeneity in corporate finance: The link between managerial ownership and corporate performance. *Journal of Financial Economics*, **103** (1), 149–168. 15

Core, J. E., Guay, W. and Larcker, D. F. (2008). The power of the pen and executive compensation. *Journal of financial economics*, **88** (1), 1–25. 22, 39

De Angelis, D. and Grinstein, Y. (2016). Relative performance evaluation in CEO compensation: A non-agency explanation. *SSRN Working Paper 2432473*. 9, 27

DeMarzo, P. M., Fishman, M. J., He, Z. and Wang, N. (2012). Dynamic agency and the q theory of investment. *Journal of Finance*, **67** (6), 2295–2340. 6

Dikolli, S. S., Hofmann, C. and Pfeiffer, T. (2012). Relative performance evaluation and peer-performance summarization errors. *Review of Accounting Studies*, **18** (1), 34–65. 11, 18

Eisfeldt, A. L. and Kuhnen, C. M. (2013). CEO turnover in a competitive assignment framework. *Journal of Financial Economics*, **109** (2), 351–372. 6

Fich, E. M. and Shivdasani, A. (2006). Are busy boards effective monitors? *Journal of Finance*, **61** (2), 689–724. 22

Francis, B., Hasan, I., Mani, S. and Ye, P. (2016). Relative peer quality and firm performance. *Journal of Financial Economics*, **122** (1), 196 – 219. 5, 28

Garvey, G. and Milbourn, T. (2003). Incentive compensation when executives can hedge the market: Evidence of relative performance evaluation in the cross section. *Journal of Finance*, **58** (4), 1557–1582. 4, 22, 23

Gayle, G.-L. and Miller, R. A. (2009). Has moral hazard become a more important factor in managerial compensation? *American Economic Review*, **99** (5), 1740–1769. 15

— and — (2015). Identifying and testing models of managerial compensation. *Review of Economic Studies*, **82** (3), 1074–1118. 20

Gibbons, R. and Murphy, K. J. (1990). Relative performance evaluation for chief executive officers. *Industrial & Labor Relations Review*, **43** (3), 30S–51S. 6, 15

Godfrey, D. and Bourchier, J. (2015). Discussion of tsr alternatives. *GRG Remuneration Insight*, **77**. 4

Gong, G., Li, L. Y. and Shin, J. Y. (2011). Relative performance evaluation and related peer groups in executive compensation contracts. *Accounting Review*, **86** (3), 1007–1043. 6

Haubrich, J. G. (1994). Risk aversion, performance pay, and the principal-agent problem. *Journal of Political Economy*, **102** (2), 258–276. 20

Hayes, R. M. and Schaefer, S. (2009). CEO pay and the lake wobegon effect. *Journal of Financial Economics*, **94** (2), 280–290. 5, 28

Hemmer, T. (2015). Optimal dynamic relative performance evaluation. *Rice University Working Paper*. 5, 28

Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, **10** (1), 74–91. 1, 6, 10, 14, 15, 30

Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, **13** (2), 324–340. 11

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, **66** (1), 169–182. 1, 17

Holmström, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, **55** (2), 303–328. 10, 15, 17

Hugessen (2016). Assessing relative tsr for your company: A brief overview. *Hugessen Consulting White Paper*. 37

Hvide, H. K. (2002). Tournament rewards and risk taking. *Journal of Labor Economics*, **20** (4), 877–898. 4, 26

Ittner, C. D. and Larcker, D. F. (2002). Determinants of performance measure choices in worker incentive plans. *Journal of Labor Economics*, **20** (S2), S58–S90. 16

Janakiraman, S. N., Lambert, R. A. and Larcker, D. F. (1992). An empirical investigation of the relative performance evaluation hypothesis. *Journal of Accounting Research*, **30** (1), 53–69. 1, 4, 6, 10, 21, 31

Jensen, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *the Journal of Finance*, **48** (3), 831–880. 22

Keil, J. (2017). The trouble with approximating industry concentration from compustat. *Journal of Corporate Finance*, **45**, 467–479. 22, 39

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, **89** (5), 841–864. 4, 26

Lee, C. M., Ma, P. and Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, **116** (2), 410–431. 3, 11, 12, 47, 48

—, — and — (2016). The search for peer firms: When do crowds provide wisdom? *Harvard Business School Working Paper*. 3, 12, 14

Lewellen, S. (2015). Executive compensation and industry peer groups. *London Business School Working Paper*. 16

— and Metrick, A. (2010). Corporate governance and equity prices: Are results robust to industry adjustments. *Yale University Working*. 13

Loehmann, G. (2016). Revisiting relative tsr. *Compensia White Paper*. 37

Margiotta, M. M. and Miller, R. A. (2000). Managerial compensation and the cost of moral hazard. *International Economic Review*, **41** (3), 669–719. 15

Medland, C., McElheran, A., Stancel, A., Conradi, A. and Anderson, J. (2016). Choosing the right performance peer group. *Meridian Compensation Partners White Paper*. 37

Mengistae, T. and Colin Xu, L. (2004). Agency theory and executive compensation: The case of chinese state-owned enterprises. *Journal of Labor Economics*, **22** (3), 615–637. 16

Mercer (2013). Total shareholder return: Making it work. *Mercer Compensation White Paper*. 37

Milbourn, T. T. (2003). CEO reputation and stock-based compensation. *Journal of Financial Economics*, **68** (2), 233–262. 15

Morse, A., Nanda, V. and Seru, A. (2011). Are incentive contracts rigged by powerful CEOs? *Journal of Finance*, **66** (5), 1779–1821. 5, 27, 30

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, pp. 1–18. 29

Pakela, S., Lane, B. and Scheiring, B. (2017). Effectively administering a relative tsr program: Learning and best practices. *Pay Governance LLC White Paper*. 37

Patel, H. and Edwards, K. (2012). A more thoughtful approach to using relative tsr in performance plans. *Towers Watson White Paper*. 37

Radford (2016). Relative tsr is the top metric for performance-based equity compensation. *Aon Hewitt Radford White Paper*. 37

Reda, J. F. and Tonello, M. (2015). The conference board CEO and executive compensation practices 2015 edition key findings. *SSRN Working Paper 2702563*. 11

Scharfstein, D. S. and Stein, J. C. (1990). Herd behavior and investment. *American Economic Review*, **80** (3), 465–479. 5, 28

Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship. *Bell Journal of Economics*, pp. 55–73. 1

Shin, J.-E. (2016). An evaluation of compensation benchmarking peer groups based on mutual peer-designating behaviors. *SSRN Working Paper 2823592*. 27

Sinkular, J. and Kennedy, J. (2016). Analysis of recent annual incentive financial goals. *Pay Governance LLC White Paper*. 17

Sirras, T. and Sullivan, B. (2012). The problem with relative total shareholder returns. *Workspan Magazine*. 37

Sues, J., Ulaj, A. and Graskamp, E. (2016). The 2015 top 250 report: Long-term incentive grant practices for executives. *Frederick W Cook & Co White Paper*. 37

Swinford, D. (2015). The limits of using tsr as an incentive measure. *Pearl Meyer Partners White Paper*. 37

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, **26** (2), 305–321. 36

Tran, J., Wong, S. and Goforth, M. (2016). Compensation consultant market share rankings 2015. *Equilar White Paper*. 37

Vnuk, M., Bisman, S. and Carter, C. (2012). Long-term incentive plans with relative performance measurements. *Compensation Advisory Partners White Paper*. 37

Vuolteenaho, T. (2002). What drives firm-level stock returns? *Journal of Finance*, **57** (1), 233–264. 19

Walker, D. I. (2016). The way we pay now: Understanding and evaluating performance-based executive pay. *Journal of Law, Finance, and Accounting*, **1** (2), 395–449. 4

Wu, M. G. H. (2016). Optimal risk trade-off in relative performance evaluation. *SSRN Working Paper 2288042*. 13

Yermack, D. (1996). Higher market valuation of companies with a small board of directors. *Journal of financial economics*, **40** (2), 185–211. 22

Appendix A. A Simple Model of Benchmark Selection

To analyze why boards may prefer index-based benchmarks, We extend the basic model in Section 3 to endogenize the choice of benchmark precision into the board's problem. Whereas the baseline model assumes that the quality of the benchmarking technology available to the board (σ_b^2) is exogenously determined (i.e., Eqn. 5 and 8), assume instead that improving the benchmark peers' quality (lowering σ_b^2) entails costly effort on the part of the board, and model the cost function as quadratic in peer precision.

The board's optimal selection of a benchmark, characterized by its measurement error variance (σ_b^2), is the solution to the utility maximization problem based on the board's indirect utility function from substituting Eqns. 6 and 7 into Eqn. 5:

$$\sigma_b^{2*} = \arg\max_{\sigma_b^2} f(\sigma_b^2; \theta, \kappa, \sigma^2) = \arg\max_{\sigma_b^2} \frac{1}{2\kappa + 2\kappa^2\eta(\sigma_b^2 + \sigma^2)} - \underline{w} - \frac{1}{2}\theta\left(\frac{1}{\sigma_b^2}\right)^2.$$

Thus, obtaining a precise estimate for the common component of firm performance (low σ_b^2) is more costly with θ, a cost shifter to capture differential cost of effort or monitoring among boards.

Because the objective function exhibits increasing differences in σ_b^2 with respect to each of the state variables (i.e., $\frac{\partial^2 f}{\partial\sigma_b^2\partial\theta} > 0$, $\frac{\partial^2 f}{\partial\sigma_b^2\partial\kappa} > 0$, and $\frac{\partial^2 f}{\partial\sigma_b^2\partial\sigma^2} > 0$), by Topkis' Theorem (Topkis, 1978), the model predicts that the level of benchmark precision is:

1. decreasing in the board's monitoring quality ($\frac{\partial\sigma_b^{2*}}{\partial\theta} > 0$);
2. decreasing with the level of volatility in firm performance $\left(\frac{\partial\sigma_b^{2*}}{\partial\sigma^2} > 0\right)$; and
3. increasing with the CEO's quality or ability $\left(\frac{\partial\sigma_b^{2*}}{\partial\kappa} > 0\right)$.

The intuition for the first prediction is that a board will be more likely to exert effort to search for better benchmarks when board members are higher-quality monitors (e.g., less distracted or less captured). The intuition for the latter two predictions is that boards are more likely to exert effort to produce better benchmarks when the marginal benefits are higher: that is, because their efforts contribute more to firm performance (lower σ^2) or when managers are more likely to exert effort as a result of better filtering (e.g., because they are more talented).

Table A.I.
Compensation Consultants' Descriptions of the Objectives of Relative TSR

This table presents passages from prominent compensation consulting firms' white papers about the objectives and implementation of relative total shareholder returns (rTSR) in executive performance-based incentive plans. The Market Share column draws on Equilar's 2015 Market Share Rankings, which are based on S&P 500 board engagements (Tran *et al.*, 2016). Note that the concept of removing common noise and shareholder alignment are one and the same. Without loss of generality, under a linear one-factor structure, $p = a + c + e$, where p is returns, c is the common component of returns, and e is idiosyncratic noise in returns, relative returns $(p - c)$ is equivalent to the excess returns of the firm $(a + e)$, i.e., the manager's "alpha".

Firm	**Market Share**	**Description**
Frederic Cook	25.8%	TSR, specifically relative TSR, has emerged as the metric of choice under Say-on-Pay. For shareholders, there is an elegance to TSR in that it demonstrates the return relative to alternative investments. It is also the singular definition of corporate performance used by ISS, as well as the sole performance metric required by the SEC for pay and performance disclosure under Dodd-Frank. As such, some companies view relative TSR as a means to satisfy shareholder, ISS and SEC preferences (Sues *et al.*, 2016).
Meridian Compensation	13%	Defining performance in terms of relative ranking against peers is easier - all a committee has to do is express a philosophical preference about what portion of peers a company has to outperform in order to earn a PSU payout at some level. The market does all of the heavy lifting, taking into account the macroeconomic and other market factors affecting the business (Medland *et al.*, 2016).
Pay Governance	12.8%	Relative TSR is a performance metric most often used in LTI performance plans. Its use as a metric has nearly doubled over the past 5 years and is now used by approximately 50% of companies spanning all sizes and industries. ...the appeal of this metric for shareholders and directors alike is its alignment with shareholder value creation and the absence of having to establish long-term performance goals (Pakela *et al.*, 2017).
Pearl Meyer Partners	7.9%	Measuring TSR on a relative basis levels the playing field by removing overall market movements and industry cycles from the evaluation of executive performance (Swinford, 2015).
Semler Brossy	5.1%	The theory behind relative shareholder return as an incentive metric is sound: Executives earn rewards only when shareholders experience above-market returns (Sirras and Sullivan, 2012).
Tower Watson	5.1%	There are numerous theoretical and pragmatic arguments for the use of relative TSR as a long-term measure. Relative TSR is viewed as objective, transparent and easy to communicate to participants. It supports shareholder alignment and incorporates relative measurement (Patel and Edwards, 2012).
Exequity	4.2%	[rTSR] is intended to align executive wealth with the shareholder experience (Burney, 2016).
CA Partners	4%	Companies tend to use industry peer groups or indices when economic factors have a unique impact on the industry. Companies that use a broader market group, such as a general industry index like the S&P 500, believe that companies compete broadly for investor dollars (Vnuk *et al.*, 2012).
Mercer	3.5%	In an ideal world, companies would use a peer group of like organizations that are subject to the same external influences so that share price movements genuinely reflect decisions made by management (Mercer, 2013).
Compensia	2.6%	[rTSR] provides an unambiguous link to shareholder value, with outcomes that are not driven by overachievement against low expectations (Loehmann, 2016).
Hugessen Consulting	NA	[rTSR] counterbalances the general market movement "wind-fall" problems associated with stock options. Satisfies motivation and retention objectives in both up and down markets. [rTSR] may result in a closer measure of management performance (measuring "alpha"), at least theoretically (Hugessen, 2016)
Radford (Aon Hewitt)	NA	Performance-based equity incentive plans with Relative TSR metrics offer companies a wide range of important benefits, including: Reduced forecasting requirements for long-term performance goals, especially in an uncertain macro-economic environment. Clearer links between final executive compensation payouts and shareholder value creation. Reduced use of redundant performance metrics between annual cash incentive plans and long-term equity incentive plans, improving the overall risk profile of executive compensation programs (Radford, 2016).

Table A.II.
Sample Selection

This table reports the selection criterion used to generate the final samples used in Tables 4, 5 and 6.

Main Sample Selection	Firm-year Observations	Firm-year-month Observations	Unique Firms
(1) Firms in ISS Incentive Lab data that include CEO grant data between fiscal year 2004 and 2013	12,216		1,668
(2) Less firms without CEO grants based on an RP component	(8,998)		
	3,218		751
(3) Less firms whose relative benchmark cannot be identified	(685)		
	2,533		645
(4) Less firms that do not use stock price as the relevant RP performance measure	(486)		
	2,047		554
(5) Less firms without CIK-GVKEY matches	(226)		
	1,821		487
(6) Merged with monthly return data from CRSP		21,710	
(7) Less observations with missing SBP data	(635)	(6,654)	(131)
(8) Less observations before calendar year 2006	(50)	(764)	(4)
(9) Less observations that use both, index and specific peers, in a given fiscal year	(85)	(1,107)	(11)
(10) Less observations with fewer than 10 monthly returns in the time-series regressions	(13)	(77)	(11)
Final Sample	1,038	13,108	330

Table A.III.
Descriptive Statistics

Panel A reports summary statistics on the variables used in Tables 7, 8, and 9. Panel B reports the correlation matrix of the same variables. Observations are at the annual (fiscal) firm-benchmark level. Significance levels of the correlation coefficients in Panel B are indicated by *, **, *** for 10%, 5%, and 1% respectively.

Variables are defined as follows: the variable names from the relevant databases are reported in brackets. Using Compustat, we define the following variables on firm characteristics: *ROA* is the ratio of net income to total assets [ni/at]; *Log Market Cap* is the log of the firm's market capitalization ($Millions) as of the fiscal year end [mkvalt]; and *Book-to-Market* is the book value of common equity ($Millions) [ceq] divided by market capitalization ($Millions) [mkvalt]. Census-based HHI Index is the US census-based Herfindahl-Hirschman Index available from Keil (2017). Using Execucomp, we define the following variables on CEO characteristics: *CEO Total Pay* is the CEO's total compensation (in $Thousands) [tdc1]; *CEO Expected Pay* is obtained following Core *et al.* (2008) by regressing the natural logarithm of *CEO Total Pay* on proxies for economic determinants such as firm size, growth opportunities, stock return, accounting return, and industry controls and exponentiating the expected value from the determinant model, *CEO Excess Pay* is obtained by subtracting *CEO Expected Pay* from *CEO Total Pay*; *CEO Tenure* is the current year minus the year in which the CEO joined the firm [becameceo]; and *CEO Age* is the age of the CEO [age]. Using MSCI GMI's databases on companies and directorships, we define the following variables on board characteristics: % *Busy Directors* is the percentage of the firm's directors with more than four board seats at public firms; *Board Size* is the number of directors on the board; *Director Workload* is the number of full board meetings held over the prior fiscal year [BDMTGS] divided by the number of directors and % *Age 65+ Directors* is the fraction of board members who are aged 66 or greater. Using CRSP, we define *Return Volatility* as the standard deviation of monthly cum-dividend returns [ret] of a firm over the fiscal year. Finally, *Index* is a dummy variable that equals 1 if the firm uses an index as its relative performance benchmark in a given fiscal year.

Panel A: Distributional Statistics

	Obs	*Mean*	*Std.Dev.*	*P25*	*Median*	*P75*
ROA	1,175	0.049	0.053	0.024	0.044	0.078
Index	1,175	0.341	0.474	0.000	0.000	1.000
CEO Total Pay	1,175	9.162	6.946	4.684	7.146	11.579
CEO Expected Pay	1,083	7.448	4.697	4.090	6.282	9.015
CEO Excess Pay	1,083	1.746	5.208	−0.609	0.946	2.940
CEO Tenure	1,175	5.415	4.593	2.000	4.000	7.000
CEO Age	1,175	56.445	5.192	53.000	57.000	60.000
% *Busy Directors*	1,175	0.022	0.048	0.000	0.000	0.000
Board Size	1,175	10.597	2.060	9.000	10.000	12.000
Director Workload	1,175	0.807	0.350	0.583	0.727	1.000
% *Age 65+ Directors*	1,175	0.311	0.319	0.214	0.333	0.500
Log Market Cap	1,175	9.038	1.272	8.123	8.886	9.754
Census-based HHI Index	1,175	0.072	0.038	0.051	0.060	0.082
Return Volatility	1,175	0.079	0.047	0.047	0.068	0.098
Book-to-Market	1,175	0.520	0.312	0.302	0.483	0.683

Table A.III.
Continued

Panel B: Correlation Matrix

ROA	1.00														
Index	-0.04	1.00													
CEO Total Pay	0.17^{***}	0.16^{***}	1.00												
CEO Expected Pay	0.21^{***}	0.01	0.66^{***}	1.00											
CEO Excess Pay	0.06^{**}	0.19^{***}	0.74^{***}	-0.02	1.00										
CEO Tenure	0.02	0.07^{**}	0.02	-0.04	0.06^{**}	1.00									
CEO Age	0.07^{**}	0.03	0.05^{*}	0.08^{**}	0.01	0.37^{***}	1.00								
% Busy Directors	0.01	0.05^{*}	0.05^{*}	0.06^{**}	-0.00	-0.03	-0.07^{**}	1.00							
Board Size	-0.01	0.06^{**}	0.22^{***}	0.33^{***}	-0.01	-0.07^{**}	0.05^{*}	0.05^{*}	1.00						
Director Workload	-0.08^{***}	0.09^{***}	0.00	-0.06^{**}	-0.00	-0.06^{**}	-0.09^{***}	0.06^{**}	-0.35^{***}	1.00					
% Age 65+ Directors	0.03	0.09^{***}	0.07^{**}	0.12^{***}	-0.03	0.07^{**}	0.12^{***}	-0.09^{***}	0.08^{***}	-0.09^{***}	1.00				
Log Market Cap	0.31^{***}	0.05	0.62^{***}	0.81^{***}	0.12^{***}	-0.05^{*}	0.06^{**}	0.06^{**}	0.44^{***}	-0.09^{***}	0.12^{***}	1.00			
Census-based HHI Index	0.07^{**}	0.05	0.13^{***}	0.17^{***}	0.03	-0.04	0.07^{**}	0.03	0.16^{***}	-0.07^{**}	-0.09^{***}	0.11^{***}	1.00		
Return Volatility	-0.26^{***}	-0.00	-0.09^{***}	-0.13^{***}	-0.02	0.04	-0.08^{***}	0.02	-0.21^{***}	0.12^{***}	-0.03	-0.34^{***}	-0.01	1.00	
Book-to-Market	-0.46^{***}	-0.07^{**}	-0.13^{***}	-0.13^{***}	-0.07^{**}	-0.03	0.03	-0.06^{**}	0.01	0.09^{***}	0.03	-0.25^{***}	-0.13^{***}	0.23^{***}	1.00

Table A.IV.
Summary Statistics of Compensation Benchmarking Peers

This table reports the number of relative performance (RP) and compensation benchmarking (CB) peers, and the extent to which these two peer sets overlap for each fiscal year using the ISS Incentive Labs data prior to any sample selection restrictions. Columns 1 and 2 report the mean number of chosen peers in the performance benchmarking peer group and the compensation benchmarking peer group, respectively. Column 3 reports the mean number of firms in both peer groups. Columns 4 and 5 report the fraction of overlapping peers in the performance benchmarking peer group and the compensation benchmarking peer group, respectively.

Fiscal Year	# of RP Peers (1)	# of CB Peers (2)	# of Overlap Peers (3)	Proportion RP Peers Also in CB Peers (4)	Proportion of CB Peers Also in RP Peers (5)
2006	15.18	20.55	8.66	0.63	0.47
2007	15.90	25.44	9.18	0.67	0.47
2008	16.81	26.93	8.54	0.67	0.49
2009	16.56	25.55	8.68	0.63	0.46
2010	17.30	26.06	8.66	0.62	0.46
2011	18.21	25.79	8.70	0.62	0.46
2012	17.24	21.09	7.78	0.62	0.43
2013	17.84	21.83	6.83	0.60	0.40
2014	17.37	19.60	6.25	0.55	0.37

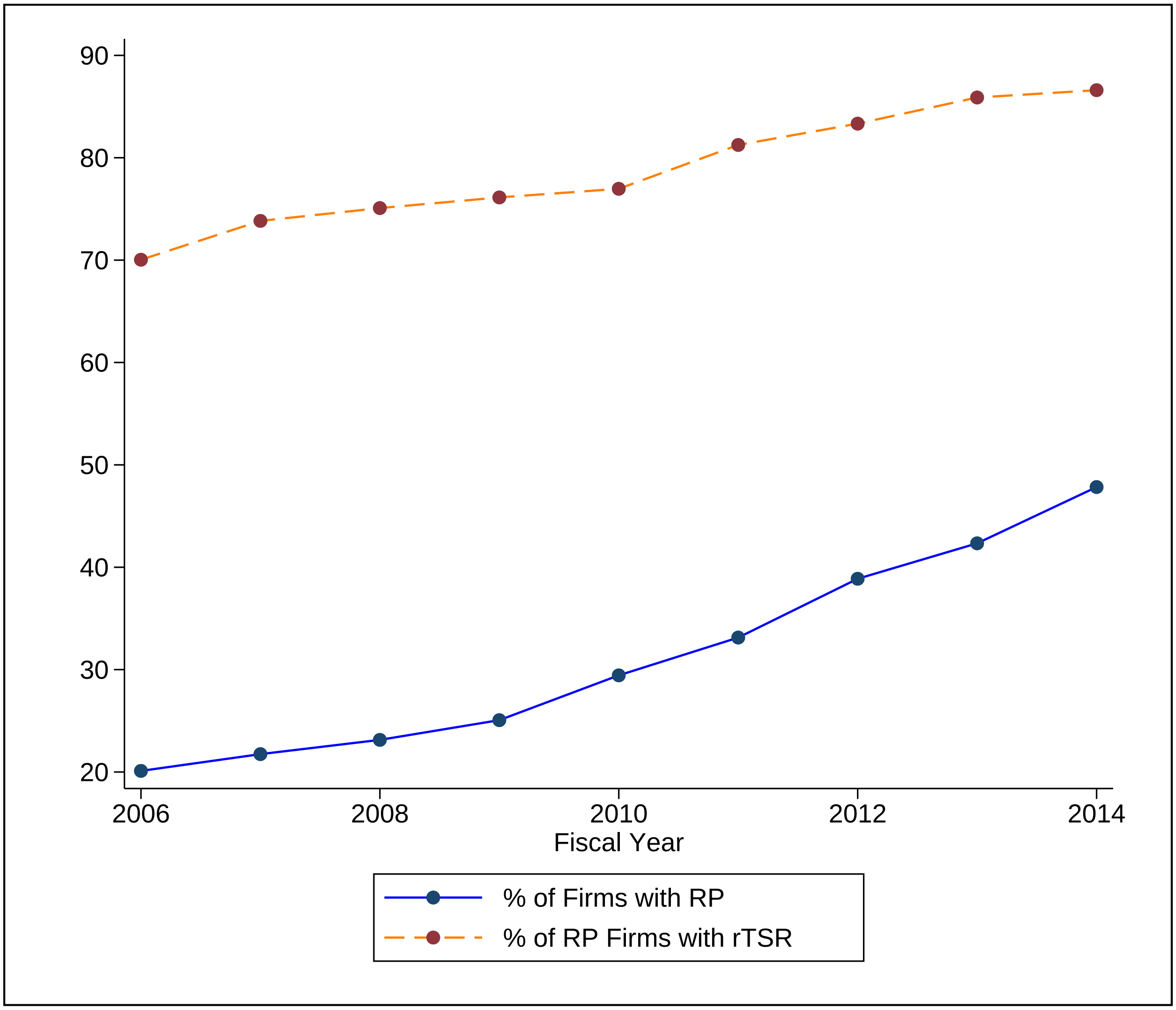

Figure 1. Fraction of Firms Using Relative Performance Contracts 2006-2014 The solid line plots the fraction of firms in the ISS Incentive Labs sample prior to any sample selection restrictions that disclose awarding at least one performance grant based on relative performance (RP) in a given fiscal year; the dotted line plots the fraction of firms with at least one RP-based performance grant that use rTSR as the metric of relative performance.

Table 1.
Summary Statistics on CEO Grants
2006-2014

Panel A reports summary statistics for all compensation grants awarded to the CEO in fiscal years 2006-2014 using the ISS Incentive Labs data prior to any sample selection restrictions. We report the total number of unique firms, the average number of grants awarded to the CEO in each year, the average of the proportion of each award payout type (cash, option, or stock) to the total number of grants awarded to the CEO, and the average of the proportion of each performance evaluation type (absolute performance, relative performance, a mix of the two, and time-based) to the total number of grants awarded to the CEO. Panels B and C report the same summary statistics for sub-samples conditional on CEO grants with a relative performance component and a rTSR component respectively.

			Payout Type [Grant-level]			Evaluation Type [Grant-level]			
Fiscal Year	Unique # of Firms	Unique # of Grants	Cash	Option	Stock	Abs	Abs/Rel	Rel	Time
Panel A: All CEO Grants									
2006	1,278	2.86	0.35	0.29	0.36	0.42	0.04	0.04	0.49
2007	1,283	3.06	0.35	0.26	0.39	0.44	0.05	0.04	0.48
2008	1,249	3.06	0.35	0.25	0.40	0.44	0.05	0.04	0.47
2009	1,153	3.13	0.35	0.24	0.41	0.43	0.05	0.04	0.47
2010	1,165	3.30	0.34	0.21	0.45	0.43	0.06	0.05	0.46
2011	1,159	3.29	0.33	0.20	0.47	0.44	0.07	0.05	0.43
2012	1,173	3.31	0.35	0.18	0.47	0.46	0.09	0.06	0.40
2013	1,155	3.31	0.34	0.17	0.49	0.46	0.10	0.06	0.38
2014	1,108	3.56	0.35	0.15	0.49	0.47	0.11	0.06	0.36
Panel B: CEO Grants with RP Component									
2006	257	1.22	0.35	0.02	0.62	-	0.55	0.45	-
2007	279	1.27	0.36	0.02	0.62	-	0.54	0.46	-
2008	289	1.24	0.29	0.02	0.69	-	0.52	0.48	-
2009	289	1.29	0.32	0.01	0.67	-	0.53	0.47	-
2010	343	1.24	0.28	0.01	0.72	-	0.52	0.48	-
2011	384	1.23	0.23	0.01	0.76	-	0.52	0.48	-
2012	456	1.27	0.21	0.01	0.78	-	0.56	0.44	-
2013	489	1.22	0.19	0.00	0.81	-	0.59	0.41	-
2014	530	1.28	0.17	0.00	0.82	-	0.63	0.37	-
Panel C: CEO Grants with rTSR Component									
2006	180	1.18	0.24	0.02	0.73	-	0.49	0.51	-
2007	206	1.18	0.27	0.01	0.72	-	0.50	0.50	-
2008	217	1.18	0.20	0.01	0.79	-	0.49	0.51	-
2009	220	1.21	0.22	0.01	0.77	-	0.48	0.52	-
2010	264	1.18	0.19	0.00	0.81	-	0.47	0.53	-
2011	312	1.17	0.16	0.00	0.83	-	0.47	0.53	-
2012	380	1.17	0.15	0.01	0.84	-	0.53	0.47	-
2013	420	1.13	0.13	0.00	0.86	-	0.57	0.43	-
2014	459	1.18	0.12	0.00	0.88	-	0.62	0.38	-

Table 2.
Importance of CEO Relative Performance Incentives

This table reports summary statistics on the relative performance incentive ratio and the relative performance stock incentive ratio of compensation grants awarded to CEOs in fiscal years 2006-2014 using the ISS Incentive Labs data prior to any sample selection restrictions. The RP incentive ratio measures the incremental potential incentive when the CEO meets all RP-based targets; it is calculated as (expected incentive-plan-based compensation if all targets are met)/(expected incentive-plan-based compensation if all other targets excluding RP metric-based targets are met). The rTSR incentive ratio measures the incremental potential incentive when the CEO meets all RP-based stock price targets (i.e. rTSR); it is calculated as (expected incentive-plan-based compensation if all targets are met)/(expected incentive-plan-based compensation if all other targets excluding rTSR targets are met). The amount of expected incentive-plan-based compensation is calculated using the values reported in the Grants of Plan-Based Awards Table in the proxy statement which includes both annual and long-term incentive plans. Specifically, it is computed by adding the target dollar value of Estimated Future Payouts Under Non-Equity Incentive Plan Awards and Grant Date Fair Value of Stock and Option Awards (which are based on meeting the performance target). For grants that use multiple performance metrics, we calculate the weighted portion of expected compensation that corresponds to each performance metric. We assume that each performance metric is weighted equally in the calculation of the grant. Column 3 reports the average expected incentive-plan-based compensation. Columns 4 and 5 report the portion of column 3 attributable to RP-based metrics and rTSR metrics, respectively. Column 6 reports the average proportion of RP-based compensation attributable to stock price-based metrics. Columns 7 and 8 report the accompanying incentive ratios.

		Expected Incentive-Plan-Based Compensation				**Incentive Ratios**	
Fiscal Year	Unique # of Firms	Total	RP	rTSR	Fraction	RP	rTSR
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2006	235	6,640,461	1,768,730	1,233,457	0.64	1.63	1.38
2007	265	7,170,858	1,962,250	1,369,169	0.67	1.55	1.34
2008	277	7,266,247	2,050,361	1,563,382	0.70	1.61	1.42
2009	280	6,078,541	1,787,294	1,336,082	0.71	1.65	1.40
2010	332	6,707,443	1,898,931	1,404,596	0.71	1.59	1.39
2011	377	7,160,502	1,820,329	1,389,227	0.75	1.55	1.39
2012	446	7,445,771	2,077,538	1,596,345	0.78	1.58	1.43
2013	478	7,727,030	2,082,448	1,650,666	0.81	1.56	1.41
2014	524	7,950,395	2,005,443	1,505,951	0.81	1.50	1.35

Table 3.
Summary Statistics on Types of Relative Performance Benchmarks
2006-2014

This table summarizes the percentages of rTSR-based grants associated with different types of relative benchmarks for fiscal years 2006-2014 using the ISS Incentive Labs data prior to any sample selection restrictions. Columns 2 and 3 report the unique numbers of firms and grants respectively. Columns 4-8 report the percentages of RP grants that use each type of benchmark: specific peers, the S&P500 index, the S&P1500 index, other indexes (typically industry-based), and unspecified. Because each grant can be associated with multiple types of benchmarks, the values across columns 4-8 can exceed one. Column 9 reports the average number of peer firms chosen as benchmarks for RP grants associated with specific peers.

			Relative Performance Benchmark Type					
Fiscal Year	Unique # of Firms	Unique # of Grants	Specific Peer	S&P500	S&P1500	Other Index	Not Specified	# of Peers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2006	257	313	0.49	0.14	0.05	0.17	0.20	14.79
2007	279	355	0.55	0.16	0.05	0.14	0.14	15.80
2008	289	358	0.57	0.18	0.04	0.17	0.10	16.71
2009	289	373	0.55	0.17	0.05	0.13	0.15	16.43
2010	343	424	0.56	0.14	0.04	0.15	0.17	17.31
2011	384	471	0.55	0.15	0.05	0.15	0.16	18.09
2012	456	579	0.51	0.15	0.04	0.16	0.18	17.07
2013	489	596	0.49	0.18	0.05	0.17	0.15	17.49
2014	530	678	0.48	0.16	0.05	0.19	0.15	16.99

Table 4.
Assessing Firms' Chosen RP Benchmarks: Slope Coefficient

This table estimates and compares the cross-sectional average slope coefficient values (β) from time-series regressions of the form

$$R_{it} = \alpha_i + \beta_i R_{p_{it}} + \epsilon_{it},$$

using CRSP monthly returns data. Column 1 reports the across-firm average slope coefficient from time-series regressions, regressing base firm i's returns on the contemporaneous returns of a portfolio of peers. Column 2 reports the p-value of the null test of $\beta = 1$, and column 3 reports the average number of observations per firm.

Results are reported for the sample of base firms whose chosen benchmarks are identifiable in the data from ISS Incentive Lab. We use return data from 2006-2013 for firms for which there are at least 10 observations. The first row reports on all firms in our sample that satisfy these filters; the second row estimates the same regressions on the subset that select specific peers as benchmarks; the third row estimates the same regressions on the subset that select a stock index as a benchmark.

Standard errors are reported in brackets and significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

Sample	β	p-value $H_0 : \beta = 1$	Mean Obs per Firm
	(1)	(2)	(3)
All	1.0255***	0.3272	36.820
(N=356)	[0.0258]		
Specific Peers	1.0052***	0.8765	40.552
(N=201)	[0.0329]		
Index	1.0520***	0.1864	31.981
(N=155)	[0.0387]		

Table 5.
Assessing Firms' Chosen RP Benchmarks: R^2

This table estimates and compares average R^2 values from time-series regressions of the form

$$R_{it} = \alpha_i + \beta_i R_{p_{it}} + \epsilon_{it}$$

using CRSP monthly returns data. In Panel A, columns 1, 2, and 3 report across-firm average R^2s from time-series regressions, regressing base firm i's returns on the contemporaneous returns of a portfolio of peers. Column 1 uses the median returns of firms' chosen relative performance benchmarks; column 2 uses the returns of the S&P500 index; and column 3 uses the mean returns of search-based peers (Lee *et al.*, 2015). Column 4 reports the differences between R^2 values reported in columns 2 and 1; column 5 reports the differences between columns 3 and 1. Column 6 reports the average number of observations per firm. Panel B re-estimates Panel A with compensation-benchmarking peers (CBP) as an additional alternative normative benchmark.

Results are reported for the sample of base firms whose chosen benchmarks are identifiable in the data from ISS Incentive Lab. We use return data from 2006-2013 for firms for which there are at least 10 observations. The first row reports on all firms in our sample that satisfy these filters; the second row estimates the same regressions on the subset that select specific peers as benchmarks; the third row estimates the same regressions on the subset that select a stock index as a benchmark.

To facilitate comparisons, all the regressions are conducted using the same underlying set of firms. The reported N in parentheses represents the number of firms-benchmark combinations contained in each sample. Standard errors are reported in brackets and significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

Panel A: Baseline Sample

Sample	RP Benchmarks (1)	S&P500 (2)	SBP (3)	(2)-(1) (4)	(3)-(1) (5)	Mean Obs per Firm (6)
All	0.483***	0.328***	0.518***	-0.155***	0.035***	36.820***
(N= 356)	[0.012]	[0.010]	[0.012]	[0.011]	[0.008]	[1.375]
Specific Peers	0.548***	0.328***	0.560***	-0.220***	0.013	40.552***
(N= 201)	[0.015]	[0.012]	[0.015]	[0.014]	[0.009]	[1.898]
Index	0.400***	0.329***	0.464***	-0.071***	0.064***	31.981***
(N= 155)	[0.019]	[0.017]	[0.019]	[0.013]	[0.015]	[1.919]

Panel B: Compensation Benchmarking Peers Subsample

Sample	RP Benchmarks (1)	CBP (2)	SBP (3)	(2)-(1) (4)	(3)-(1) (5)	Mean Obs per Firm (6)
All	0.497***	0.496***	0.525***	-0.001	0.028***	37.547***
(N= 322)	[0.013]	[0.012]	[0.013]	[0.007]	[0.008]	[1.433]
Specific Peers	0.559***	0.534***	0.568***	-0.025***	0.009	40.837***
(N= 190)	[0.015]	[0.015]	[0.015]	[0.008]	[0.009]	[1.921]
Index	0.409***	0.442***	0.464***	0.033***	0.055***	32.811***
(N= 132)	[0.020]	[0.020]	[0.021]	[0.011]	[0.015]	[2.079]

Table 6.
Estimating Measurement Error Variances and Calibrated Performance Implications: Firm Benchmarks vs. S&P500 and SBPs

This table reports method of moments estimates of Eqns. 9 and 11 using pooled firm-month observations. $\sigma^2_{b,sp500}$ is the variance of the measurement error of the common factor using the S&P500 as the relative performance benchmark. $\sigma^2_{b,firm}$ is the variance of the measurement error of the common factor using the firm's chosen performance peers. $\sigma^2_{b,sbp}$ is the variance of the measurement error of the common factor using the firm's search based-peers (Lee *et al.*, 2015). σ^2 is the variance of the firm's idiosyncratic performance where performance is measured via CRSP monthly stock returns, and peer performance is measured at the median of the peer set's returns. κ is the cost of effort parameter in the standard principal-agent model. The estimates are based on the assumption that the manager's CARA risk aversion $\eta = 0.625$.

The four rows in Panel A report the individual parameter estimates of the measurement error variances. Panel B reports the differences in the measurement error variances of the firms' chosen benchmarks relative to two normative benchmarks: the S&P500 and SBPs. Panel C reports in annualized basis points the performance implications of using the firms' chosen benchmarks relative to the two normative benchmarks described in Eqn. 12.

Column 1 reports estimates for the entire sample using the same sampling criterion as Table 5. Column 2 reports estimates for the sub-sample in which the base firm chooses specific firms as the performance benchmark. Column 3 reports estimates for the sub-sample in which the base firm chooses an index (e.g., S&P500, S&P1500) as its performance benchmark. Standard errors are reported in brackets and calculated via the delta method where appropriate. Significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

	All Firms	Specific Peers	Index Peers
	(1)	(2)	(3)
Panel A: Pooled Estimates			
$\sigma^2_{b,sp500} + \sigma^2$	60.197***	63.310***	55.080***
	[1.937]	[2.519]	[3.011]
$\sigma^2_{b,firm} + \sigma^2$	44.489***	41.161***	49.962***
	[1.593]	[1.877]	[2.867]
$\sigma^2_{b,sbp} + \sigma^2$	40.742***	39.418***	42.920***
	[1.367]	[1.712]	[2.269]
κ	0.170***	0.176***	0.162***
	[0.008]	[0.010]	[0.011]
Panel B: Δ Estimates			
$\sigma^2_{b,firm} - \sigma^2_{b,sp500}$	-15.709***	-22.149***	-5.118***
	[0.952]	[1.388]	[1.041]
$\sigma^2_{b,firm} - \sigma^2_{b,sbp}$	3.747***	1.743**	7.042***
	[0.907]	[0.853]	[1.946]
Panel C: Performance Implications			
$\mathbb{E}[p(\sigma^2_{firm}) - p(\sigma^2_{sp500})]$	281.83***	385.73***	97.22***
	[25.03]	[41.29]	[22.12]
$\mathbb{E}[p(\sigma^2_{firm}) - p(\sigma^2_{sbp})]$	-91.73***	-44.41***	-162.62***
	[22.08]	[21.53]	[44.91]
Observations	13,108	8,151	4,957

Table 7.
Explaining Index Benchmark Selection

This table reports the marginal effects, evaluated at the sample mean for continuous variables and at zero for indicator variables, from probit regressions of the firm's choice of an index as its relative performance benchmark on CEO, board of directors, and firm characteristics. Observations are at the annual firm-benchmark level and all variables are defined in Table A.III. All specifications include time fixed effects. Columns 3 and 4 includes industry-fixed effects using the 2-digit Global Industry Classification Standard definitions. Standard errors are clustered at the firm level and reported below the point estimates in brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

	$Pr(Index) = 1$			
	(1)	(2)	(3)	(4)
CEO Characteristics				
CEO Total Pay	0.016***		0.012***	
	[0.005]		[0.004]	
CEO Expected Pay		-0.006		-0.003
		[0.011]		[0.013]
CEO Excess Pay		0.020***		0.014***
		[0.005]		[0.005]
CEO Tenure	0.007	0.006	0.007	0.007
	[0.005]	[0.005]	[0.005]	[0.006]
CEO Age	-0.000	0.000	0.005	0.004
	[0.005]	[0.005]	[0.006]	[0.006]
Board and Firm Characteristics				
% Busy Directors	0.402	0.340	0.548	0.420
	[0.482]	[0.486]	[0.467]	[0.481]
Board Size	0.033**	0.034**	0.027*	0.029**
	[0.015]	[0.016]	[0.014]	[0.015]
Director Workload	0.212***	0.218***	0.203***	0.207***
	[0.069]	[0.072]	[0.073]	[0.074]
% Age 65+ Directors	0.182	0.213	0.258*	0.271*
	[0.149]	[0.154]	[0.149]	[0.154]
Log Market Cap	-0.072**	-0.004	-0.027	0.019
	[0.031]	[0.045]	[0.031]	[0.046]
Return Volatility	-0.197	0.110	0.232	0.343
	[0.552]	[0.562]	[0.559]	[0.579]
Book-to-Market	-0.141	-0.127	-0.040	-0.028
	[0.088]	[0.092]	[0.092]	[0.097]
Industry Characteristics				
Census-based HHI Index	0.309	0.634	-0.168	0.104
	[0.611]	[0.715]	[0.611]	[0.693]
Time FE	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Observations	1,175	1,081	1,175	1,081
Pseudo R^2	0.0640	0.0746	0.1871	0.1955

Table 8.
Explaining Index Benchmark Selection: Compensation Consultant Effects

This table reports the marginal effects, evaluated at the sample mean for continuous variables and at zero for indicator variables, from probit regressions of the firm's choice of an index as its relative performance benchmark on CEO, board of directors, and firm characteristics. Observations are at the annual firm-benchmark level and all variables are defined in Table A.III. All specifications include time and industry-fixed effects using the 2-digit Global Industry Classification Standard definitions. Columns 1 and 2 include compensation consultant fixed effects. Columns 3 and 4 report sub-sample results conditioned on whether the compensation consulting firm has a preference for an index benchmark versus specific peers. Index preference is determined based on whether an individual compensation consulting firm's fixed effects is above the median of all compensation consultant fixed effects in column 1. The sum of the observations from columns 3 and 4 do not equal the number of observations in column 2 because 11 observations are perfectly predicted and dropped in column 2. The reported p-value refers to joint F-tests of the significance of the compensation consultant fixed effects. Standard errors are clustered at the firm level and reported below the point estimates in brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

	$Pr(Index) = 1$			
	(1)	(2)	(3)	(4)
	All Sample	All Sample	Subsample: Index Preferring Comp Consultants	Subsample: Specific Preferring Comp Consultants
CEO Characteristics				
CEO Total Pay	0.012***			
	[0.004]			
CEO Expected Pay		-0.005	0.001	-0.005
		[0.012]	[0.016]	[0.006]
CEO Excess Pay		0.014***	0.015**	0.003
		[0.005]	[0.006]	[0.002]
CEO Tenure	0.006	0.006	0.010	0.000
	[0.005]	[0.005]	[0.008]	[0.001]
CEO Age	0.006	0.005	0.009	-0.000
	[0.006]	[0.006]	[0.008]	[0.001]
Board and Firm Characteristics				
% Busy Directors	0.548	0.416	0.284	-0.007
	[0.465]	[0.482]	[0.765]	[0.097]
Board Size	0.031**	0.033**	0.014	0.009
	[0.013]	[0.013]	[0.023]	[0.009]
Director Workload	0.221***	0.237***	0.216**	0.045
	[0.074]	[0.076]	[0.104]	[0.041]
% Age 65+ Directors	0.260*	0.263*	0.233	0.072
	[0.151]	[0.157]	[0.204]	[0.072]
Log Market Cap	-0.037	0.019	0.012	0.014
	[0.031]	[0.044]	[0.062]	[0.021]
Return Volatility	0.401	0.526	1.192	0.057
	[0.530]	[0.552]	[0.910]	[0.132]
Book-to-Market	-0.006	0.011	0.006	0.009
	[0.087]	[0.092]	[0.142]	[0.024]
Industry Characteristics				
Census-based HHI Index	-0.245	0.069	-0.267	0.058
	[0.555]	[0.593]	[1.126]	[0.139]
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Comp Consultant FE	Yes	Yes	No	No
p-value of F-test (Comp FE)	0.0000	0.0000		
Observations	1,162	1,070	639	442
Pseudo R^2	0.2634	0.2709	0.2173	0.3673
Mean of Dep. Var.			20.7%	43.2%

Table 9.
Index Benchmarks and Future ROA

This table reports OLS regressions of firms' ROA on an indicator of having chosen an index as the relative performance benchmark along with CEO, board of directors, firm, and industry characteristics controls as defined in Table A.III. Observations are at the annual firm-benchmark level. Column 2 incrementally includes time fixed effects. Column 3 incrementally includes industry-fixed effects using the 2-digit Global Industry Classification Standard definitions. Column 4 incrementally includes CEO characteristics (CEO total pay, CEO tenure, and CEO age). Column 5 incrementally includes board and firm characteristics (% busy directors, board size, director workload, % age 65+ directors, and return volatility). Column 6 incrementally includes industry characteristics (census-based HHI index). Column 7 incrementally includes compensation consultant fixed effects. The reported p-value refers to joint F-tests of the significance of the compensation consultant fixed effects. Standard errors are clustered at the firm level and reported below the point estimates in brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

	ROA						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Index	-0.0087**	-0.0085**	-0.0096**	-0.0096**	-0.0080**	-0.0080**	-0.0074*
	[0.0041]	[0.0041]	[0.0042]	[0.0041]	[0.0039]	[0.0038]	[0.0038]
Log Market Cap	0.0087***	0.0087***	0.0091***	0.0106***	0.0123***	0.0123***	0.0125***
	[0.0021]	[0.0021]	[0.0021]	[0.0027]	[0.0027]	[0.0027]	[0.0028]
Book-to-Market	-0.0697***	-0.0692***	-0.0612***	-0.0608***	-0.0573***	-0.0574***	-0.0587***
	[0.0083]	[0.0084]	[0.0088]	[0.0086]	[0.0089]	[0.0088]	[0.0087]
CEO Characteristics	No	No	No	Yes	Yes	Yes	Yes
Board and Firm Characteristics	No	No	No	No	Yes	Yes	Yes
Industry Characteristics	No	No	No	No	No	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes	Yes	Yes	Yes
Comp Consultant FE	No	No	No	No	No	No	Yes
p-value of F-test (Comp FE)							0.3091
Observations	1,175	1,175	1,175	1,175	1,175	1,175	1,175
Adj R^2	0.2516	0.2614	0.2995	0.3086	0.3234	0.3231	0.3261