

rTSR: When Do Relative Performance Metrics Capture Relative Performance?

Paul Ma
Jee-Eun Shin
Charles C.Y. Wang

Working Paper 19-112



rTSR: When Do Relative Performance Metrics Capture Relative Performance?

Paul Ma

University of Minnesota

Jee-Eun Shin

University of Toronto

Charles C.Y. Wang

Harvard Business School

Working Paper 19-112

Copyright © 2019 by Paul Ma, Jee-Eun Shin, and Charles C.Y. Wang

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

rTSR: When Do Relative Performance Metrics Capture Relative Performance?*

Paul Ma
*University of Minnesota
Carlson School of Management*

Jee-Eun Shin
University of Toronto

Charles C.Y. Wang
Harvard Business School

May 2019

Abstract

Relative total shareholder returns (rTSR) has become the predominant metric to isolate managers' idiosyncratic performance. Among firms that explicitly use rTSR in relative performance contracts, 60%—those that choose specific peers as benchmarks—select rTSR metrics that do a remarkable job of filtering out the systematic component of returns in adherence to the informativeness principle. However, firms that choose index-based benchmarks retain substantial systematic noise in their rTSR metrics. We document that the selection of noisy benchmarks is associated with compensation consultants' preferences, which are uncorrelated with observable firm attributes. Firms with weak governance are more likely to choose indexes, not because of opportunism, but because they do not adequately scrutinize outside experts' advice. Collectively, our findings provide a new explanation for why some executives are evaluated based on systematic noise, and novel evidence on how compensation consultants can impact firms.

JEL: G30, J33, M12, M52

Keywords: Relative TSR; relative performance evaluation; systematic risk; board of directors; compensation consultants; style effects

*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 Ana Albuquerque, Brian Cadman, Mark Lang, David Larcker, Chandra Kanodia, Michael Minnis, Pervin Shroff, and conference and seminar participants at Harvard Business School and the Minnesota Empirical Accounting Conference. A related working paper, which inspired this work, was presented at FARS, Georgia State University, 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 thank Srikant Datar of Harvard Business School, Paula Price of Macy's, Yulan Shen of Biogen, Barry Sullivan of Semler Brossy, Terry Adamson, Jon Burg, and Daniel Kapinos of Aon Hewitt, and Nikhil Lele and Trent Tishkowski 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

Over the last decade, relative total shareholder returns (rTSR)—that is, the firm’s own TSR relative to an index or group of peer firms—has become perhaps the single most widely used performance metric by which companies and their executives are judged by market participants. For example, since 2006 the SEC has required all firms to disclose rTSR in their annual reports to shareholders. 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 a firm’s rTSR relative to executive compensation to formulate its say-on-pay recommendations. Activist investors often focus on poor rTSR as evidence of poor management quality or poor performance (e.g., [Brav, Jiang, Partnoy and Thomas, 2008](#)). Finally, the growing preference for rTSR as a performance metric is also evident in the trend towards linking rTSR to performance-based executive contracts: the use of rTSR in these contracts has increased from 70% in 2006 to 87% in 2014 (see [Figure 1](#)).¹

The increasing popularity of rTSR appears to reflect an attempt to filter out the market- or industry-level noise from the evaluation of managerial and firm performance, consistent with the informativeness principle in [Holmström \(1979\)](#).² For example, in a comment letter about the evaluation of managerial performance to the SEC, the Investor Responsibility Research Center Institute, a not-for-profit organization sponsoring and disseminating unbiased research on corporate governance issues, stated that: “TSR is heavily influenced by market and industry factors outside of control of management. It is not a sufficiently robust metric to measure overall longer-term enterprise health and sustained performance. Relative TSR provides a better measure of management performance and strategy success...” ([Leeflang et al., 2014](#)). Compensation consultants, who help boards choose performance evaluation metrics and design executive compensation contracts, cite

¹According to a 2017 Equilar report, “relative total shareholder return continues to be the most popular measurement tying CEO pay to performance in the S&P500.” <https://corpgov.law.harvard.edu/2019/04/11/executive-long-term-incentive-plans/#more-116884>

²The informativeness principle in [Holmström \(1979\)](#) states that any observable metric of performance informative of the agent’s effort choice should be incorporated into the compensation contract in order to improve the effort-insurance trade-off. A particular application of this principle is the use of relative performance evaluation (RPE), whereby the agent’s performance is evaluated relative to a benchmark designed to filter out systematic performance uninformative of the agent’s effort ([Holmström, 1982](#); [Holmström and Milgrom, 1987](#))

similar rationale for the use of rTSR. For example, Pearl Meyer & Partners noted: “[m]easuring 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). rTSR, therefore, counterbalances windfalls that can result from general market movements, such as in the case of stock options. As a consequence, as noted by Hugessen Consulting, such a metric “satisfies motivation and retention objectives in both up and down markets” and “may result in a closer measure of management performance” (Hugessen, 2016).

Given the growing popularity of rTSR and apparent desire to remove systematic risk from the assessment of managerial performance, this paper empirically examines the properties of the rTSR metrics specifically chosen by boards to evaluate and compensate CEOs, as well as the determinants and consequences of these choices. In order to compensate managers on the basis of idiosyncratic performance, the relative performance metrics tied to compensation should, in principle, be free of systematic noise. Thus, our study also sheds new light on the longstanding unsettled academic debate beginning with Antle and Smith (1986) regarding the degree to which managers are evaluated on the systematic component of performance.

Our analyses focus on the sample of firms that explicitly tie executive compensation to rTSR, for whom the quality or the informativeness of rTSR are expected to be of greater importance (Gibbons and Murphy, 1990; Gong, Li and Shin, 2011). Our first research question examines to what extent boards’ chosen rTSR metrics subscribe to the informativeness principle: that is, to what extent are managers being evaluated by the systematic component of their firms’ total shareholder returns (TSRs). Since, by definition, rTSR is the firm’s TSR minus a peer group’s TSR, a board influences the property of the chosen rTSR metric through its choice of peers. Thus, this first research question is tantamount to asking to what extent firms’ chosen rTSR peers capture the systematic component of firms’ TSRs. Our second research question explores the reasons for and consequences of firms choosing lower-quality rTSR peers, i.e., those that perform poorly in capturing and filtering out the systematic component of performance.

We begin by testing two properties that would hold true if firms’ chosen rTSR peers perfectly capture systematic risk. First, the structure of the relative performance metric ($TSR_i - 1 \times TSR_{common}$) implies that a firm’s TSR should have a beta of 1 with respect to its systematic component. Thus, if a firm’s chosen peers perfectly capture the systematic component of its TSR, then its TSR should

exhibit a beta of 1 with respect to its peers' TSR ("benchmark-return-beta"). Second, by definition, the systematic component of a firm's TSR should capture all of the contemporaneous variations in firm performance that is non-idiosyncratic to the firm. Consequently, if a firm's chosen peers perfectly capture the systematic component of its TSR, then no alternative peer groupings should yield greater explanatory power of its TSR.

Our analyses produce three main sets of findings. Our first finding is that 60% of firms—who choose a customized set of peer firms ("specific peers")—select rTSR peers that perform well in capturing the systematic components of firms' TSRs, consistent with the informativeness principle. These firms select rTSR peers that exhibit benchmark-return-betas of 1, and explain the firms' contemporaneous TSR as well as (or better than) the best available alternative peer sets suggested by the literature: search-based peers (SBPs) and firms' self-chosen compensation-benchmarking peers.³

On the other hand, 40% of firms—who choose a broad index-based benchmark—do not fare well in capturing the systematic component of firm performance. Although boards appear to be selecting indexes in order to help filter out systematic risk from TSR (since the chosen indexes also yield benchmark-return-betas of 1), on average these indexes perform significantly worse relative to normative alternatives (Kleinberg, Liang and Mullainathan, 2017). For example, SBPs explain 16% more of the time-series variation in firms' monthly stock returns (i.e., time-series regression R^2). This finding is perhaps surprising considering the readily availability of firms' chosen compensation-benchmarking peers which explain 8% more of the TSR variation than the index-based benchmarks.

Our second set of findings shed light on why firms may choose index-based benchmarks. Evaluating a host of explanations offered by the optimal contracting and governance literature, we find that the observed selection of index-based benchmarks is systematically associated with proxies for governance weaknesses, such as abnormal compensation, larger board size, and heavier director workload. However, we also find that compensation consultants systematically exhibit strong preferences towards either index- or specific-peers- benchmarks, which are orthogonal to all firm characteristics—including governance. While firms hire index- and specific-peer-preferring consul-

³Lee, 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.

tants in roughly equal proportions, we do not find evidence of firms screening consultants on the basis of these preferences. Instead, our evidence suggests that firms are much more likely to overturn the consultant's default preferences when the consultant prefers index-based benchmarks. This is due to stronger governance companies being more likely to overrule consultants' preferences for index-based benchmarks, while poor governance companies are more likely to follow the consultant's default tendencies. Collectively, we interpret the evidence as suggesting that firms' choice of index benchmarks for managerial evaluation could arise in part due to weaker boards' failure to carefully scrutinize the default tendencies of compensation consultants.

We do not find evidence for other explanations offered by the literature that may explain why certain firms would select less precise rTSR benchmarks: (1) managers' efforts being correlated with 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) either boards' attempt at opportunism or alternatively avoidance of the appearance of opportunism (Murphy, 2002); (4) boards' objective in provisioning tournament incentives (Lazear and Rosen, 1981; Hvide, 2002); or (5) boards' provisioning of incentives on the basis of aspiration (Hemmer, 2015; Francis, Hasan, Mani and Ye, 2016). At a basic level, none of the alternative explanations predict that firms' chosen rTSR benchmarks would exhibit a benchmark-return-beta of 1 or the specific patterns we document with respect to the propensity to accept or reject compensation consultant's preferences.

In the final analysis, we explore the potential consequences of firms' rTSR benchmark choice for firm performance. We find that the choice of an index-based benchmark is cross-sectionally associated with lower ROA. However, this association could also reflect unobservable differences across firms beyond their benchmark choice. To address these concerns, we complement the baseline results through i) instrumental variables approaches (two-stage least squares, generalized methods of moments, and intent-to-treat), and ii) the novel methodology recently proposed by Oster (2017) to bound the effects of omitted variables on the bias of OLS estimates. Triangulating our results in these ways, we find a range of estimates under plausible assumptions consistent with the selection of index-based benchmarks leading to lower ROA. These performance consequences could be consistent with managers' response to weaker explicit (Mehran, 1995; Core and Larcker, 2002; Brick *et al.*, 2012) or implicit incentives (career concerns) (e.g., Avery, Chevalier and Schaefer, 1998; Chevalier

and Ellison, 1999; Brickley, Linck and Coles, 1999; Fee and Hadlock, 2003; Milbourn, 2003).⁴

These findings contribute to the broad performance evaluation literature that examines 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; Lambert and Larcker, 1987; Aggarwal and Samwick, 1999a; Bertrand and Mullainathan, 2001; Albuquerque, 2009; Jenter and Kanaan, 2015; Lewellen, 2015). This literature has traditionally focused on managerial compensation, and, although extensive, has also been inconclusive. The mixed findings in part stem from the “sparseness of firms’ public disclosure of the details of explicit executive compensation plans” (Antle and Smith, 1986), which has forced researchers to make assumptions about the form of the compensation contract as well as the properties of the performance metric used in the determination of compensation (e.g., the actual performance metric used to evaluate managers as well as the potential set of peers by which performance may be benchmarked).⁵ Our work approaches these fundamental questions of interest by honing in on the properties of the rTSR metrics tied to compensation, which is made possible by the SEC disclosure reforms in 2006. In order to compensate managers on the basis of idiosyncratic performance, the relative performance metrics tied to compensation should, in principle, be free of systematic noise. Thus, our work suggests that a potentially important reason why some managers may be compensated for systematic noise is that the explicit *relative* performance metrics tied to compensation retain a significant amount of systematic noise.

Our work also contributes to the literature that examines the consequences of the quality of board monitoring (e.g., Core, Holthausen and Larcker, 1999; Fich and Shivdasani, 2006). While prior literature has shown that better board monitoring can reduce the degree to which managers are evaluated and rewarded on the basis of systematic risk (Bertrand and Mullainathan, 2001; Garvey and Milbourn, 2006), our findings add to the understanding of the mechanism through which this

⁴Ghosh and John (2000) provides experimental evidence that sales agents increase their effort in response to less noisy performance measurements. As a recent example that highlights the importance of career concerns, in its 2017 proxy fight against Proctor and Gamble management, Trian argued that the company’s directors have performed poorly by referencing the company’s low TSR relative to the company’s peers over each director’s tenure. <https://trianpartners.com/content/uploads/2017/01/Trian-PG-White-Paper-9.6.17-1.pdf>

⁵Gibbons and Murphy (1990) notes that “theory predicts that compensation will depend on relative performance but gives little guidance as to the form of the contract.” Lambert and Larcker (1987) derives the particular functional form assumptions required for compensation to be linear in their own performance and peer performance and highlights the limits to inference from linear compensation regressions. Finally, Dikolli, Hofmann and Pfeiffer (2012) notes the additional concern that the method of peer aggregation in measuring the systematic component of performance can also introduce measurement errors, impacting inferences regarding firms’ practices of relative performance evaluation.

occurs. Our evidence suggests that naïve deference to compensation consultants' fixed preferences for benchmark types can explain why certain boards, through its choice of index benchmarks, evaluate managers based on systematic risk in TSR. Boards can provide better monitoring, therefore, through more careful scrutiny of outside experts' advice.

Our work also contributes to the literature that seeks to understand the role of compensation consultants in executive compensation (e.g., [Conyon, Peck and Sadler, 2009](#); [Cadman, Carter and Hillegeist, 2010](#); [Armstrong, Ittner and Larcker, 2012](#); [Murphy and Sandino, 2010](#)). In the context of peer selection for relative performance, we show that compensation consultants' preferences (which are independent of firm characteristics) play a significant role ([Bertrand and Schoar, 2003](#)). Moreover, our analyses provide additional evidence that speaks to the unresolved debate about *how* compensation consultants influence compensation. Prior literature offers two views on the role of compensation consultants: one view is that compensation consultants have distinct styles that are sought out by particular firms; another view is that compensation consultants are substitutes who respond to their own economic incentives and the economic circumstances or the incentives of the firm ([Cai, Kini and Williams, 2016](#)). Our findings offer another possibility: compensation consultants have certain styles that can manifest among firms who accept the default recommendations without much scrutiny.

Finally, our evidence contributes to the unsettled debate in the literature about whether peer selection in executive compensation is explained by opportunism or efficiency (e.g., [Bizjak, Lemmon and Naveen, 2008](#); [Faulkender and Yang, 2013](#); [Cadman and Carter, 2013](#); [Albuquerque, De Franco and Verdi, 2013](#); [Gong, Li and Shin, 2011](#); [Cremers and Grinstein, 2013](#); [Bizjak, Kalpathy, Li and Young, 2016](#)). Two recent studies suggest that the selection of rTSR performance benchmarks may be opportunistically motivated. Focusing on firms that select specific peers only, [Gong, Li and Shin \(2011\)](#) finds that while firms' chosen relative-performance peers have higher returns co-movement than random peers, consistent with the informativeness principle, they also exhibit weaker returns than the base firm, suggestive of rent extraction. In a contemporaneous study, [Bizjak, Kalpathy, Li and Young \(2016\)](#) suggests, through simulations, that the ex-ante award value of relative-performance-based compensation is greater for index-based benchmarks relative to credible alternatives.⁶ Our study shows that, in constructing rTSR measures for executive performance

⁶In particular, [Bizjak et al. \(2016\)](#) simulates stock returns and shows that for firms using index-based benchmarks,

contracts, the majority of firms do a remarkable job of filtering out the systematic component of TSR, consistent with the informativeness principle. It also suggests that the selection of index-based benchmarks is less a deliberate act of opportunism but more the inadvertent interaction of board-level governance quality and compensation consultant preferences. Thus, these findings are neither consistent with the view that the executive pay-setting and evaluation process is broken and compromised by powerful executives and captured boards (Morse, Nanda and Seru, 2011; Bebchuk, Cremers and Peyer, 2011) nor the efficiency view that the features of executive compensation reflect the outcomes of optimal contract design (Janakiraman *et al.*, 1992; Aggarwal and Samwick, 1999b; Garvey and Milbourn, 2003).

The remainder 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. Section 3 describes the testable implications for informativeness based on the basic principal-agent framework and provides empirical evidence on the properties of firms' chosen rTSR benchmarks. Section 4 assesses the determinants of firms' benchmark selection choice. Section 5 investigates the potential consequences of rTSR benchmark selection. Section 6 concludes.

2 Data and Descriptive Evidence of rTSR Usage

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.

the ex-ante difference in compensation between using the firm's chosen benchmarks and four alternative peer group is 4.6% of expected award size (Table 8). For firms using specific peers, the ex-ante difference in economically insignificant at approximately 0.09% (Table 5). They also show on an ex-post basis that for firms using either specific peers or index-based benchmarks, compensation was in fact lower using the chosen benchmark relative to alternatives (Tables 6 and 9).

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.

2.1 Growing Prominence of rTSR

Table 1, Panel A, provides summary statistics on 34,321 CEO grants awarded by 1,547 unique firms in fiscal years 2006-2014. 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 2014.⁷

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.⁸ Columns 4

⁷This increase in the explicit use of relative performance grants is consistent with practitioners' observations. See <https://corpgov.law.harvard.edu/2014/11/14/relative-total-shareholder-return-performance-awards>.

⁸Expected compensation is calculated using values reported in the Grants of Plan-Based Awards Table by adding

and 5 estimate the allocated expected compensation stemming from meeting RP-based targets and from meeting rTSR-based targets respectively.⁹ 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%.¹⁰

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

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.

⁹We calculate the weighted portion of expected compensation that corresponds to each performance metric, and assume that each performance metric is weighted equally in the determination of the grant.

¹⁰The 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.

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.

3 Assessing Properties of RP Benchmarks

Given the rising importance of rTSR as a metric for judging and incentivizing managerial performance and apparent desire to isolate the idiosyncratic component of manager’s performance, we seek to examine to what extent the specific rTSR metrics chosen by boards are consistent with the informativeness principle (Holmström, 1979; Holmström and Milgrom, 1987). To guide our analysis, we begin by providing a simple theoretical framework illustrating the empirical properties of a relative performance metric that filters out the systematic component of performance.

3.1 Theoretical Foundations

The theoretical foundation of our empirical examination stems from Holmström and Milgrom (1987) and Gibbons and Murphy (1990), in which there is a risk-neutral principal (the board) and a risk-averse agent (the CEO), and firm performance follows a factor structure consisting of (i) unobserved managerial effort $[a]$, (ii) an observable systematic shock that is beyond the manager’s control $[c \sim_{iid} N(0, \sigma_c^2)]$, and (iii) a firm-specific idiosyncratic shock $[\epsilon \sim_{iid} N(0, \sigma^2)]$

$$p = a + c + \epsilon. \tag{1}$$

The empirical counterpart to p includes stock returns or accounting performance measures.

Under linear contracts, exponential utility, and a quadratic cost of managerial effort, one of the main theoretical implications is that the board should evaluate managerial performance and design compensation contracts on the basis of *relative* performance: $p - c$. That is, to motivate optimal effort by the manager, the principal should design a contract that rewards managers for effort by perfectly eliminating the systematic shocks in firm performance since they are beyond the manager’s control. This implication would apply either in a setting in which explicit compensation (e.g., the usage of rTSR-based performance contracts) or career concerns (e.g., market participants’ evaluation of managers through rTSR) are tied to relative performance.

In practice, boards do not perfectly observe the common risk component of performance, and measure it, through the choice of peers, with some error:

$$\hat{c} = c + \omega_b, \tag{2}$$

where the measurement error (ω_b) has finite variance σ_b^2 . In this framework, better peers should exhibit lower measurement error variance (lower σ_b^2), and perfect measurement of the common risk component of performance is the special case where $\sigma_b^2 = 0$. Put differently, the better the peer selection, the more informative the metric ($p - \hat{c}$) is for managerial performance (a), and the greater the manager’s chosen effort. The intuition is that, all else equal, a lower σ_b^2 means that a smaller proportion of the firm’s performance is unpredictable—even after filtering out the common component of performance. Thus, the decreased risk in the compensation, coupled with the participation constraint, causes the optimal contract to be more sensitive to performance and thereby increasing the manager’s incentives to exert effort.

The above theoretical framework suggests that a relative performance metric that perfectly filters out the systematic component of performance should exhibit the following two testable properties.

Proposition 1 *The performance metric p should exhibit a slope of 1 with respect to the peer benchmark (\hat{c}) used to capture the systematic component of performance.*

This property follows directly from the factor structure of the performance measure (Eqn. (1)).¹¹

¹¹That 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

Thus, the performance measure should also exhibit a slope of 1 in a regression on the peer benchmark that perfectly captures the common risk component of performance.¹²

Proposition 2 *No other measurements of the common risk component of performance, such as alternative peer benchmarks, can better explain the variation in the firm’s performance. Equivalently, no other peer formulations should produce greater explanatory power (i.e., regression R^2) for firm performance (p) than the chosen peer benchmarks (\hat{c}).*

This property follows from the fact that, by definition, the systematic component should capture all of the contemporaneous variation in firm performance that is non-idiosyncratic to the firm. Consequently, if a firm’s chosen peers perfectly capture the systematic component of its performance, then no alternative peer groupings should yield greater explanatory power. More precisely, there is a one-to-one mapping between measurement errors in capturing the systematic component of performance and the regression R^2 . This is because the R^2 —the squared correlation coefficient ($\rho_{p,\hat{c}}^2$)—of the time-series return regression $p_t = \delta_1 + \delta_2\hat{c}_t + \epsilon_t$ can be expressed as a function of the model’s primitives:

$$\hat{R}^2 = \rho_{p,\hat{c}}^2 = \frac{\sigma_c^2 \sigma_c^2}{(\sigma_c^2 + \sigma^2)(\sigma_c^2 + \sigma_b^2)}. \quad (3)$$

Thus, for a given firm—i.e., fixing σ^2 and σ_c^2 —lower σ_b^2 corresponds to higher R^2 .

3.2 Empirical Approach

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

$$R_{it} = \alpha_i + \beta_i R_{pit} + \epsilon_{it} \quad (4)$$

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.

¹²Though not tested, because they do not observe the actual performance metric used by boards, Equation 4 of Janakiraman *et al.* (1992) derives the same prediction.

where R_{it} is firm i 's monthly cum-dividend returns in period t and R_{pit} 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), (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.I.

In estimating Eqn. (4), 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). 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_{pit} . 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.

3.3 Slope Coefficient Results

To examine the first property of the common component of firm performance, we analyze the slopes from estimating Eqn. 4, 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 boards choosing these benchmarks to capture and to filter out the systematic component of TSR.¹³

3.4 R^2 Results

To examine the second property of the common component of firm performance, we analyze the R^2 s from estimating Eqn. 4, summarized in Table 5. Specifically, we examine the return-regression R^2 s produced by 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.¹⁴

Table 5, Panel A, compares the R^2 s generated by firms’ selected benchmarks to those obtained from both the S&P500 index and SBPs.¹⁵ 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,

¹³To assess whether these results are different from chance, in un-tabulated analysis, we randomly draw ten random peers for each firm and compute the average slope coefficients over 1,000 draws (with replacement). In these simulations, we find that firms with index-based benchmarks (specific peers benchmarks) have an average slope of 0.78 (0.85), both statistically different from 1 at the 1% level.

¹⁴Among 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.

¹⁵We 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.

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 exhibiting a better ability to capture the common component of performance. However, in columns 1, 3, and 5, the first row shows that SBPs produce R^2 s (of 51.8%) that are significantly higher (at the 1% level) than those produced by firms' chosen RP benchmarks, suggesting that an alternative peer formulation performs significantly better in capturing the common component of performance relative to the firms' chosen RP benchmarks.

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^2 s 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%.¹⁶ However, we find that the degree of imprecision in firms' chosen benchmarks 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^2 s 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.

Table 5, Panel B, compares the R^2 s 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^2 s produced by firms' selected RP benchmarks are statistically no different from the R^2 s produced by firms' compensation-benchmarking

¹⁶Though index-based benchmarking firms include those that use the S&P500, the difference in R^2 s stems from the narrower industry indexes that are also part of the group.

peers, the subsamples of specific peers and index-based benchmarks reveal very different patterns. Whereas specific peers produce R^2 s 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^2 s that are on average 7.5% lower than those produced by compensation-benchmarking peers, a difference that is statistically significant at the 1% level.

3.5 Robustness Checks

In additional tests, we assess the robustness of the findings in a few ways. We provide additional support to the inference, based on our slope coefficient tests, that firms are attempting to remove systematic noise in TSR. In un-tabulated results, we compare the variance in firms’ chosen rTSR metric and the variance in their realized TSR. These tests are motivated by an implication of the canonical model: evaluating managers by removing systematic noise is desirable only when the noise reduction in systematic risk outweighs the incremental measurement errors introduced by peer benchmarking ($\sigma_c^2 > \sigma_b^2$).¹⁷ In other words, utilizing relative performance metrics may not be sensible if “the cure is worse than the disease.” Consistent with the R^2 results, however, 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 their realized relative performance ($Var(rTSR)$) than their realized absolute performance ($Var(TSR)$). This is true both for firms that choose index-based benchmarks (which exhibit 30% 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).

Also consistent with firms’ desire to filter out systematic noise, we find in untabulated results that out of 26 total firms that switched benchmarks during our sample period, 73% (19 firms) switched from an index benchmark to specific peers. 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). Consistent with our main results, we find that, prior to their switch, index benchmarks generated worse time-series R^2 relative to SBPs; however, after the

¹⁷ Wu (2016) refers to it as the “boundary condition” for relative performance evaluation. 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.

switch to specific peers, their R^2 s becomes insignificantly different from SBPs.

Finally, we examine robustness of the R^2 tests to alternative rules for aggregating peer performance, namely the mean and the 75th 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 (75th 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 75th percentile of specific peer performance. On the other hand, similar to the baseline results using the median, we do not find statistically different R^2 s using the mean of chosen peers' TSRs compared to those produced by SBPs.¹⁸

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 systematic component of firms' returns, these specific-peer benchmarks exhibit a slope of 1 and exhibit R^2 s of magnitudes similar to those of SBPs. This finding may be particularly surprising because it suggests that, although there have been significant findings in the literature of additional trade-offs facing the board that might weaken the need for informative performance measures, 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 systematic component of returns. This finding is particularly interesting 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. 4. In the following section, we explore possible explanations for firms' benchmarking choice.

¹⁸In un-tabulated results, we also examine how results differ by using a third 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.

4 Understanding rTSR Benchmark Choice

We now turn to understand why some boards select specific peers in designing rTSR metrics, while other boards select index-based benchmarks, which are significantly less effective in capturing the common component of firms' TSRs. Put differently, why did certain boards choose to evaluate managers on the basis of noisy measures of relative performance while other boards chose relatively precise measures. Answers to these questions can shed light on the managerial evaluation and compensation design process.

4.1 Empirical Drivers of rTSR Benchmark Choice

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. Consistent with this explanation, studies find that better board monitoring reduces the degree to which managers are evaluated and rewarded on the basis of systematic risk ([Bertrand and Mullainathan, 2001](#); [Garvey and Milbourn, 2006](#)).

Beyond governance, certain fundamental economic attributes of the firm may also render less precise relative performance benchmarks optimal. For example, a high degree of volatility in firm performance or a firm's high growth rate could render the effort to select precise RP benchmarks (or more generally the filtration of systematic shocks) less advantageous ([Gibbons and Murphy, 1990](#); [Albuquerque, 2013](#)). Additionally, [Janakiraman *et al.* \(1992\)](#) and [Aggarwal and Samwick \(1999a\)](#) suggest that in oligopolistic industries, where manager's efforts are correlated with the performance of its peer benchmarks, precise relative performance benchmarks would lead managers to sabotage their industry competitors rather than improve their own performance. As a result, it may be optimal to partially reward CEOs for the systematic shock in order to encourage firms to soften product market competition. Thus, one prediction of such a theory is that firms with greater market power are 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 6. 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*.¹⁹ We also include a census-based Herfindahl-Hirschman Index measure of SIC-based industry concentration (*Census-based HHI Index*) as a measure of competition and market power (Aggarwal and Samwick, 1999b).²⁰ The specifics of variable construction are explained in Table A.II.

Column 1, Table 6 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 Abnormal 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 Abnormal 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 by *CEO Abnormal 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

¹⁹Our 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. Masulis, Wang, Xie and Zhang (2018) argues that older directors display monitoring deficiencies.

²⁰Following 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>.

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 Abnormal 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 Examining 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 can play an important role in the design of CEO compensation packages (e.g. Conyon, Peck and Sadler, 2009; Murphy and Sandino, 2010; Cai, Kini and Williams, 2016). In addition, our background 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 independent of the economic circumstances of their clients.²¹ [Cai et al. \(2016\)](#) provides a particularly acute example where Towers Perrin was accused to have given similar advice about workplace diversity to clients across multiple industries.

Table 7 begins by estimating the main specifications of Table 6, but includes compensation-consultant-fixed effects ([Bertrand and Schoar, 2003](#)).²² 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 Abnormal 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 the χ^2 test 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 controlling for firm-level covariates. We also assess the joint significance of compensation consultants using permutation tests, which [Fee, Hadlock and Pierce \(2013\)](#) argues is a more robust approach than [Bertrand and Schoar \(2003\)](#). In particular, we simulate a placebo distribution of the χ^2 statistic by randomly scrambling the assignment of compensation consultants (without replacement) to firms each draw. After each draw, we estimate a regression of firms' index selection choice on year, industry, and compensation consultant fixed effects, then obtaining the χ^2 statistic from a test of the joint significance of the resultant compensation consultant fixed effects. We perform this procedure 1,000 times and plot the resulting null distribution along with the actual test statistic in Figure 2 Panel A. The figure shows that the test statistic based on the actual assignment of compensation consultants to firms is completely outside of the empirically simulated null distribution based on random assignment. Overall, these empirical results corroborate the qualitative evidence from our interviews, which

²¹We conducted interviews with six compensation consultants and three 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.

²²In our sample, there are 15 compensation consultant firm groups and consultant switches are observed in 20% of the firm-years. [Cai et al. \(2016\)](#) reports higher separation rates but they study a broader sample unrestricted from rTSR-based contracting.

point to different compensation consultant “styles” with respect to the type of relative-performance benchmarks.

One potential concern with this conclusion is that we have not controlled for all possible covariates correlated with firm’s benchmark choice that might also drive the matching process between compensation consultants and firms. Hence it is possible that the tendencies we attribute to compensation consultants instead reflect the firm’s ex-ante (unobserved) desire to match with such a consultant in the first place. To assess this concern, we re-estimate compensation fixed effects without any of the covariates in column 2 except for industry and time fixed effects and the χ^2 remains significant. We then partition the sample by the index preference of the compensation consulting firms into index-preferring (i.e., those compensation consultants whose fixed effect is above the median of all compensation consultants) or specific-peer-preferring (i.e., those compensation consultants whose fixed effect is below or equal to the median of all compensation consultants).²³ Remarkably, column 3 shows that none of the covariates we examine in column 2 are significantly associated with the compensation consultant’s default tendencies towards index or specific peer benchmarks.

Moreover, in a test for the joint significance of CEO characteristics, board and firm characteristics, and industry characteristics for explaining a firm’s choice of compensation consultant styles, we fail to reject to the null that these covariates are jointly insignificant (p -value of joint significance is 0.31). Again, we assess the joint significance of firm characteristics in explaining firms’ choice of compensation consultant styles using permutation tests. In particular, we simulate a placebo distribution of the χ^2 statistic by randomly scrambling the assignment of compensation consultant preferences (without replacement) to firms each draw. After each draw, we estimate a regression of firms’ benchmark preference on firm characteristics and year and industry fixed effects, then we obtain the χ^2 statistic from a test of the joint significance of the firm characteristics. We perform this procedure 1,000 times and plot the resulting null distribution along with the actual test statistic in Figure 2 Panel B. The figure shows that test statistic based on the actual assignment of compensation consultants to firms is at the center of the empirically simulated null distribution, failing to reject the null that compensation consultants’ preferences for index-based versus specific-peer-based

²³The rank ordering of the compensation consultant fixed effects are unchanged in all cases, with and without covariates in Table 7.

benchmarks for evaluating managerial relative performance are unrelated to firms' characteristics.

Our empirical evidence supports the view that boards, in selecting compensation consultants, do not appear to take into consideration the consultants' preferences for relative performance benchmark types. This is consistent with our understanding—from extant literature, interviews with compensation committee members, and boards' public disclosures of compensation consultant selection policies—of how boards select compensation consultants.²⁴ For example, [Ogden and Watson \(2012\)](#) suggests that independence of a compensation consultant is one of the most crucial attributes considered by board members. Moreover, in considering a consultant's fit, boards also consider the consultant's familiarity with the firm's business environment and ability to communicate effectively and objectively with the board ([Pfizer Inc., 2016](#)). Finally, the ideas that specialists on the supply side exhibit “styles” that are i) unrelated to their clientele attributes and ii) can affect clients' outcomes are not new. For example, [Fracassi, Petry and Tate \(2016\)](#) finds evidence of systematic optimism and pessimism among credit analysts, which in turn impacts corporate policies. Another example is the recent literature in health economics, which suggest that there are substantial differences in physicians' diagnostic testing practices that are unrelated to demand-side factors (such as patient characteristics) but impact patients' experiences or outcomes ([Song *et al.*, 2010](#); [Finkelstein *et al.*, 2016](#); [Molitor, 2018](#); [Cutler *et al.*, 2019](#)).

To better understand the influence that these compensation consultant preferences could exert, we further seek to understand the observed systematic associations between firms' choice of index-based benchmarks and characteristics reflecting board-level governance weaknesses. One hypothesis (the “lack of scrutiny” hypothesis) is that, given the preferences of certain compensation consultants, a board that does not carefully scrutinize the consultant's recommendations (e.g., a weak governance board) would be more likely to accept them as they are. On the other hand, a board that scrutinizes the consultant's choice would be more likely to choose specific peers, and thus more likely to overrule a compensation consultant's default preference for index-based peers. An alternative hypothesis (the “deliberate choice” hypothesis) for why firms' choice of index-based benchmarks is systematically associated with characteristics reflecting board-level governance weaknesses could be due to the inherent preferences of firms for benchmarks types. In particular, weak governance boards could have preferences for index-based peers and, irrespective of compensation consultants' preferences, ensure

²⁴In our sample, 97% of compensation consultants are retained by the board.

the selection of index-based peers ex-post. Similarly, stronger governance boards have preferences for custom-tailored specific peers and, irrespective of compensation consultants' preferences, ensure the selection of specific peers ex-post.

To distinguish between these two hypotheses for why we observe a systematic association between firms' choice of index-based benchmarks and characteristics reflecting board-level governance weaknesses, we examine these associations in the subsample of firms that choose index-preferring consultants and the subset of firms that choose specific-peer-preferring consultants. The two hypotheses predict different patterns in these subsamples. Under the deliberate choice hypothesis, firms exercise their preference irrespective of compensation consultant styles, and we should therefore observe the association between board-level governance weaknesses and the choice of an index in both subsamples. However, under the lack of scrutiny hypothesis, we should only observe an association between board-level governance weaknesses and the choice of an index within the subsample of firms that hire index-preferring consultants. This is because 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 quality and the choice of an index for firms that hire index-preferring consultants. 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. Moreover, under the lack of scrutiny hypothesis, we should expect a greater likelihood of defaulting 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.

Table 7, Columns 4 and 5, estimate the specification of column 2 using the subsample of firms that hired index-preferring consultants and the subsample of firms that hired specific-peer-preferring consultants, respectively. 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 the lack of scrutiny hypothesis, we find that the significance in *CEO Abnormal*

Pay and *Director Workload* are concentrated in the set of firms whose compensation consultants prefer index-based benchmarks (column 4).²⁵ 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 5), inconsistent with the deliberate choice hypothesis. While *Board Size* and *% Age 65+ Directors* lose significance in both subsamples, the size of the coefficients are much larger in column 4 versus column 5. Furthermore, we report in the last row of Table 7 that, consistent with the lack of scrutiny hypothesis, the propensity to overrule compensation consultant’s preferences is nearly three times as high in the set of firms who hire index-preferring compensation consultants. Whereas firm-year observations matched to specific-peer-preferring consultants overrule their consultants only 20% of the time, firms matched to index preferring consultants overrule about 60% of time. Given that the assignment of index-preferring consultants to firms appears conditionally unconfounded, these descriptive statistics also suggest the magnitude of being assigned to an index-preferring consultant on the choice of an index is large. In un-tabulated results, we find being assigned to an index-preferring consultant increases the probability of selecting an index by 23.7%: a 60% proportional response relative to the 40% baseline likelihood of selecting indexes.

One potential concern with this analysis is that we have not examined all possible covariates that could be correlated with the matching process between compensation consultants and firms. In Section 5, we also include compensation consultant fixed effects in regressions with firm performance. To the extent that such fixed effects are unrelated to firm performance, which is the case, it alleviates the potential concern that unobservables correlated with firm performance also drive firms’ selection of compensation consultants.

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 and in juxtaposition with our empirical findings.

²⁵The fact that Abnormal Pay is only significant in column 4 and not 5 further suggests that the variable is capturing governance, instead of a potential mechanical effect on pay through the choice of an index as suggested in Bizjak *et al.* (2016).

4.3.1 Opportunism

One possibility is that the selection of indexes reflects the board’s deliberate trade-off between informativeness and opportunism (Gong *et al.*, 2011). Bizjak *et al.* (2016) reports that the selection of an index results in a simulated average 4% increase in the relative-performance-based portion of compensation relative to credible alternative peer sets. However the analysis in Table 7 suggests that boards’ lack of scrutiny over the default preferences of their outside expert rather than deliberate opportunism as the governing reason behind the choice of an index.²⁶

Alternatively, it is possible that the selection of index-based benchmarks could be due to concerns that the selection of specific peers in relative-performance contracts may provide the *appearance* of opportunism (Murphy, 2002). Thus, if firms with characteristics associated with poor governance are also more sensitive to the external perception of poor governance, they may prefer index benchmarks. However, our empirical evidence is not consistent with this alternative as the primary driver of firm behavior. Under such a hypothesis, one would expect governance characteristics to be associated with the choice of index-based benchmarks *regardless* of the compensation consultant’s preferences. 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.

4.3.2 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.

²⁶Furthermore, a simple back-of-the-envelope estimate suggests that a manager with exponential normal utility would not willfully accept an alternative contract that increases expected compensation by 4% but also its variance by 16% (based on an un-tabulated estimate of relative measurement error variances between index benchmarks and SBPs implied from Tables 5). Assuming $\eta = 0.02$ (the lower bound estimate of risk aversion in Gayle and Miller (2015)), $\mu = 1,500,000$ (Table 2), the manager would prefer the safer contract unless the baseline variance is small ($\frac{\sigma}{\mu} < 0.007$).

In particular, [Lazear and Rosen \(1981\)](#) shows 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).

4.3.3 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.

This theory does not appear to explain the role of performance peers in relative performance contracts. For one, market participants' comment letters, compensation consultants' whitepapers, and interviews with practitioners, all suggest that the primary objective of rTSR is to provide a measure of firm-specific performance or shareholder value improvement that removes the effect of systematic shocks, rather than as a measure that motivates performance via aspiration. 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.

Overall, at a basic level, none of the alternative explanations specifically predict that firms' chosen rTSR benchmarks to exhibit a benchmark-return-beta of 1. While we have not explored all possible explanations, 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 Consequences of Noisy Benchmarks: An Exploratory Analysis

In our final analysis, we explore the potential firm performance consequences of evaluating managers on the basis of noisy benchmarks. In theory, there are at least two channels why noisy benchmarks can affect firm performance. One channel is through managers' *explicit* incentives due to explicit contracts (Mehran, 1995; Lazear, 2000; Core and Larcker, 2002; Brick *et al.*, 2012). A less noisy relative performance metric allows the board to increase the pay-performance sensitivity and increases the manager's incentives to exert effort. Another channel is through managers' *implicit* incentives due to their career concerns (e.g., Avery, Chevalier and Schaefer, 1998; Chevalier and Ellison, 1999; Brickley, Linck and Coles, 1999; Fee and Hadlock, 2003; Milbourn, 2003). The intuition is that a less noisy metric is more informative of the manager's talent, which in turn motivates the manager to increase effort in order to signal the perceived talent (Holmström, 1999).

We perform a comprehensive, albeit exploratory, empirical analysis of the performance consequences of index selection through three different approaches. First, we estimate, through standard OLS, the cross-sectional association between realized operating performance and index-peer selection after controlling for other determinants of performance. However, this reduced-form analysis treats firms that use specific-peer benchmarks as counterfactuals to firms that use index-based benchmarks and is therefore subject to endogeneity concerns. To address these concerns, our second approach estimates the performance consequences of index selection through two-stage least squares (2SLS). We do so by leveraging the observation (from Table 7) that compensation consultants have fixed preferences for benchmark types and that the assignment of these preferences to firms appears to be conditionally unconfounded. Finally, we complement the prior analyses by implementing the analytical framework proposed by Altonji, Elder and Taber (2005) and Oster (2017), which allows for an explicit assessment of the extent to which omitted variables could influence the OLS treatment effects. Together, the various approaches serve to provide an overall assessment of the direction of the effect of benchmark selection in rTSR metric design for managerial incentives and firm performance.

5.1 Reduced Form OLS Estimates

Columns 1 and 2, Table 8, report results of OLS regressions of firm performance, in terms of ROA, on the choice of index-based benchmarks, after controlling for the same CEO, board, firm, industry characteristics, and time and industry fixed effects as in Table 7. Column 1 of the table reports that firms choosing index-based performance benchmarks are associated, on average, with 89 basis points lower ROA, which is statistically significant at the 5% level. This effect size represents a fifth of a standard deviation in ROA. In column 2 we include compensation consultant fixed effects, which lowers the estimated coefficient slightly but does not change its significance.

As in the prior analyses in Table 7, we test for the joint significance of these fixed effects and fail to reject the null of no compensation-consultant effects. This result has the following implications. First, this implies that the direction of causality does not flow through a direct treatment effect of compensation consultants to firm performance independent of the choice of an index-benchmark. Second, it also rules out the reverse causality concern of firm performance as a determinant of the choice of the compensation consultant in the first place. However, there may be other omitted variables that could still account for these effects, which motivates our two-stage least squares approach.

5.2 Instrumental Variables Approaches

Our findings in Table 7 that i) compensation consultants have fixed preferences for benchmark types and that ii) these preferences are uncorrelated with observable firm characteristics raises the possibility that the assignment of these consultant preferences to firms may be conditionally unconfounded. Based on the intuition that, conditional on firm and CEO characteristics, consultants' fixed preferences for index-based benchmarks could influence performance *only through* boards' choice of an index-based benchmark, we leverage these consultant styles as instruments to estimate the performance effects of index selection via 2SLS. The first and second stages of the 2SLS estimation, using the vector of compensation consultant fixed effects as instrumental variables for index selection, are reported in columns 3 and 4 of Table 8.

Column 3 shows the results from the first stage.²⁷ Notably, the χ^2 statistic for the joint

²⁷The first stage is estimated using a linear probability model to avoid the forbidden regression problem (Kelejian, 1971; Abrevaya, Hausman and Khan, 2010). For convention, we describe the estimator as 2SLS, but we actually use

significance of the compensation consultant fixed effects is strongly significant (at the 1% level), suggesting instrument relevance. More formally, because we have a single endogenous regressor, we implement the test for weak instruments recently developed by [Olea and Pflueger \(2013\)](#), which is robust to heteroscedasticity, serial correlation, and clustering in the error term.²⁸ This test produces an F-statistic of 4.85, which is above the 5% critical value of 4.41, rejecting the null hypothesis of weak instruments. The statistical test thus confirms the intuition that compensation consultants' fixed preferences are important in determining the selection of index-based benchmarks.

Column 4 reports the results of the second stage, which suggests that the average effect of selecting an index-based benchmark on ROA is negative and statistically significant at the 10% level, consistent with the OLS estimates. These estimates are credible insofar as the instruments, in addition to satisfying the relevance condition, satisfy the exogeneity condition. We argued and showed evidence that the compensation consultants' fixed preferences for index appear to be conditionally unconfounded because they are unrelated to any observable firm characteristics that we examined (i.e., column 3 of [Table 7](#)). Moreover, column 2 of [Table 8](#) suggests that consultant fixed effects are not associated with firm performance, after controlling for the index selection choice, suggesting that compensation consultants' fixed preferences do not have a *direct* effect on firm performance, consistent with the exclusion restriction. Finally, because we have an over-identified system of equations, with one endogenous regressor and more than one instrument, we can formally test the null hypothesis that the instruments are jointly valid. Indeed, Sargan-Hansen J-test of overidentifying restrictions fails to reject the null of valid instruments (p -value=0.628). Collectively, these findings suggest that it is unlikely for the compensation consultant preferences for indexes to impact firm performance except through their impact on the benchmark selection decision.

The estimates are larger than our OLS results at about half of a standard deviation in ROA. A likely explanation is that column 4 estimates a local average treatment effect for the set of “complier” firms, which are most susceptible to the consultant's fixed preference in the index selection choice ([Angrist et al., 1996](#)). These are likely the firms for which board monitoring is a weak control on managerial actions, and where other control mechanisms—such as explicit incentives—are relatively more important. In this way, the complier treatment effects serve as an upper bound of the average

LIML to estimate the system of equations due to its better finite sample properties ([Stock and Yogo, 2002](#)).

²⁸Homoscedasticity is a required assumption for the standard rule of thumb tests illustrated in [Staiger and Stock \(1997\)](#) and [Stock and Yogo \(2002\)](#).

performance effect of index selection.

In column 5, we again take advantage of the fact that we have an over-identified system of equations and report the efficient GMM estimate of the treatment effect. We find a treatment-effect magnitude similar to the 2SLS results, but the efficient GMM estimate is significant at the 1% level as it is, by design, estimated with greater precision.

Finally, we complement these analyses by reporting in column 6 the results from an OLS regression of ROA on an indicator variable for the assignment to an index-preferring compensation consultant. This regression shows that having an index-preferring consultant is associated with 50 basis point lower ROA. Given our prior evidence that the assignment of consultants' index preference appears to be conditionally unconfounded, this result can be interpreted as an "intent-to-treat" (ITT) estimate, which can be thought of as a lower bound estimate of the average performance effect of index selection.

5.3 Assessing Omitted Variable Bias

Because the 2SLS estimates represent complier average treatment effects, it can be difficult to compare them to the OLS estimate. Thus, we perform an alternative assessment of the OLS treatment-effect estimates by assessing the degree to which omitted variables bias could impact results.

We follow the methodology proposed by [Oster \(2017\)](#), which is based on the insight ([Altonji, Elder and Taber, 2005](#)) that the amount of selection between the treatment and the observed set of controls can be informative of the degree of selection on unobservables and therefore useful in bounding the magnitudes of potential omitted variable bias in OLS estimates. Applied to our research setting, [Oster \(2017\)](#) suggests that the sensitivity of our OLS estimate of the performance-effects of index selection depends on i) the degree to which the omitted variables are correlated with the selection of an index, and ii) the extent to which the omitted variables contributes to explaining firm performance.

Table 9 reports the results from the [Oster \(2017\)](#) approach in estimating bias-adjusted performance effects of index selection. We provide a range of bias-adjusted estimates based on variations in two technical parameters: δ , which captures the degree of correlation between the omitted variables and index selection, and R_{max}^2 , which captures the importance of the omitted variables for explaining

firm performance.²⁹ In terms of the range of δ , we consider $\delta = 1$, which occurs if selection on unobservables is similar to selection on observables (the variables omitted by the researchers are as important as the included controls), and $\delta = 0.5$, which occurs if selection on unobservables is smaller than selection on observables (the included variables are more important than the omitted variables). We also report the cases where δ is negative: this represents situations where the omitted variables are associated with the index in the opposite sign of the association with the observables. We also consider the bias-adjusted estimates by varying the theoretical R_{max}^2 that would be achieved if we included the unobservables to identify the treatment effect. Oster (2017) recommends a R_{max}^2 to be set as 130% of the R^2 achieved with the observed controls ($R^2 \approx 0.45$). For completeness, we also report results assuming a theoretical R_{max}^2 that is 200% of the in-sample R^2 ($R^2 \approx 0.70$) and also assuming a theoretical R_{max}^2 of 100%.³⁰ Panel A shows that when $R_{max}^2 = 1.3X$, the estimated performance-effect of index selection remains significant for all four values of δ , and the effect magnitudes range from -70 basis points to -110 basis points. These results suggest that the baseline OLS estimates (e.g., Table 8 column 2) are robust to omitted variables that can account for an additional 30% of the variation in firm performance relative to the included set of controls. Panel B shows that when $R_{max}^2 = 2X$, the estimated parameters range from -40 basis points to -180 basis points and lose significance when $\delta = -1$, suggesting that the OLS results remain fairly robust even when inclusion of omitted variables can double the amount of variation in firm performance explained. Finally, Panel C shows that when the $R_{max}^2 = 1$, the theoretical upper bound in which the inclusion of omitted variables explains 100% of the variation in firm performance. Even in this extreme scenario, we continue to find significance for $\delta = 0.5$ and $\delta = 1$ and in economic magnitudes similar to the IV estimates. When $\delta < 0$, however, the estimates become smaller and lose significance, but remain negative. Overall, while the degree of bias depends on how much information one believes the unobservables are capturing, our results suggest that the performance-effects of index selection remain negative and significant under a wide range of plausible assumptions on the degree of selection on unobservables as well as the amount of information captured by unobservables. Together with our baseline OLS results, the two-stage least

²⁹Formally, δ is the parameter such that $\delta \frac{\sigma_{index,observable}^2}{\sigma_{observables}^2} = \frac{\sigma_{index,unobservables}^2}{\sigma_{unobservables}^2}$. Moreover, R_{max}^2 is the maximum R^2 that could be achieved if we included all the unobservables to identify the treatment effect.

³⁰Of the non-experimental published studies analyzed, Oster (2017) reports that 45% would survive a R_{max}^2 of 130% of the R^2 with full controls, which falls to 27% when $R_{max}^2 = 200\%$ of the R^2 with full controls, and finally to between 9% to 16% when $R_{max}^2 = 1$.

squares, GMM, and ITT estimates, our exploratory analyses provide evidence that the selection of an index leads to lower ROA.

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 assesses the extent to which boards' choices of rTSR benchmarks evaluate managers on the basis of the systematic component of TSR. We find that over half of the firms that tie CEO performance-based contracts to rTSR—those that choose specific peers as benchmarks—do a remarkable job of measuring and filtering out the systematic risk in TSR. 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](#)) and helpful in dispelling any perceptions that the selection of specific peers could be misconstrued as opportunism.

However, firms that choose index-based benchmark evaluate managers on the basis of rTSR metrics that retain a significant amount of systematic risk. The selection of indexes appears driven by compensation consultants' default tendencies in conjunction with a lack of scrutiny on the part of boards with governance weaknesses. These findings provide new evidence on both the role that compensation consultants play in the managerial evaluation process, and highlights a novel aspect of boards' monitoring role: scrutinizing outside experts' advice. It also offers a new explanation for why managers may be evaluated based on systematic noise and the economic consequences for firm performance.

A limitation of our study is that we restrict attention to firms that explicitly tie rTSR to compensation. However, the prevalence use of rTSR by market participants suggests that there could be important implicit incentive effects, consistent with our empirical results, independent of the explicit compensation contract. For example, if shareholders, board members, or the executive labor

market assess managers' competence at least partly on the basis of rTSR, managers' reputational, career, or status 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

- ABREVVAYA, J., HAUSMAN, J. A. and KHAN, S. (2010). Testing for causal effects in a generalized regression model with endogenous regressors. *Econometrica*, **78** (6), 2043–2061. [29](#)
- 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](#), [5](#), [18](#)
- and — (1999b). The other side of the trade-off: The impact of risk on executive compensation. *Journal of Political Economy*, **107** (1), 65–105. [7](#), [19](#)
- ALBUQUERQUE, A. (2009). Peer firms in relative performance evaluation. *Journal of Accounting and Economics*, **48** (1), 69–89. [5](#)
- ALBUQUERQUE, A. M. (2013). Do growth-option firms use less relative performance evaluation? *The Accounting Review*, **89** (1), 27–60. [18](#)
- , DE FRANCO, G. and VERDI, R. S. (2013). Peer choice in CEO compensation. *Journal of Financial Economics*, **108** (1), 160–181. [6](#)
- 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. [19](#)
- 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. [28](#), [31](#), [54](#)
- ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, **91** (434), 444–455. [30](#)
- 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. [2](#), [5](#)
- ARMSTRONG, C. S., ITTNER, C. D. and LARCKER, D. F. (2012). Corporate governance, compensation consultants, and ceo pay levels. *Review of Accounting Studies*, **17** (2), 322–351. [6](#)
- AVERY, C., CHEVALIER, J. A. and SCHAEFER, S. (1998). Why do managers undertake acquisitions? an analysis of internal and external rewards for acquisitiveness. *Journal of Law, Economics, & Organization*, **14** (1), 24–43. [4](#), [28](#)
- BECHUK, L. A., CREMERS, K. M. and PEYER, U. C. (2011). The CEO pay slice. *Journal of Financial Economics*, **102** (1), 199–221. [7](#), [33](#)
- BENNETT, B., BETTIS, J. C., GOPALAN, R. and MILBOURN, T. (2017). Compensation goals and firm performance. *Journal of Financial Economics*, **124** (2), 307–330. [13](#)
- BERTRAND, M. and MULLAINATHAN, S. (2001). Are CEOs rewarded for luck? the ones without principals are. *Quarterly Journal of Economics*, **116** (3), 901–932. [5](#), [18](#)
- and SCHOAR, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, **118** (4), 1169–1208. [6](#), [21](#)
- 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*. [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](#), [25](#), [26](#)

- , LEMMON, M. L. and NAVEEN, L. (2008). Does the use of peer groups contribute to higher pay and less efficient compensation? *Journal of Financial Economics*, **90** (2), 152–168. [6](#)
- 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. [1](#)
- BRICK, I. E., PALMON, O. and WALD, J. K. (2012). Too much pay-performance sensitivity? *Review of Economics and Statistics*, **94** (1), 287–303. [4](#), [28](#)
- BRICKLEY, J. A., LINCK, J. S. and COLES, J. L. (1999). What happens to CEOs after they retire? new evidence on career concerns, horizon problems, and CEO incentives. *Journal of Financial Economics*, **52** (3), 341–377. [5](#), [28](#)
- CADMAN, B. and CARTER, M. E. (2013). Compensation peer groups and their relation with CEO pay. *Journal of Management Accounting Research*, **26** (1), 57–82. [6](#)
- , — and HILLEGEIST, S. (2010). The incentives of compensation consultants and ceo pay. *Journal of Accounting and Economics*, **49** (3), 263–280. [6](#)
- CAI, C., KINI, O. and WILLIAMS, R. (2016). Do compensation consultants have distinct styles? *SSRN Working Paper 2724072*. [6](#), [20](#), [21](#)
- CHENG, S. (2008). Board size and the variability of corporate performance. *Journal of Financial Economics*, **87** (1), 157–176. [19](#)
- CHEVALIER, J. and ELLISON, G. (1999). Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, **114** (2), 389–432. [4](#), [28](#)
- CONYON, M. J., PECK, S. I. and SADLER, G. V. (2009). Compensation consultants and executive pay: Evidence from the united states and the united kingdom. *Academy of Management Perspectives*, **23** (1), 43–55. [6](#), [20](#)
- 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. [19](#), [41](#)
- , HOLTHAUSEN, R. W. and LARCKER, D. F. (1999). Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics*, **51** (3), 371–406. [5](#)
- and LARCKER, D. F. (2002). Performance consequences of mandatory increases in executive stock ownership. *Journal of Financial Economics*, **64** (3), 317–340. [4](#), [28](#)
- CREMERS, M. and GRINSTEIN, Y. (2013). Does the market for ceo talent explain controversial ceo pay practices? *Review of Finance*, **18** (3), 921–960. [6](#)
- CUTLER, D. M., SKINNER, J. S., STERN, A. D. and WENNBURG, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *NBER Working Paper*. [23](#)
- DE ANGELIS, D. and GRINSTEIN, Y. (2016). Relative performance evaluation in CEO compensation: A non-agency explanation. *SSRN Working Paper 2432473*. [9](#), [27](#)
- 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. [5](#)
- FAULKENDER, M. and YANG, J. (2013). Is disclosure an effective cleansing mechanism? the dynamics of compensation peer benchmarking. *Review of Financial Studies*, **26** (3), 806–839. [6](#)
- FEE, C. E. and HADLOCK, C. J. (2003). Raids, rewards, and reputations in the market for managerial talent. *Review of Financial Studies*, **16** (4), 1315–1357. [5](#), [28](#)

- , — and PIERCE, J. R. (2013). Managers with and without style: Evidence using exogenous variation. *The Review of Financial Studies*, **26** (3), 567–601. [21](#)
- FICH, E. M. and SHIVDASANI, A. (2006). Are busy boards effective monitors? *Journal of Finance*, **61** (2), 689–724. [5](#), [19](#)
- FINKELSTEIN, A., GENTZKOW, M. and WILLIAMS, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The quarterly journal of economics*, **131** (4), 1681–1726. [23](#)
- FRACASSI, C., PETRY, S. and TATE, G. (2016). Does rating analyst subjectivity affect corporate debt pricing? *Journal of Financial Economics*, **120** (3), 514–538. [23](#)
- FRANCIS, B., HASAN, I., MANI, S. and YE, P. (2016). Relative peer quality and firm performance. *Journal of Financial Economics*, **122** (1), 196 – 219. [4](#), [27](#)
- 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](#), [7](#), [18](#), [20](#)
- GARVEY, G. T. and MILBOURN, T. T. (2006). Asymmetric benchmarking in compensation: Executives are rewarded for good luck but not penalized for bad. *Journal of Financial Economics*, **82** (1), 197–225. [5](#), [18](#)
- GAYLE, G.-L. and MILLER, R. A. (2015). Identifying and testing models of managerial compensation. *Review of Economic Studies*, **82** (3), 1074–1118. [26](#)
- GHOSH, M. and JOHN, G. (2000). Experimental evidence for agency models of salesforce compensation. *Marketing Science*, **19** (4), 348–365. [5](#)
- GIBBONS, R. and MURPHY, K. J. (1990). Relative performance evaluation for chief executive officers. *Industrial & Labor Relations Review*, **43** (3), 30S–51S. [2](#), [5](#), [10](#), [18](#)
- 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. [2](#), [6](#), [26](#)
- HEMMER, T. (2015). Optimal dynamic relative performance evaluation. *Rice University Working Paper*. [4](#), [27](#)
- HOLMSTRÖM, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, **10** (1), 74–91. [1](#), [10](#), [17](#), [33](#)
- HOLMSTRÖM, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, **13** (2), 324–340. [1](#)
- HOLMSTRÖM, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, **66** (1), 169–182. [28](#)
- HOLMSTRÖM, B. and MILGROM, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, **55** (2), 303–328. [1](#), [10](#)
- HUGESSEN (2016). Assessing relative tsr for your company: A brief overview. *Hugessen Consulting White Paper*. [2](#)
- HVIDE, H. K. (2002). Tournament rewards and risk taking. *Journal of Labor Economics*, **20** (4), 877–898. [4](#), [26](#)
- 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. [4](#), [7](#), [12](#), [18](#)
- JENSEN, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *the Journal of Finance*, **48** (3), 831–880. [19](#)

- JENTER, D. and KANAAN, F. (2015). CEO turnover and relative performance evaluation. *Journal of Finance*, **70** (5), 2155–2184. [5](#)
- KEIL, J. (2017). The trouble with approximating industry concentration from compustat. *Journal of Corporate Finance*, **45**, 467–479. [19](#), [41](#)
- KELEJIAN, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association*, **66** (334), 373–374. [29](#)
- KLEINBERG, J., LIANG, A. and MULLAINATHAN, S. (2017). The theory is predictive, but is it complete?: An application to human perception of randomness. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, ACM, pp. 125–126. [3](#)
- LAMBERT, R. A. and LARCKER, D. F. (1987). An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research*, pp. 85–125. [5](#)
- LAZEAR, E. (2000). Performance pay and productivity. *American Economic Review*, **90** (5), 1346–1361. [28](#)
- LAZEAR, E. P. and ROSEN, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, **89** (5), 841–864. [4](#), [26](#), [27](#)
- 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](#), [13](#), [14](#), [50](#)
- , — and — (2016). The search for peer firms: When do crowds provide wisdom? *Harvard Business School Working Paper*. [3](#), [14](#), [17](#)
- LEEFLANG, K., O’BYRNE, S. and VAN CLIEAF, M. (2014). The alignment gap between creating value, performance measurement, and long-term incentive design. [1](#)
- LEWELLEN, S. (2015). Executive compensation and industry peer groups. *London Business School Working Paper*. [5](#)
- and METRICK, A. (2010). Corporate governance and equity prices: Are results robust to industry adjustments. *Yale University Working*. [15](#)
- MASULIS, R. W., WANG, C., XIE, F. and ZHANG, S. (2018). Directors: Older and wiser, or too old to govern? *SSRN Working Paper*. [19](#)
- MEHRAN, H. (1995). Executive compensation structure, ownership, and firm performance. *Journal of financial economics*, **38** (2), 163–184. [4](#), [28](#)
- MILBOURN, T. T. (2003). CEO reputation and stock-based compensation. *Journal of Financial Economics*, **68** (2), 233–262. [5](#), [28](#)
- MOLITOR, D. (2018). The evolution of physician practice styles: evidence from cardiologist migration. *American Economic Journal: Economic Policy*, **10** (1), 326–56. [23](#)
- MORSE, A., NANDA, V. and SERU, A. (2011). Are incentive contracts rigged by powerful CEOs? *Journal of Finance*, **66** (5), 1779–1821. [7](#), [33](#)
- MURPHY, K. J. (2002). Explaining executive compensation: Managerial power versus the perceived cost of stock options. *University of Chicago Law Review*, **69**, 847. [4](#), [26](#)
- and SANDINO, T. (2010). Executive pay and “independent” compensation consultants. *Journal of Accounting and Economics*, **49** (3), 247–262. [6](#), [20](#)
- ODGEN, S. and WATSON, R. (2012). Remuneration committees, pay consultants and the determination of executive directors’ pay. *British Journal of Management*, **23** (4), 502–517. [23](#)

- OLEA, J. L. M. and PFLUEGER, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, **31** (3), 358–369. [30](#)
- OSTER, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, pp. 1–18. [4](#), [28](#), [31](#), [32](#), [54](#)
- PFIZER INC. (2016). *Criteria for the Selection of a Compensation Committee Consultant*. Tech. rep., Pfizer Corporation. [23](#)
- REDA, J. F. and TONELLO, M. (2015). The conference board CEO and executive compensation practices 2015 edition key findings. *SSRN Working Paper 2702563*. [13](#)
- SHIN, J.-E. (2016). An evaluation of compensation benchmarking peer groups based on mutual peer-designating behaviors. *SSRN Working Paper 2823592*. [27](#)
- SONG, Y., SKINNER, J., BYNUM, J., SUTHERLAND, J., WENNBERG, J. E. and FISHER, E. S. (2010). Regional variations in diagnostic practices. *New England Journal of Medicine*, **363** (1), 45–53. [23](#)
- STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, pp. 557–586. [30](#)
- STOCK, J. H. and YOGO, M. (2002). Testing for weak instruments in linear iv regression. [30](#)
- SWINFORD, D. (2015). The limits of using tsr as an incentive measure. *Pearl Meyer Partners White Paper*. [2](#)
- WU, M. G. H. (2016). Optimal risk trade-off in relative performance evaluation. *SSRN Working Paper 2288042*. [16](#)
- YERMACK, D. (1996). Higher market valuation of companies with a small board of directors. *Journal of financial economics*, **40** (2), 185–211. [19](#)

Table A.I.
Sample Selection

Panel A of this table reports the selection criterion used to generate the final samples used in Tables 4 and 5. Panel B reports the selection criterion used to generate the final samples used in Tables 6, 7, 8, and 9.

Panel A: Properties of rTSR Benchmarks Sample

<i>Main Sample Selection</i>	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		(6,654)	(131)
(8) Less observations before calendar year 2006		(764)	(4)
(9) Less observations that use both, index and specific peers, in a given fiscal year		(1,107)	(11)
(10) Less observations with fewer than 10 monthly returns in the time-series regressions		(77)	(11)
Final Sample		13,108	330

Panel B: Benchmarking Choice Sample

<i>Main Sample Selection</i>	Firm-year Observations	Unique Firms
(1) Firm-year observations after step (5) from above	1,821	487
(2) After confining to firm-benchmark sample used in Tables 4, 5	1,444	330
Final Sample after merging in additional firm characteristics as described in Table A.II	1,175	299

Table A.II.
Descriptive Statistics

Panel A reports summary statistics on the variables used in Tables 6, 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 $\text{Log}(CEOTenure_{i,t})$, $\text{Log}(Sales_{i,(t-1)})$, *Book-to-Market*_{*i,(t-1)*}, a dummy equal to 1 if the firm is included in the S&P500, lagged and contemporaneous annual stock return, and *ROA*, and industry controls. The expected value from the determinant model is exponentiated (*CEO Expected Pay*), and *CEO Abnormal 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

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

Table A.II.
Continued

Panel B: Correlation Matrix

<i>ROA</i>	1.00																		
<i>Index</i>	-0.04	1.00																	
<i>CEO Total Pay</i>	0.17***	0.16***	1.00																
<i>CEO Expected Pay</i>	0.21***	0.01	0.66***	1.00															
<i>CEO Abnormal Pay</i>	0.06**	0.19***	0.74***	-0.02	1.00														
<i>CEO Tenure</i>	0.02	0.07**	0.02	-0.04	0.06**	1.00													
<i>CEO Age</i>	0.07**	0.03	0.05*	0.08**	0.01	0.37***	1.00												
<i>% Busy Directors</i>	0.01	0.05*	0.05*	0.06**	-0.00	-0.03	-0.07**	1.00											
<i>Board Size</i>	-0.01	0.06**	0.22***	0.33***	-0.01	-0.07**	0.05*	0.05*	1.00										
<i>Director Workload</i>	-0.08***	0.09***	0.00	-0.06**	-0.00	-0.06**	-0.09***	0.06**	-0.35***	1.00									
<i>% Age 65+ Directors</i>	0.03	0.09***	0.07**	0.12***	-0.03	0.07**	0.12***	-0.09***	0.08***	-0.09***	1.00								
<i>Log Market Cap</i>	0.31***	0.05	0.62***	0.81***	0.12***	-0.05*	0.06**	0.06**	0.44***	-0.09***	0.12***	1.00							
<i>Census-based HHI Index</i>	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						
<i>Return Volatility</i>	-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					
<i>Book-to-Market</i>	-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.III.
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

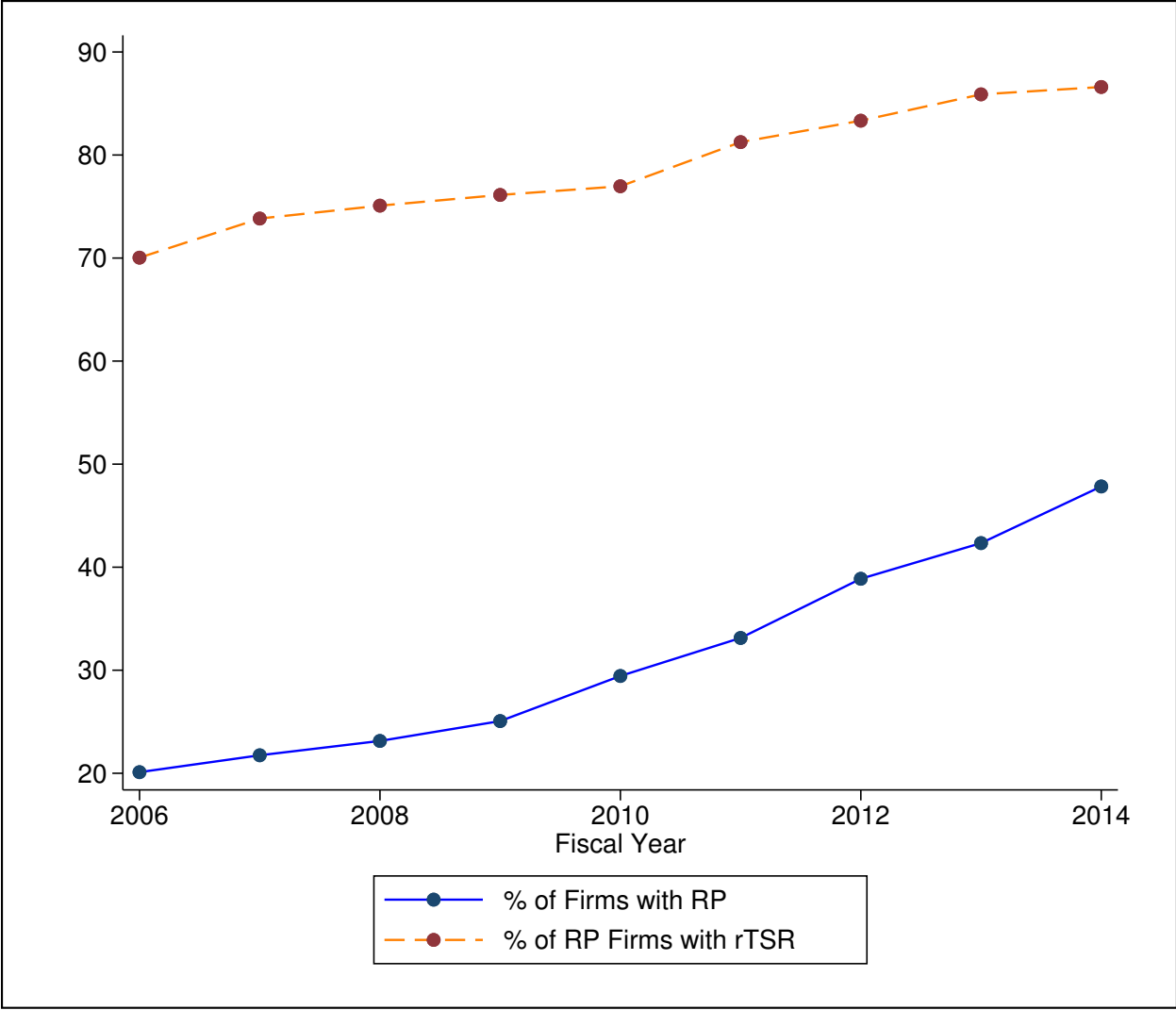
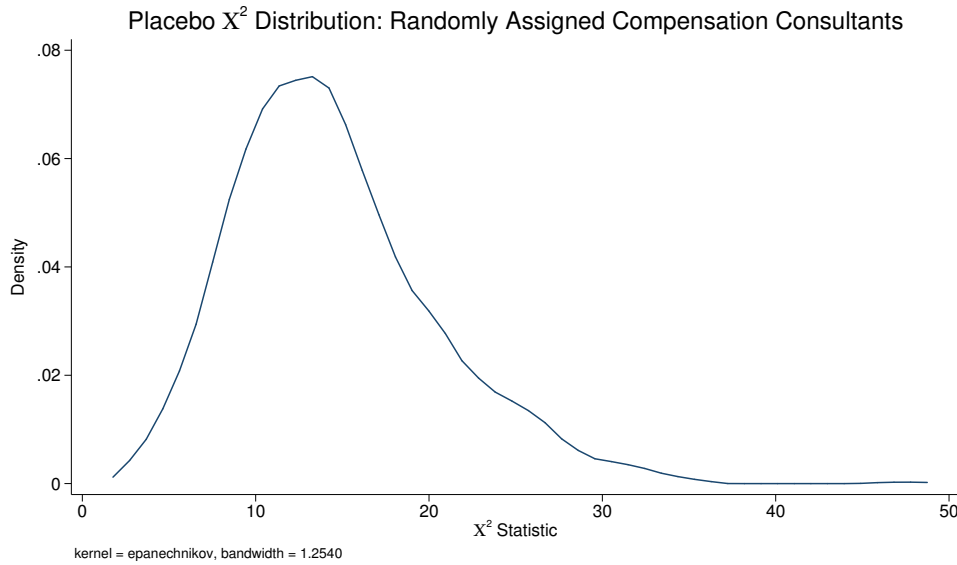
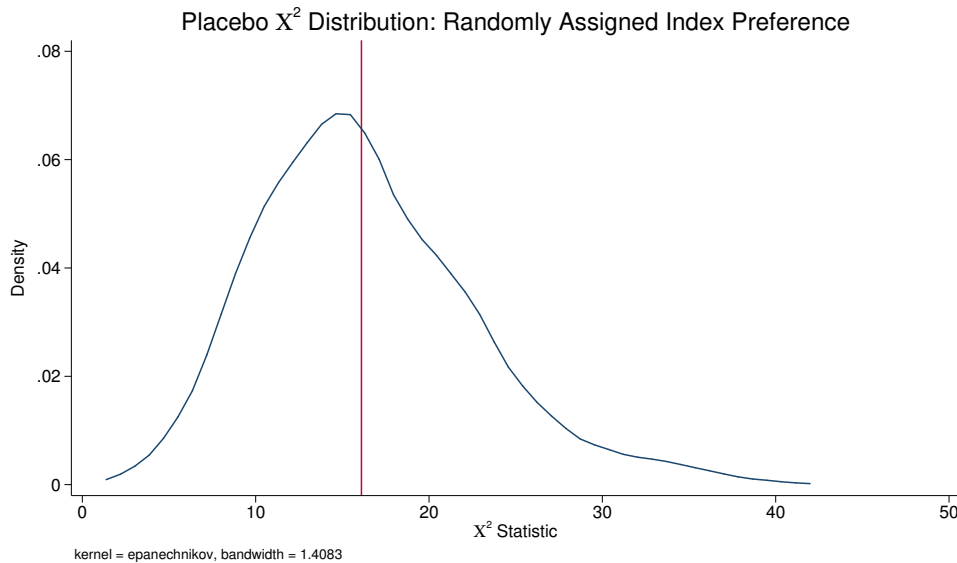


Figure 1. Fraction of Firms Using Relative Performance Contracts 2006-2014 The solid line plots the fraction of firms in the ISS Incentive Labs sample prior to any sample selection restrictions that disclose awarding at least one performance grant based on relative performance (RP) in a given fiscal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



A: Null Distribution of χ^2 Tests of Joint Significance of Compensation Consultant Fixed Effects in Index Selection



B: Null Distribution of Joint Significance of Firm Characteristics in Compensation-Consultant Preferences

Figure 2. Placebo χ^2 Distribution Figure A plots the placebo distribution of the χ^2 test of joint significance of compensation consultant fixed effects in a regression upon index selection with only compensation consultant, year, and industry fixed effects. The distribution is based on 1,000 draws where within each draw the assignment of compensation consultants to firms is randomly scrambled without replacement. The vertical red line reports the χ^2 statistic under the actual assignment of consultants to firms. Figure B plots the placebo distribution of the χ^2 test of joint significance of firm characteristics in a specification identical to column 3 of Table 7. The distribution is based on 1,000 draws where within each draw the assignment of an index preferring consultant is randomly scrambled without replacement. The vertical red line reports the χ^2 statistic under the actual assignment of index-preferring consultants to firms.

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.

Fiscal Year	Unique # of Firms	Unique # of Grants	Payout Type [Grant-level]			Evaluation Type [Grant-level]			
			Cash	Option	Stock	Abs	Abs/Rel	Rel	Time
<i>Panel A: All CEO Grants</i>									
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
<i>Panel B: CEO Grants with RP Component</i>									
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	-
<i>Panel C: CEO Grants with rTSR Component</i>									
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.

Fiscal Year	Unique # of Firms	Expected Incentive-Plan-Based Compensation				Incentive Ratios	
		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.

Fiscal Year	Unique # of Firms	Unique # of Grants	Relative Performance Benchmark Type					# of Peers
			Specific Peer	S&P500	S&P1500	Other Index	Not Specified	
(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 (N=356)	1.0255*** [0.0258]	0.3272	36.820
Specific Peers (N=201)	1.0052*** [0.0329]	0.8765	40.552
Index (N=155)	1.0520*** [0.0387]	0.1864	31.981

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^2 s 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 (N= 356)	0.483*** [0.012]	0.328*** [0.010]	0.518*** [0.012]	-0.155*** [0.011]	0.035*** [0.008]	36.820*** [1.375]
Specific Peers (N= 201)	0.548*** [0.015]	0.328*** [0.012]	0.560*** [0.015]	-0.220*** [0.014]	0.013 [0.009]	40.552*** [1.898]
Index (N= 155)	0.400*** [0.019]	0.329*** [0.017]	0.464*** [0.019]	-0.071*** [0.013]	0.064*** [0.015]	31.981*** [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 (N= 322)	0.497*** [0.013]	0.496*** [0.012]	0.525*** [0.013]	-0.001 [0.007]	0.028*** [0.008]	37.547*** [1.433]
Specific Peers (N= 190)	0.559*** [0.015]	0.534*** [0.015]	0.568*** [0.015]	-0.025*** [0.008]	0.009 [0.009]	40.837*** [1.921]
Index (N= 132)	0.409*** [0.020]	0.442*** [0.020]	0.464*** [0.021]	0.033*** [0.011]	0.055*** [0.015]	32.811*** [2.079]

Table 6.
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.II. 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)
<u>CEO Characteristics</u>				
<i>CEO Total Pay</i>	0.016*** [0.005]		0.012*** [0.004]	
<i>CEO Expected Pay</i>		-0.006 [0.011]		-0.003 [0.013]
<i>CEO Abnormal Pay</i>		0.020*** [0.005]		0.014*** [0.005]
<i>CEO Tenure</i>	0.007 [0.005]	0.006 [0.005]	0.007 [0.005]	0.007 [0.006]
<i>CEO Age</i>	-0.000 [0.005]	0.000 [0.005]	0.005 [0.006]	0.004 [0.006]
<u>Board and Firm Characteristics</u>				
<i>% Busy Directors</i>	0.402 [0.482]	0.340 [0.486]	0.548 [0.467]	0.420 [0.481]
<i>Board Size</i>	0.033** [0.015]	0.034** [0.016]	0.027* [0.014]	0.029** [0.015]
<i>Director Workload</i>	0.212*** [0.069]	0.218*** [0.072]	0.203*** [0.073]	0.207*** [0.074]
<i>% Age 65+ Directors</i>	0.182 [0.149]	0.213 [0.154]	0.258* [0.149]	0.271* [0.154]
<i>Log Market Cap</i>	-0.072** [0.031]	-0.004 [0.045]	-0.027 [0.031]	0.019 [0.046]
<i>Return Volatility</i>	-0.197 [0.552]	0.110 [0.562]	0.232 [0.559]	0.343 [0.579]
<i>Book-to-Market</i>	-0.141 [0.088]	-0.127 [0.092]	-0.040 [0.092]	-0.028 [0.097]
<u>Industry Characteristics</u>				
<i>Census-based HHI Index</i>	0.309 [0.611]	0.634 [0.715]	-0.168 [0.611]	0.104 [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 7.
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 (except column 3) 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.II. 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. Column 3's dependent variable is an indicator which equals one if the compensation consultant has a preference for index-based benchmark. 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 ($\chi^2=50.81$) in a specification identical to column 1 without any CEO, board, firm, or industry characteristics (but including industry and time fixed effects). Columns 4 and 5 report sub-sample results conditioned on whether the compensation consulting firm has a preference for an index benchmark versus specific peers. The observations in columns 2 and 3 are not identical because 11 observations are perfectly predicted and dropped. The reported p -value refers to joint χ^2 test 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.

	$Pr(Index) = 1$		$Pr(IndexPreferring) = 1$	$Pr(Index) = 1$	
	(1)	(2)	(3)	(4)	(5)
	All Sample	All Sample	All Sample	Subsample: Index Preferring Comp Consultants	Subsample: Specific Preferring Comp Consultants
<u>CEO Characteristics</u>					
<i>CEO Total Pay</i>	0.012*** [0.004]				
<i>CEO Expected Pay</i>		-0.005 [0.012]	0.007 [0.012]	0.001 [0.016]	-0.005 [0.006]
<i>CEO Abnormal Pay</i>		0.014*** [0.005]	0.001 [0.004]	0.015** [0.006]	0.003 [0.002]
<i>CEO Tenure</i>	0.006 [0.005]	0.006 [0.005]	0.002 [0.006]	0.010 [0.008]	0.000 [0.001]
<i>CEO Age</i>	0.006 [0.006]	0.005 [0.006]	-0.003 [0.005]	0.009 [0.008]	-0.000 [0.001]
<u>Board and Firm Characteristics</u>					
<i>% Busy Directors</i>	0.548 [0.465]	0.416 [0.482]	-0.458 [0.531]	0.284 [0.765]	-0.007 [0.097]
<i>Board Size</i>	0.031** [0.013]	0.033** [0.013]	-0.004 [0.015]	0.014 [0.023]	0.009 [0.009]
<i>Director Workload</i>	0.221*** [0.074]	0.237*** [0.076]	-0.094 [0.061]	0.216** [0.104]	0.045 [0.041]
<i>% Age 65+ Directors</i>	0.260* [0.151]	0.263* [0.157]	0.225 [0.147]	0.233 [0.204]	0.072 [0.072]
<i>Log Market Cap</i>	-0.037 [0.031]	0.019 [0.044]	0.037 [0.045]	0.012 [0.062]	0.014 [0.021]
<i>Return Volatility</i>	0.401 [0.530]	0.526 [0.552]	-0.313 [0.550]	1.192 [0.910]	0.057 [0.132]
<i>Book-to-Market</i>	-0.006 [0.087]	0.011 [0.092]	0.018 [0.092]	0.006 [0.142]	0.009 [0.024]
<u>Industry Characteristics</u>					
<i>Census-based HHI Index</i>	-0.245 [0.555]	0.069 [0.593]	-0.071 [0.801]	-0.267 [1.126]	0.058 [0.139]
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Comp Consultant FE	Yes	Yes	No	No	No
p -value of χ^2 (Comp FE)	0.0000	0.0000			
Observations	1,162	1,070	1,081	639	442
Pseudo R^2	0.2634	0.2709	0.0607	0.2173	0.3673
Mean of Dep. Var.	34.1%	34.5%	59.2%	43.2%	20.8%

Table 8.
Index Benchmarks and ROA

Columns 1 to 5 of this table reports OLS, 2SLS and GMM estimates of 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.II. Observations are at the annual firm-benchmark level. Columns 1 and 2 are OLS based where column 1 (2) excludes (includes) compensation consultant fixed effects. Columns 3 and 4 report the first and second stage of the 2SLS estimates using compensation consultant fixed effects as instruments in the first stage. Column 5 uses the GMM estimator to optimally weigh the instrument moments. Column 6 estimates the intent to treat effect of being assigned to an index-preferring consultant as defined in Table 7. All columns include time and industry-fixed effects using the 2-digit Global Industry Classification definitions. 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.

	OLS		2SLS		GMM	OLS
	(1)	(2)	First Stage	Second Stage	(5)	(6)
<i>Index</i>	-0.009** [0.004]	-0.007** [0.004]			-0.027*** [0.010]	
<i>Index-Preferring Consultant</i>						-0.005* [0.003]
<u>CEO Characteristics</u>						
<i>CEO Expected Pay</i>	-0.001 [0.001]	-0.001 [0.001]	-0.003 [0.009]	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.001]
<i>CEO Abnormal Pay</i>	-0.000 [0.000]	-0.000 [0.000]	0.011*** [0.003]	-0.000 [0.000]	0.000 [0.000]	-0.000 [0.000]
<i>CEO Tenure</i>	0.000 [0.000]	0.000 [0.000]	0.006 [0.004]	0.000 [0.000]	0.000 [0.000]	0.000 [0.000]
<i>CEO Age</i>	0.001* [0.000]	0.001* [0.000]	0.003 [0.005]	0.001* [0.000]	0.001* [0.000]	0.001 [0.000]
<u>Board and Firm Characteristics</u>						
<i>% Busy Directors</i>	-0.013 [0.035]	-0.014 [0.035]	0.357 [0.363]	-0.008 [0.034]	-0.001 [0.033]	-0.019 [0.035]
<i>Board Size</i>	-0.002** [0.001]	-0.002** [0.001]	0.024** [0.012]	-0.002** [0.001]	-0.002** [0.001]	-0.003*** [0.001]
<i>Director Workload</i>	-0.006 [0.005]	-0.005 [0.005]	0.198*** [0.067]	-0.003 [0.005]	-0.006 [0.005]	-0.008 [0.005]
<i>% Age 65+ Directors</i>	0.009 [0.011]	0.007 [0.011]	0.227* [0.131]	0.012 [0.011]	0.009 [0.011]	0.008 [0.011]
<i>Log Market Cap</i>	0.011*** [0.004]	0.011*** [0.004]	0.013 [0.035]	0.011*** [0.004]	0.011*** [0.004]	0.011*** [0.004]
<i>Return Volatility</i>	-0.256*** [0.058]	-0.249*** [0.058]	0.574 [0.506]	-0.250*** [0.056]	-0.247*** [0.051]	-0.261*** [0.059]
<i>Book-to-Market</i>	-0.046*** [0.009]	-0.048*** [0.009]	0.003 [0.073]	-0.046*** [0.009]	-0.045*** [0.008]	-0.046*** [0.009]
<u>Industry Characteristics</u>						
<i>Census-based HHI Index</i>	-0.026 [0.035]	-0.031 [0.037]	0.235 [0.544]	-0.024 [0.036]	-0.045 [0.037]	-0.027 [0.036]
Time FE	Yes	Yes		Yes	Yes	Yes
Industry FE	Yes	Yes		Yes	Yes	Yes
Comp Consultant FE	No	Yes		No	No	No
p -value of χ^2 (Comp FE)		0.331				
p -value of J-statistic				0.628		
Observations	1,083	1,083	1,083	1,083	1,083	1,083
Adj R^2	0.346	0.349	0.273	0.331	0.320	0.343

Table 9.

Bias-Adjusted Performance-Effects OLS Estimates based on Degree of Selection Between Observables and Unobservables

This table reports biased-adjusted OLS estimates of index benchmarks on firms' ROA using controls as defined in Table 7. Following Altonji, Elder and Taber (2005) and Oster (2017), δ is the proportionality of selection between observables and unobservables as defined through $[\delta \frac{\sigma_{index,observable}^2}{\sigma_{observables}^2} = \frac{\sigma_{index,unobservables}^2}{\sigma_{unobservables}^2}]$. Zero selection ($\delta = 0$) corresponds to column 2 of Table 8. R_{max}^2 is the theoretical R^2 including both observables and unobservable as controls. Following Oster (2017), the first row assumes that R_{max}^2 equals 1.3X the \tilde{R}^2 of column 2 in Table 8 (≈ 0.45). The second row assumes that R_{max}^2 equals 2X \tilde{R}^2 (≈ 0.70). The third row assumes that $R_{max}^2 = 1$. All columns include compensation consultant, time, and industry-fixed effects using the 2-digit Global Industry Classification definitions. Standard errors are clustered bootstrapped with 1,000 repetitions and reported below the point estimates in brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

	Selection Between Observables and Unobservables (δ)			
	$\delta = 0.5$	$\delta = 1$	$\delta = -0.5$	$\delta = -1$
	(1)	(2)	(3)	(4)
Panel A: $R_{max}^2 = 1.3\tilde{R}^2$				
<i>Index</i>	-0.010*** [0.003]	-0.011*** [0.003]	-0.008*** [0.003]	-0.007*** [0.003]
Panel B: $R_{max}^2 = 2\tilde{R}^2$				
<i>Index</i>	-0.013*** [0.004]	-0.018*** [0.006]	-0.006** [0.003]	-0.004 [0.003]
Panel C: $R_{max}^2 = 1$				
<i>Index</i>	-0.016*** [0.005]	-0.029** [0.013]	-0.005 [0.003]	-0.002 [0.004]
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Comp Consultant FE	Yes	Yes	Yes	Yes
CEO Characteristics	Yes	Yes	Yes	Yes
Board and Firm Characteristics	Yes	Yes	Yes	Yes
Industry Characteristics	Yes	Yes	Yes	Yes
Observations	1,083	1,083	1,083	1,083