HARVARD | BUSINESS | SCHOOL

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

Funding for this research was provided in part by Harvard Business School.

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

September 2019

Abstract

We develop a measurement-error framework for assessing the quality of relative-performance metrics designed to filter out the systematic component of performance, and analyze relative total shareholder return (rTSR)—the predominant metric market participants use to isolate managers' idiosyncratic performance—chosen by boards to evaluate managers. Among firms that explicitly use rTSR in relative performance contracts, 60%—those that choose specific peers—select rTSR metrics that do a remarkable job of filtering out the systematic component of returns in adherence to the informativeness principle. Firms that choose index-based benchmarks retain substantial systematic noise in their rTSR metrics. The choice 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. These findings provide novel evidence on why some executives are evaluated on systematic noise and on compensation consultants' impact on firms.

JEL: G30, J33, M12, M52
Keywords: Relative TSR; measurement error; 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, University of Chicago, 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, Nikhil Lele and Trent Tishkowski of Ernst and Young, Stephen O'Byrne of Shareholder Value Advisors, and Ben Burney of Exequity 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 *et al.*, 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). 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 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

[1]According 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

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 examination of rTSR sheds new light on the longstanding unsettled academic debate—beginning with Antle and Smith (1986)—about the degree to which and the potential reasons why 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 more important (Gibbons and Murphy, 1990; Gong *et al.*, 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.

To examine the first research question, we develop a measurement-error framework for analyzing the extent to which firms' chosen rTSR peers capture the common component of performance. We show that, under fairly general assumptions, it is possible to derive empirical measures that allow for the analysis of the distributional properties of the measurement errors of a particular set of peers. Based on this framework, we derive a necessary and sufficient condition for identifying peers that fully capture the systematic component of performance: no alternative peer groupings would yield lower measurement-error variances. To show that a particular set of chosen peers contain a significant degree of systematic noise, it suffices to identify an alternative benchmark that exhibits a significantly lower measurement-error variance. As considering all possible peer formulations would be intractable, we utilize as an alternative benchmark search-based peers (SBPs) (Lee *et al.*, 2015, 2016), which have been shown to be superior to other state-of-the-art peer identification schemes at explaining variation in firms' stock returns, valuation multiples, and fundamental performance characteristics.

Finding that firms' chosen rTSR peers exhibit significantly greater measurement-error variance than do SBPs suggests that the chosen peers measure the common component of firm performance with a significant degree of error. On the other hand, finding that firms' chosen rTSR peers do not exhibit significantly greater measurement-error variance than do SBPs would be consistent with the chosen peers measuring the common component of firm performance with little error. To be sure, we also derive a necessary condition for capturing systematic performance: peer benchmarks must exhibit a benchmark-return-beta of 1. We use this supplemental test to validate the relative-performance benchmarks chosen by firms.

Based on this evaluation framework, we document the following empirical findings in addressing the first research question. 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 on average, and that remove about 93% of the systematic noise contained in randomly selected peers. This result may be particularly surprising in light of the prevailing view that the selection of tailored peers could facilitate managerial opportunism (Walker, 2016).

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), and although they exhibit significantly lower measurement-error variances compared to randomly chosen peers, on average these index-based benchmarks only remove about 52% of the noise in TSR.

Our second set of empirical analyses 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 consultants 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 *et al.*, 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); or (5) boards' provisioning of incentives on the basis of aspiration (Hayes and Schaefer, 2009; Hemmer, 2015; Francis *et al.*, 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 *et al.*, 1998; Chevalier and Ellison, 1999; Brickley *et al.*, 1999; Fee and Hadlock, 2003; Milbourn, 2003).[2]

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 or turnover, 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). 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. We also contribute to this literature by developing a measurement-error framework for analyzing the quality of relative-performance benchmarks chosen to remove the systematic component of performance. We show that, under fairly general assumptions, it is possible to empirically estimate and compare the variance and mean of measurement errors from a particular set of peers.

[2]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

Our work also contributes to the literature that examines the consequences of the quality of board monitoring (e.g., Core *et al.*, 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 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 *et al.*, 2009; Cadman *et al.*, 2010; Armstrong *et al.*, 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 *et al.*, 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 performance peer selection in executive compensation is explained by opportunism or efficiency.[3] Recent studies suggest that the selection of rTSR performance benchmarks may be opportunistically motivated (Gong *et al.*, 2011; Bizjak *et al.*, 2016). For example, focusing on firms that select specific peers only, Gong *et al.* (2011) finds that while firms' chosen relative-performance peers have higher returns co-movement than random peers, consistent with the informativeness principle, they also

[3]There is also a parallel unsettled empirical literature studying the selection of compensation peer benchmarks (primarily used to assess the manager's outside option). Bizjak *et al.* (2008); Albuquerque *et al.* (2013); Cadman and Carter (2013) finds evidence that the selection of compensation peers reflect CEO talent whereas Faulkender and Yang (2013); Cremers and Grinstein (2013); De Vaan *et al.* (2019) argues in favor of opportunism.

exhibit weaker returns than the base firm, suggestive of rent extraction. Our measurement-error framework suggests that the mean difference between firm and peer performance (i.e., the average level of the relative performance metric) could result from "pure" or innocuous measurement errors. Moreover, our findings also suggest that the selection of index-based benchmarks by boards 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 *et al.*, 2011; Bebchuk *et al.*, 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).

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 examines the measurement-error 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&P 500, most of the S&P Midcap 400, and a small proportion of the S&P Small-Cap 600. Thus, roughly speaking, each annual cross-section encompasses the largest 1,000 firms listed on the U.S. stock market in terms of market capitalization. Our analysis focuses on the sample from 2006 onward; mandatory disclosure of compensation details began in 2006, and coverage of firms is more comprehensive after that year.

For each grant, ISS Incentive Lab collected information on the form of the payout (cash, stock options, or stock units); conditions for payout (tenure [Time], fulfillment of absolute performance criteria [Abs], relative performance criteria [Rel], or a combination of the two [Abs/Rel]); and specific accounting- or stock-based performance metrics associated with performance-based grants. Finally, ISS Incentive Lab collected information on the specific peer firms or indexes selected for purposes of awarding grants based on relative performance.[4]

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.

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.[5] Columns 4 and 5 estimate the allocated expected compensation stemming from meeting RP-based targets and

[4] For example, in 2008, Consolidated Edison selected Ameren, America Electric Power, Centerpoint Energy, Constellation Energy, Dominion Resources, DTE, Duke Energy, Edison International, Entergy, Exelon, FirstEnergy, FPL, NiSource, Pepsco, PG&E, PPL, Progress Energy, Sempra Energy, Southern Company, and Excel Energy as its peers for measuring relative shareholder returns.

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

from meeting rTSR-based targets respectively.[6] Overall, RP-based incentives comprise a significant proportion of the total expected plan-based compensation, with rTSR accounting for the vast majority of RP-based incentives (73% on average as reported in column 6).

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

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&P 500 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&P1500 index, and other indexes (typically industry-based). Columns 4-7 report the percentages of RP grants that use each type of benchmark in a given fiscal year. Column 8 reports the percentage of RP grants whose benchmark cannot be identified. Because each grant can be associated with multiple types of benchmarks, the sum of the values across columns 4-8 can exceed one. Finally, column 9 reports the average number of peer

[6] We 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.

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

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). Specifically, we examine the extent to which benchmarks chosen by boards filter out the systematic component of TSR.

3.1 Theoretical Foundations

We begin by developing a measurement-error framework for understanding how well chosen RP peers capture the common component of firm performance. Our starting point is the factor structure for a given firm's performance,

$$p_t = a + \mathbf{b}'\mathbf{f}_t + \epsilon_t \quad (1)$$

$$= a + c_t + \epsilon_t \quad (2)$$

where p_t refers to a firm performance metric (e.g., TSR), a is a fixed constant, $\mathbf{f}_t$ is a vector of factor returns, and $\mathbf{b}'$ is a vector of factor-return sensitivities, and ϵ_t represents idiosyncratic shocks to firm performance that are uncorrelated with factor returns. That performance can be decomposed into a linear factor structure (Eqn. (1)) is without loss of generality: given a set of factor returns, a unique linear structure is guaranteed by the projection theorem. Furthermore, *any* linear factor structure can be re-expressed as a single "common" component (Eqn. (2)): with arbitrarily many factors, the common component is simply $c_t = \mathbf{b}'\mathbf{f}_t$.

We make a couple of observations about this structure. First, such a linear structure with a perfectly observable common component is standard in theoretical models of incentive compensation design (e.g., Holmström and Milgrom, 1987; Gibbons and Murphy, 1990). Second, this type of linear factor structure is also consistent with the relative performance metrics observed in practice, like rTSR, which are expressed as the difference between firm and benchmark performance.

In practice, boards do not perfectly observe the common component of performance, but they attempt to measure it, through the choice of peers, with some error:

$$\hat{c}_t = c_t + \omega_{b,t}, \tag{3}$$

where the measurement error (ω_b) is assumed to have 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 $\sigma_b^2 = 0$. The linear decomposition of firm performance is also used in standard incentive models (e.g., Holmström and Milgrom, 1987), which also serve to provide economic intuition for how these measurement errors could affect managerial incentives. In these models, 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.

Therefore, the quality of peers chosen for relative performance metrics is determined by their measurement-error variance. Under the linear decomposition of Eqn. (2), we show that it is possible to derive an empirical metric that allows for a *relative* comparison of measurement-error variances between different measurements (i.e., alternative peer formulations) of the common component of performance.

To see this, note that by combining Eqn. (2) and Eqn. (3), a firm's performance relative to its chosen peers becomes a function of the measurement errors:

$$p_t - \hat{c}_t = a + \omega_{b,t} + \epsilon_t. \tag{4}$$

Note here that the measurement error can have *any* statistical structure (i.e., need not be "classical"). However, by the decomposition property, ϵ_t is uncorrelated with the measurement error. Thus, for two peer benchmarks $\hat{c}_1$ and $\hat{c}_2$ for a firm's common component of performance, the variances of the firm's relative performance identify the ordering of measurement-error variances:

$$\begin{aligned} Var(p_t - \hat{c}_{1,t}) &= Var(\omega_{b_1,t}) + Var(\epsilon_t), \\ Var(p_t - \hat{c}_{2,t}) &= Var(\omega_{b_2,t}) + Var(\epsilon_t). \end{aligned} \tag{5}$$

Because the variance in the idiosyncratic component of performance is common to both equations above, a peer group better captures the common component of firm performance if, and only if, it exhibits a lower variance in relative performance:

$$Var(\omega_{b_1,t}) < Var(\omega_{b_2,t}) \quad \Longleftrightarrow \quad Var(p_t - \hat{c}_{1,t}) < Var(p_t - \hat{c}_{2,t}). \tag{6}$$

This framework suggests that a relative performance benchmark c^* that perfectly filters out the systematic component of performance should exhibit the following testable property.

Proposition 1 *No other measurements of the common risk component of performance, such as alternative peer benchmarks, can produce lower measurement-error variance. Equivalently, no other peer formulations should produce a lower variance in relative performance than* $Var(p_t - {c^*}_t)$.

3.2 Discussion of the Measurement-Error Framework

Proposition 1 provides a *necessary* and *sufficient* condition for identifying perfect measurements of the common component of firm performance. It suggests that, to show that a particular chosen peer benchmark $\hat{c}$ contains significant measurement errors, it suffices to identify an alternative benchmark that exhibits a substantially lower $Var(p_t - \hat{c}_t)$.

However, it is significantly more difficult to show that a chosen set of peers contain little to no measurement error, as it would require researchers to argue that *no* alternative peer sets would yield significantly lower $Var(p_t - \hat{c}_t)$. As considering all possible peer formulations would be intractable, our approach is to rely on the literature and consider the peer formulation that has been shown to best explain contemporaneous firm performance. To the extent that firms' chosen relative peers produces similar, if not better, measurement-error variances compared to this benchmark, we would consider the chosen peers to contain little measurement error.

To be sure, it remains possible that relative-performance peers that perform as well as the "state-of-the-art" benchmarks contain a significant amount of measurement errors. One way to validate whether such peers are in fact consistent with the common component of a firm's performance is to check the following *necessary* condition.

Proposition 2 *The performance metric p should exhibit a regression 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)). Thus, a peer benchmark that fails to exhibit a slope of one would imperfectly capture the common risk component of performance. Together, Propositions 1 and 2 provide a framework for testing and validating the extent to which firms' chosen performance benchmarks capture the common component of firm performance.

The above analyses suggest that the distributional properties of the relative performance metric $(p_t - \hat{c}_t)$ are informative of the properties of the chosen benchmark's measurement-error properties. In addition to the variance, the first moment of the relative performance also reflects chosen peers' measurement errors. In particular, Eqn. (4) suggests that the average difference between firm performance and peer performance is driven by the mean of the benchmark's measurement error:

$$\mathbb{E}[p_t - \hat{c}_1] = a + \mathbb{E}[\omega_{b_1,t}]. \tag{7}$$

Although the literature has interpreted a positive mean to be indicative of rent extraction, or a "strategic" form of measurement error (e.g., Gong *et al.*, 2011; Bizjak *et al.*, 2016), our analysis suggests that a large and positive mean could also result from "pure" or innocuous measurement errors. We examine these properties empirically in the next section.

3.3 Measurement Errors of Firms' Chosen Peers

Following Proposition 1, we test the extent to which firms' chosen rTSR peers capture the common component of their TSR by examining the properties of their rTSR variance. In particular, we compare whether firms' chosen peers produce significantly greater measurement-error variances compared to their search-based peer firms (SBPs). 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—as an approximation of the lower bound on measurement errors 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.[8] In addition, we also compare the measurement-error variances of firms' chosen benchmarks to that produced by a set of randomly selected peers (i.e., chosen without any thought), which represent an upper bound on measurement errors.

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.

Table 4, Panel A, reports our estimates of firms' chosen peers' measurement-error variances. In estimating $Var(p_t - \hat{c}_t)$, we use the median of the peer set's returns for firms that select a set of

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

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 returns. Our choice reflects how the selected peers are *used* in RP contracts and thus how they relate to realized firm performance *ex-post*.[9]

We find that, overall, firms' chosen peers produce measurement-error variances that are significantly—at least 12%—higher than that of SBPs.[10] As an alternative benchmark, we also report $Var(p_t - \hat{c}_t)$ for randomly selected firms. In particular, for each firm-benchmark in the sample, we compute $\hat{c}$ based on the median of ten randomly drawn CRSP peers that existed during the base firm's sampling period.[11] Not surprisingly, the results suggest that overall random peers produce significantly greater measurement-error variances than both firms' chosen peers (at least 49% greater) and SBPs (at least 70% greater).

Finally, the last column of Table 4, Panel A, reports a summary performance metric of firms' chosen peers that describes the percentage of systematic noise embedded in a set of randomly selected peers that is eliminated as a result of the boards' peer selection efforts. This metric is computed as

$$\frac{Var(p - \hat{c}_{chosen}) - Var(p - \hat{c}_{chosen})}{Var(p - \hat{c}_{random}) - Var(p - \hat{c}_{sbp})}. \tag{8}$$

Applying Eqn. (5), this ratio simplifies to

$$\frac{\sigma^2_{b,random} - \sigma^2_{b,chosen}}{\sigma^2_{b,random} - \sigma^2_{b,sbp}}. \tag{9}$$

[9]Choosing the ex-post realization allows for potential private information about future co-movement to be incorporated into the board's decision. Ultimately, however, turnover in chosen peers for rTSR benchmarks is uncommon and the results here are not sensitive to using the prior-year stock returns.

[10]Because $Var(p_t - \hat{c}_t)$ identifies measurement-error variances up to a fixed constant, it provides a lower bound on the proportional improvement of an alternative peer set, since $\frac{\sigma^2_{b,chosen}+\sigma^2_{\epsilon}}{\sigma^2_{b,sbp}+\sigma^2_{\epsilon}} > 1 \implies \frac{\sigma^2_{b,chosen}}{\sigma^2_{b,sbp}} > \frac{\sigma^2_{b,chosen}+\sigma^2_{\epsilon}}{\sigma^2_{b,sbp}+\sigma^2_{\epsilon}}$.

[11]The results we report on random peer benchmarks are based on the average across 1,000 random peer draws (with replacement) per firm. In un-tabulated analysis, we also draw random indexes (there are a total of 77 unique indexes in our sample) instead of random sets of peers as well as experimenting with alternative peer set sizes (100 peers) and obtain similar results.

Assuming that the measurement-error variances are bounded above by random peers (i.e., if the board gave no thought to the peer selection problem) and bounded below by SBPs, then Eqn. (9) can be interpreted as the amount of "total" noise (i.e., that would be generated by random peers) that are resolved due to the boards' peer selection efforts. Across all firms, we find that boards' choice of peers resolves about 80% of the total noise.

We also examine how these results vary for the two 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. 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 4, Panel A, rows 2 and 3, reports the measurement-error variances 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 generate greater measurement-error variances than SBPs and lower measurement-error variances than random peers.

In fact, we find that specific peer sets chosen by firms perform well in capturing the common component of performance. These specific peers produce measurement-error variances of similar magnitude to firms' SBPs. Moreover, we estimate that these chosen peers resolve about 93% of the total noise.

On the other hand, we find that index-based benchmarks perform quite poorly. They produce measurement-error variances that are at least 19% greater than SBPs. Moreover, we estimate that these chosen peers resolve only 52% of the total noise. Together, these findings suggest that the overall under-performance of firms' chosen peers relative to SBPs is concentrated in index-based benchmarks.

We note that the measurement errors in the different types of peers are also reflected in the mean of the relative performance metric. Panel B shows that across the whole sample, a firm's average stock returns is statistically no different than that of its SBPs. On the other hand, a firm's average return is significantly larger than that of its chosen peers. However, this positive and significant difference remains substantially smaller—about one-third the magnitude—compared to the average difference between a firm's returns and random peer set's returns.

As before, we examine the mean of the relative-performance metric for index-based benchmarks and specific peers separately. Consistent with the measurement-error variance results above, these findings on the mean of rTSR is driven by index-based benchmarks. The mean returns of specific peers are statistically no different than the mean returns of those firms that choose these benchmarks. However, the mean returns of index-based benchmarks are significantly lower compared to the returns of firms that choose them as rTSR benchmarks. Nevertheless, this positive and significant difference remains substantially smaller relative to that obtained using a random set of peers. Overall, these results are consistent with the idea that the degree of measurement errors in firms' chosen relative-performance benchmarks, even when they are not a result of strategic choices (e.g., attempts to extract rent), can drive differences in the average level of the relative-performance metric.

3.4 Benchmark-Return Beta

The above findings on the measurement-error properties of firms' chosen rTSR benchmarks suggest that firms choosing specific peers do a good job in measuring and filtering out the common component of their TSR. We validate these results by testing Proposition 2. If these chosen benchmarks indeed measure the common component of performance, they should exhibit a benchmark-return beta of 1. We estimate the following time-series returns regression for each firm:

$$R_{it} = \alpha_i + \beta_i R_{p_{it}} + \epsilon_{it} \tag{10}$$

where R_{it} is firm i's monthly cum-dividend returns in period t and $R_{p_{it}}$ is the benchmark peers' returns. The benchmark-return beta for a given firm is the slope coefficient on benchmark peers' returns.

Table 5, Panel A, reports the results from estimating Eqn. (10). Interestingly, we find a cross-sectional average slope coefficient β 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 specific peers and index-based peers.

Index-based peers exhibiting a slope of 1 is not inconsistent with the conclusion of Table 4, that they contain a significant amount of measurement errors. This is because exhibiting a benchmark-return beta of 1 is a *necessary* but not sufficient condition for perfectly measuring the common component of firm performance.[12] Our interpretation of these slopes, instead, is that boards' choices

[12]Under a *classical* measurement-errors structure (e.g., white noise), the slope would attenuate towards 0 as the

of rTSR benchmarks are consistent with the desire to filter out the common component of their firms' TSR.

We also report the constant coefficient from estimating Eqn. (10). Overall, we find a positive constant coefficient which is statistically significant. Moreover, this positive constant is driven by index-based benchmarks. The return-regressions using the sub-sample of specific peers produce a constant coefficient that is statistically no different from 0. These findings are consistent with those in Table 4, Panel B, which shows that that the mean rTSR, in particular among the sub-sample of firms using index-based benchmarks, is significantly different from 0. Thus, these results on the constant are also consistent with index-based peers containing a significant amount of measurement errors.

Overall, the findings of Tables 4 and Table 5 suggest that firms choose rTSR benchmarks in an attempt to filter out the common component of their TSRs. Those firms choosing specific peers do remarkably well in measuring the common component of performance. On the other hand, those firms choosing index-based benchmarks do relatively poorly, leaving substantial room for improvement.

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 ($\sigma_c^2 > \sigma_b^2$).[13] 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

measurement-error variances of the benchmarks increase. However, this is not true under the more realistic scenario of a non-classical measurement-error structure. Thus, whereas having a slope of 1 is both necessary and sufficient for identifying a perfect benchmark under a classical measurement-error structure, it is only necessary, but not sufficient, under a more general measurement-error structure.

[13] 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.

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 measurement-error variances relative to SBPs; however, after the switch to specific peers, their measurement-error variances become statistically no different from that of SBPs.

Finally, we examine the robustness of the measurement-error variance tests to alternative rules for aggregating peer performance, namely the mean and the 75^{th} percentile of peer portfolio returns, and find very similar results. Since these variations do not affect index returns, these robustness tests focus only on firms that use specific RP peers.

Overall, we draw two conclusions from the empirical findings in 4 and 5. First, the majority of firms—those choosing specific peers—select RP benchmarks that perform remarkably well. 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 alternative peers that can serve as better RP benchmarks.[14] In the following

[14]In un-tabulated results, we also examine how results differ by using a third alternative normative peer benchmark—

section, we explore possible explanations for firms' benchmarking choice.

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

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. Finally, we also find that firms' own compensation benchmark peers generate lower (higher) measurement-error variances than index-based benchmarks (specific peer-based benchmarks).

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 the 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*.[15] 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).[16] 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.[17] 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*, following Core *et al.* (2008), into *CEO Expected Pay* and *CEO Abnormal Pay*. 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

[15] Our board characteristics are motivated based on prior literature and conversations with practitioners. For example, Fich and Shivdasani (2006) suggest that "busy" boards or over-tasked board members reflect weak board monitoring quality. Jensen (1993), Yermack (1996), and Cheng (2008) argue that larger board size is suggestive of less effective board monitoring. Masulis *et al.* (2018) argues that older directors display monitoring deficiencies.

[16] Following Ali *et al.* (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`.

[17] Although the specification is probit, we do not have an incidental parameters problem because we have fixed effects at the industry rather than firm level, (Cameron and Miller, 2015).

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 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 *et al.*, 2009;

Murphy and Sandino, 2010; Cai *et al.*, 2016). Consultant are known to exhibit "styles" across various advisory services: for example, Towers Perrin was accused of giving similar advice about workplace diversity to clients across multiple industries (Cai *et al.*, 2016). Our background interviews with compensation consultants and compensation experts suggest that compensation consultants also have "styles" or preferences with respect with the choice of index versus specific relative performance peer benchmarks.[18]

Table 7 begins by estimating the main specifications of Table 6, but includes compensation-consultant-fixed effects (Bertrand and Schoar, 2003).[19] 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 *et al.* (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

[18]We conducted interviews with eight 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.

[19]In 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.

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 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 a 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).[20] 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

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

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 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.[21] 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).[22] Finally, the idea that specialists on the supply side exhibit "styles" that are i) unrelated to their clientele attributes and ii) can affect clients' outcomes is not new. For example, Fracassi *et al.* (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

[21] In our sample, 97% of compensation consultants are retained by the board.

[22] Consistent with this notion, we find that industry fixed effects are significant in column 3 of Table 7.

"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 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).[23] 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 consultants' 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 the 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.

[23] The 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 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.

4.3.1 Opportunism

One possibility is that the selection of indexes reflects the board's deliberate trade-off between informativeness and opportunism. Suggestive of rent extraction, Bizjak *et al.* (2016) reports that selected peer firms experience lower stock returns relative to the focal firm. Similarly, Gong *et al.* (2011), using analyst price forecasts, reaches a similar conclusion. Our analysis suggests that such performance differences between the firm and its chosen peers reflect the first moment of the peers' measurement error for systematic risk. Moreover, this measurement error could be purely innocuous. For example, Table 4, Panel B, shows a relatively large and significant difference between firm returns and the returns of a randomly chosen set of peers.

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. Note first that peer selection for such purposes should also, in theory, be consistent with the informativeness principle (Lazear and Rosen, 1981). As such, tournament incentives would not appear to rationalize firms' choice of indexes.

Moreover, the large number of firms included in the index-based benchmarks we observe (e.g., S&P500, S&P1500, Russell 3000) makes it unlikely that the choice of these benchmarks is characterized by some "optimal tournament" involving hundreds or thousands of heterogeneous firms. In particular, Lazear and Rosen (1981) 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 (Hayes and Schaefer, 2009; Francis *et al.*, 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 *et al.*, 1998; Chevalier and Ellison, 1999; Brickley *et al.*, 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 jam 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 *et al.* (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 show 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.[24] Notably, the χ^2 statistic for the joint

[24]The first stage is estimated using a linear probability model to avoid the forbidden regression problem (Kelejian, 1971; Abrevaya *et al.*, 2010). For convention, we describe the estimator as 2SLS, but we actually use LIML to estimate

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 developed by Olea and Pflueger (2013), which is robust to heteroscedasticity, serial correlation, and clustering in the error term.[25] 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

the system of equations due to its better finite sample properties (Stock and Yogo, 2002).

[25] Homoscedasticity 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 *et al.*, 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^2_{max}, which captures the importance of the omitted variables for explaining firm performance.[26] 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^2_{max} that would be achieved if we included the unobservables to identify the treatment effect. Oster (2017) recommends a R^2_{max} 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^2_{max} that is 200% of the in-sample R^2 ($R^2 \approx 0.70$) and also assuming a theoretical R^2_{max} of 100%.[27] Panel A shows that when $R^2_{max} = 1.3X$, 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 $R^2_{max} = 2X$, 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 $R^2_{max} = 1$, 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

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

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

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 under a measurement-error framework. 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 perceptions that the selection of specific peers could be misconstrued as opportunism (Walker, 2016).

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 to be 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 prevalent 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

Abrevaya, 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. 31

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, 20

— 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, 21

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. 20

—, 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. 21

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. 30, 33, 55

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. 32

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. 5, 30

Bebchuk, L. A., Cremers, K. M. and Peyer, U. C. (2011). The CEO pay slice. *Journal of Financial Economics*, **102** (1), 199–221. 7, 35

Bennett, B., Bettis, J. C., Gopalan, R. and Milbourn, T. (2017). Compensation goals and firm performance. *Journal of Financial Economics*, **124** (2), 307–330. 15

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, 6, 20

— and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics*, **118** (4), 1169–1208. 6, 23

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, 13, 27, 28

—, 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. 5, 30

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, 30

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, 23

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, **50** (2), 317–372. 21

Cheng, S. (2008). Board size and the variability of corporate performance. *Journal of Financial Economics*, **87** (1), 157–176. 21

Chevalier, J. and Ellison, G. (1999). Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, **114** (2), 389–432. 5, 30

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, 22

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. 21, 43

—, 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. 6

— and Larcker, D. F. (2002). Performance consequences of mandatory increases in executive stock ownership. *Journal of Financial Economics*, **64** (3), 317–340. 5, 30

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 Wennberg, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *NBER Working Paper*. 25

De Angelis, D. and Grinstein, Y. (2016). Relative performance evaluation in CEO compensation: A non-agency explanation. *SSRN Working Paper 2432473*. 9, 29

De Vaan, M., Elbers, B. and DiPrete, T. A. (2019). Obscured transparency? compensation benchmarking and the biasing of executive pay. *Management Science*. 6

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, 30

—, — and PIERCE, J. R. (2013). Managers with and without style: Evidence using exogenous variation. *The Review of Financial Studies*, **26** (3), 567–601. 23

FICH, E. M. and SHIVDASANI, A. (2006). Are busy boards effective monitors? *Journal of Finance*, **61** (2), 689–724. 6, 21

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. 25

FRACASSI, C., PETRY, S. and TATE, G. (2016). Does rating analyst subjectivity affect corporate debt pricing? *Journal of Financial Economics*, **120** (3), 514–538. 25

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, 29

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, 21, 22

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. 6, 20

GIBBONS, R. and MURPHY, K. J. (1990). Relative performance evaluation for chief executive officers. *Industrial & Labor Relations Review*, **43** (3), 30S–51S. 2, 11, 20

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, 13, 28

HAYES, R. M. and SCHAEFER, S. (2009). CEO pay and the lake wobegon effect. *Journal of Financial Economics*, **94** (2), 280–290. 4, 29

HEMMER, T. (2015). Optimal dynamic relative performance evaluation. *Rice University Working Paper*. 4, 29

HOLMSTRÖM, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, **10** (1), 74–91. 1, 10, 19, 35

HOLMSTRÖM, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, **66** (1), 169–182. 30

HOLMSTRÖM, B. and MILGROM, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, **55** (2), 303–328. 10, 11

HUGESSEN (2016). Assessing relative tsr for your company: A brief overview. *Hugessen Consulting White Paper*. 2

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, 20

JENSEN, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *the Journal of Finance*, **48** (3), 831–880. 21

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. 21, 43

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. 31

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. 30

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, **89** (5), 841–864. 4, 28, 29

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, 14, 50

—, — and — (2016). The search for peer firms: When do crowds provide wisdom? *Harvard Business School Working Paper*. 3, 14, 20

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*. 16

Masulis, R. W., Wang, C., Xie, F. and Zhang, S. (2018). Directors: Older and wiser, or too old to govern? *SSRN Working Paper*. 21

Mehran, H. (1995). Executive compensation structure, ownership, and firm performance. *Journal of financial economics*, **38** (2), 163–184. 5, 30

Milbourn, T. T. (2003). CEO reputation and stock-based compensation. *Journal of Financial Economics*, **68** (2), 233–262. 5, 30

Molitor, D. (2018). The evolution of physician practice styles: evidence from cardiologist migration. *American Economic Journal: Economic Policy*, **10** (1), 326–56. 25

Morse, A., Nanda, V. and Seru, A. (2011). Are incentive contracts rigged by powerful CEOs? *Journal of Finance*, **66** (5), 1779–1821. 7, 35

Murphy, K. J. (2002). Explaining executive compensation: Managerial power versus the perceived cost of stock options. *University of Chicago Law Review*, **69**, 847. 4, 28

— and Sandino, T. (2010). Executive pay and "independent" compensation consultants. *Journal of Accounting and Economics*, **49** (3), 247–262. 6, 23

Ogden, S. and Watson, R. (2012). Remuneration committees, pay consultants and the determination of executive directors' pay. *British Journal of Management*, **23** (4), 502–517. 25

Olea, J. L. M. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, **31** (3), 358–369. 32

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, pp. 1–18. 4, 30, 33, 34, 55

Pfizer Inc. (2016). *Criteria for the Selection of a Compensation Committee Consultant*. Tech. rep., Pfizer Corporation. 25

Reda, J. F. and Tonello, M. (2015). The conference board CEO and executive compensation practices 2015 edition key findings. *SSRN Working Paper 2702563*. 15

Shin, J.-E. (2016). An evaluation of compensation benchmarking peer groups based on mutual peer-designating behaviors. *SSRN Working Paper 2823592*. 29

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. 25

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, pp. 557–586. 32

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression. 32

Swinford, D. (2015). The limits of using tsr as an incentive measure. *Pearl Meyer Partners White Paper*. 1

Walker, D. I. (2016). The way we pay now: Understanding and evaluating performance-based executive pay. *Journal of Law, Finance, and Accounting*, **1** (2), 395–449. 3, 35

Wu, M. G. H. (2016). Optimal risk trade-off in relative performance evaluation. *SSRN Working Paper 2288042*. 18

Yermack, D. (1996). Higher market valuation of companies with a small board of directors. *Journal of Financial Economics*, **40** (2), 185–211. 21

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

Main Sample Selection	Firm-year Observations	Firm-year-month Observations	Unique Firms
(1) Firms in ISS Incentive Lab data that include CEO grant data between fiscal year 2004 and 2013	12,216		1,668
(2) Less firms without CEO grants based on an RP component	(8,998)		
	3,218		751
(3) Less firms whose relative benchmark cannot be identified	(685)		
	2,533		645
(4) Less firms that do not use stock price as the relevant RP performance measure	(486)		
	2,047		554
(5) Less firms without CIK-GVKEY matches	(226)		
	1,821		487
(6) Merged with monthly return data from CRSP		21,710	
(7) Less observations with missing SBP data		(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

Main Sample Selection	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 and 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 $Thousands) [tdc1]; *CEO Expected Pay* is obtained following Core *et al.* (2008) by regressing the natural logarithm of *CEO Total Pay* on Log($CEOTenure_{i,t}$), Log($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

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

Table A.II.
Continued

Panel B: Correlation Matrix

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

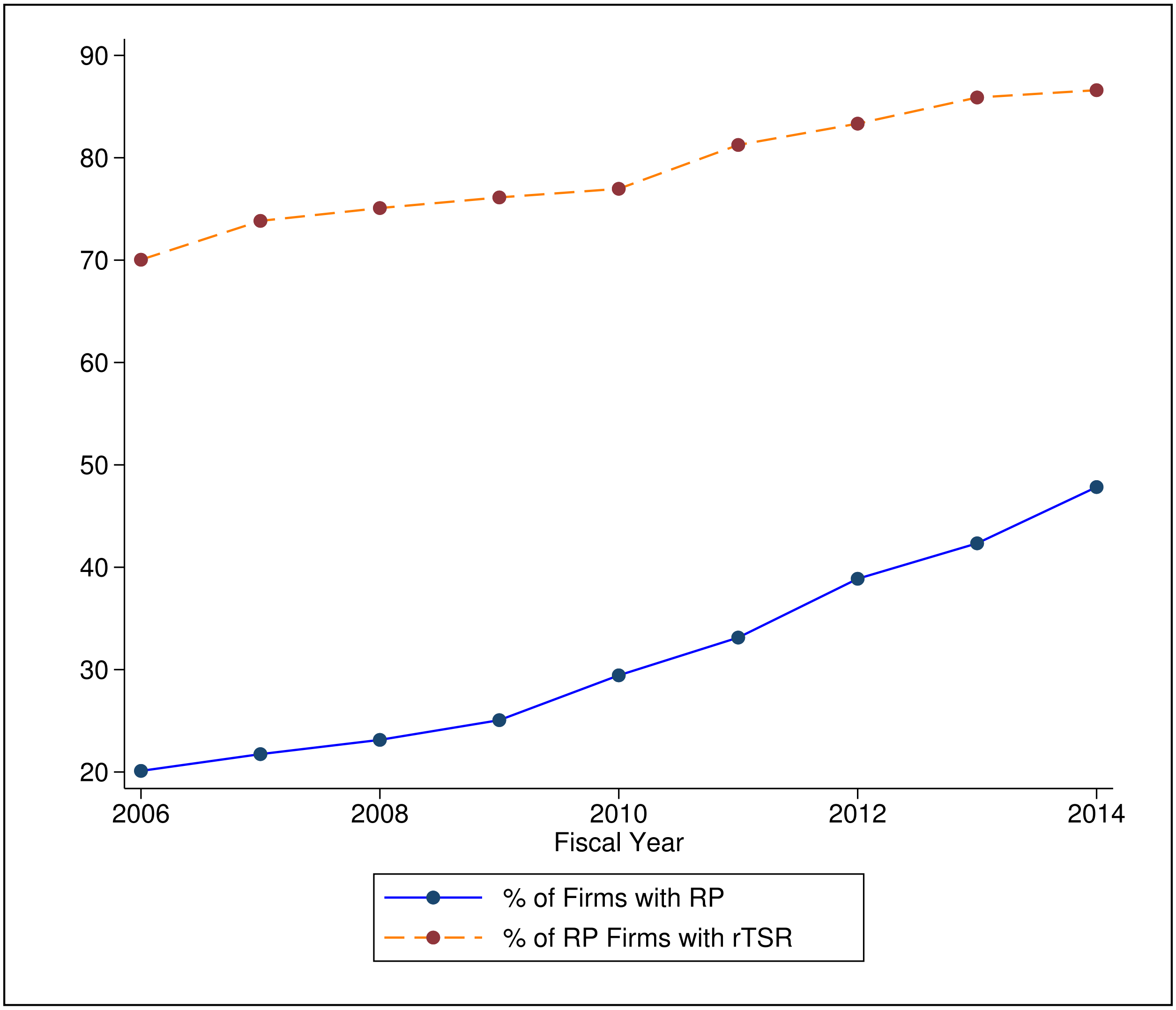
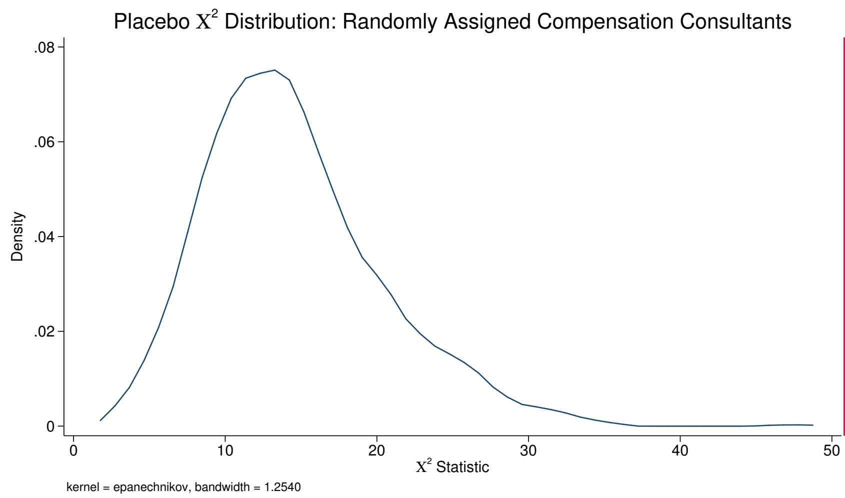
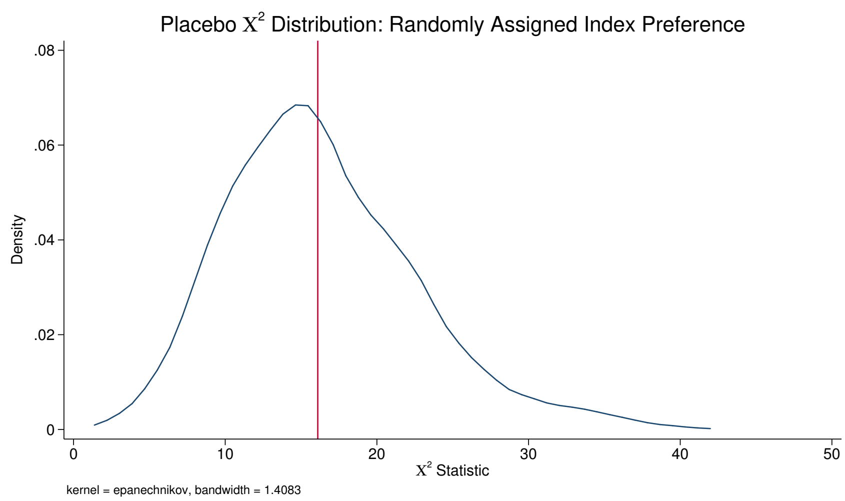


Figure 1. Fraction of Firms Using Relative Performance Contracts 2006-2014 The solid line plots the fraction of firms in the ISS Incentive Labs sample prior to any sample selection restrictions that disclose awarding at least one performance grant based on relative performance (RP) in a given fiscal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



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



Panel 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.

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

Table 2.
Importance of CEO Relative Performance Incentives

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

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

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

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

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

Table 4.
Measurement Errors of Peer Benchmarks

This table reports the distributional properties of the measurement errors for three estimates of the common component of TSR: i) a firm's chosen performance benchmark, ii) a firm's search-based peers ("SBPs" from Lee *et al.*, 2015), and iii) peers randomly selected from CRSP. Panel A of this table reports the cross-sectional average variance of $(p - \hat{c})$ where p is a firm's monthly stock returns and $\hat{c}$ is a measure of the common component of the firm's stock returns. If a firm selects specific peers, c is the median of the chosen peers' monthly returns; if a firm selects index-based benchmarks, c is the monthly index return; for SBPs, c is the portfolio monthly return for the firm's top-10 SBPs; for randomly selected peers, c is the median monthly return from 10 randomly drawn firms from CRSP. Columns 1–3 reports this variance, which estimates the measurement-error variance up to a constant (Eqn. 5), for the three peer sets. Columns 4 and 5 reports the mean difference in measurement-error variances between chosen benchmarks and their SBPs and between firms' chosen benchmarks and random peers, respectively. Column 6 reports the reduction in variance of $(p - \hat{c})$ from selecting SBPs instead of the firm's chosen performance benchmark. Column 7 reports the fraction of total noise embedded in random peers that chosen peers remove. Panel B, columns 1–3, report the cross sectional average of $(p - \hat{c})$, which approximates the mean measurement error up to a constant, for each of the three benchmark types. Columns 4 and 5 report the average difference in the mean measurement error between firms' chosen benchmarks and SBPs and between firms' chosen benchmarks and random peers. 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 (third) row is restricted to the subset that select specific peers (indexes) as benchmarks. Standard errors are reported in brackets and significance levels are indicated by *, **, *** for 10%, 5%, and 1% respectively.

Panel A: Comparing Measurement-Error Variances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\sigma^2_{b,chosen} + \sigma^2_e$	$\sigma^2_{b,sbp} + \sigma^2_e$	$\sigma^2_{b,random} + \sigma^2_e$	$\sigma^2_{b,chosen} - \sigma^2_{b,sbp}$	$\sigma^2_{b,chosen} - \sigma^2_{b,random}$	$\frac{\sigma^2_{b,sbp} - \sigma^2_{b,chosen}}{\sigma^2_{chosen} + \sigma^2_e}$	$\frac{\sigma^2_{b,random} - \sigma^2_{b,chosen}}{\sigma^2_{b,random} - \sigma^2_{b,sbp}}$
All	0.0049***	0.0043***	0.0073***	0.0006***	-0.0024***	-11.93%	80.07%
(N=356)	[0.00039]	[0.00027]	[0.00047]	[0.00045]	[0.00032]		
Specific Peers	0.0045***	0.0043***	0.0078***	0.0002	-0.0033***	-5.27%	93.24%
(N=201)	[0.00039]	[0.00035]	[0.00061]	[0.00018]	[0.00032]		
Index	0.0054***	0.0044***	0.0065***	0.0010***	-0.0011***	-19.19%	52.10%
(N=155)	[0.00072]	[0.00043]	[0.00073]	[0.00045]	[0.00010]		

Panel B: Comparing Measurement-Error Means

	(1)	(2)	(3)	(4)	(5)
	$a + \mathbb{E}[\omega_{chosen}]$	$a + \mathbb{E}[\omega_{sbp}]$	$a + \mathbb{E}[\omega_{random}]$	$\mathbb{E}[\omega_{chosen} - \omega_{sbp}]$	$\mathbb{E}[\omega_{chosen} - \omega_{random}]$
All	0.0022**	0.0005	0.0067***	0.0017***	-0.0045***
(N=356)	[0.0009]	[0.0008]	[0.0010]	[0.0004]	[0.0005]
Specific Peers	0.0015	0.0001	0.0061***	0.0013***	-0.0046***
(N=201)	[0.0011]	[0.0010]	[0.0013]	[0.0005]	[0.0008]
Index	0.0032**	0.0011	0.0075***	0.0021***	-0.0043***
(N=155)	[0.0014]	[0.0011]	[0.0015]	[0.0007]	[0.0004]

Table 5.
Assessing Firms' Chosen RP Benchmarks: Benchmark-Return Betas

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

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

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

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	α	β	p-value $H_0 : \beta = 1$
	(1)	(2)	(3)
All	0.0022*	1.0255***	0.3272
(N=356)	[0.0009]	[0.0258]	
Specific Peers	0.0018	1.0052***	0.8765
(N=201)	[0.0011]	[0.0329]	
Index	0.0026*	1.0520***	0.1864
(N=155)	[0.0013]	[0.0387]	

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

Table 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 (χ^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
CEO Characteristics					
CEO Total Pay	0.012***				
	[0.004]				
CEO Expected Pay		-0.005	0.007	0.001	-0.005
		[0.012]	[0.012]	[0.016]	[0.006]
CEO Abnormal Pay		0.014***	0.001	0.015**	0.003
		[0.005]	[0.004]	[0.006]	[0.002]
CEO Tenure	0.006	0.006	0.002	0.010	0.000
	[0.005]	[0.005]	[0.006]	[0.008]	[0.001]
CEO Age	0.006	0.005	-0.003	0.009	-0.000
	[0.006]	[0.006]	[0.005]	[0.008]	[0.001]
Board and Firm Characteristics					
% Busy Directors	0.548	0.416	-0.458	0.284	-0.007
	[0.465]	[0.482]	[0.531]	[0.765]	[0.097]
Board Size	0.031**	0.033**	-0.004	0.014	0.009
	[0.013]	[0.013]	[0.015]	[0.023]	[0.009]
Director Workload	0.221***	0.237***	-0.094	0.216**	0.045
	[0.074]	[0.076]	[0.061]	[0.104]	[0.041]
% Age 65+ Directors	0.260*	0.263*	0.225	0.233	0.072
	[0.151]	[0.157]	[0.147]	[0.204]	[0.072]
Log Market Cap	-0.037	0.019	0.037	0.012	0.014
	[0.031]	[0.044]	[0.045]	[0.062]	[0.021]
Return Volatility	0.401	0.526	-0.313	1.192	0.057
	[0.530]	[0.552]	[0.550]	[0.910]	[0.132]
Book-to-Market	-0.006	0.011	0.018	0.006	0.009
	[0.087]	[0.092]	[0.092]	[0.142]	[0.024]
Industry Characteristics					
Census-based HHI Index	-0.245	0.069	-0.071	-0.267	0.058
	[0.555]	[0.593]	[0.801]	[1.126]	[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**
			First Stage	*Second Stage*		
	(1)	(2)	(3)	(4)	(5)	(6)
Index	-0.009**	-0.007**		-0.024*	-0.027***	
	[0.004]	[0.004]		[0.012]	[0.010]	
Index-Preferring Consultant						-0.005*
						[0.003]
CEO Characteristics						
CEO Expected Pay	-0.001	-0.001	-0.003	-0.001	-0.001	-0.001
	[0.001]	[0.001]	[0.009]	[0.001]	[0.001]	[0.001]
CEO Abnormal Pay	-0.000	-0.000	0.011***	-0.000	0.000	-0.000
	[0.000]	[0.000]	[0.003]	[0.000]	[0.000]	[0.000]
CEO Tenure	0.000	0.000	0.006	0.000	0.000	0.000
	[0.000]	[0.000]	[0.004]	[0.000]	[0.000]	[0.000]
CEO Age	0.001*	0.001*	0.003	0.001*	0.001*	0.001
	[0.000]	[0.000]	[0.005]	[0.000]	[0.000]	[0.000]
Board and Firm Characteristics						
% Busy Directors	-0.013	-0.014	0.357	-0.008	-0.001	-0.019
	[0.035]	[0.035]	[0.363]	[0.034]	[0.033]	[0.035]
Board Size	-0.002**	-0.002**	0.024**	-0.002**	-0.002**	-0.003***
	[0.001]	[0.001]	[0.012]	[0.001]	[0.001]	[0.001]
Director Workload	-0.006	-0.005	0.198***	-0.003	-0.006	-0.008
	[0.005]	[0.005]	[0.067]	[0.005]	[0.005]	[0.005]
% Age 65+ Directors	0.009	0.007	0.227*	0.012	0.009	0.008
	[0.011]	[0.011]	[0.131]	[0.011]	[0.011]	[0.011]
Log Market Cap	0.011***	0.011***	0.013	0.011***	0.011***	0.011***
	[0.004]	[0.004]	[0.035]	[0.004]	[0.004]	[0.004]
Return Volatility	-0.256***	-0.249***	0.574	-0.250***	-0.247***	-0.261***
	[0.058]	[0.058]	[0.506]	[0.056]	[0.051]	[0.059]
Book-to-Market	-0.046***	-0.048***	0.003	-0.046***	-0.045***	-0.046***
	[0.009]	[0.009]	[0.073]	[0.009]	[0.008]	[0.009]
Industry Characteristics						
Census-based HHI Index	-0.026	-0.031	0.235	-0.024	-0.045	-0.027
	[0.035]	[0.037]	[0.544]	[0.036]	[0.037]	[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 *et al.* (2005) and Oster (2017), δ is the proportionality of selection between observables and unobservables as defined through $[\delta \frac{\sigma_{index,observable}}{\sigma^2_{observables}} = \frac{\sigma_{index,unobservables}}{\sigma^2_{unobservables}}]$. Zero selection ($\delta = 0$) corresponds to column 2 of Table 8. R^2_{max} is the theoretical R^2 including both observables and unobservable as controls. Following Oster (2017), the first row assumes that R^2_{max} equals 1.3X the $\tilde{R}^2$ of column 2 in Table 8 (≈ 0.45). The second row assumes that R^2_{max} equals 2X $\tilde{R}^2$ (≈ 0.70). The third row assumes that $R^2_{max} = 1$. 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: $R^2_{max} = 1.3\tilde{R}^2$				
Index	-0.010***	-0.011***	-0.008***	-0.007***
	[0.003]	[0.003]	[0.003]	[0.003]
Panel B: $R^2_{max} = 2\tilde{R}^2$				
Index	-0.013***	-0.018***	-0.006**	-0.004
	[0.004]	[0.006]	[0.003]	[0.003]
Panel C: $R^2_{max} = 1$				
Index	-0.016***	-0.029**	-0.005	-0.002
	[0.005]	[0.013]	[0.003]	[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