HARVARD | BUSINESS | SCHOOL

rTSR: Properties, Determinants, and Consequences of Benchmark Choice

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

Working Paper 19-112

rTSR: Properties, Determinants, and Consequences of Benchmark Choice

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: Properties, Determinants, and Consequences of Benchmark Choice*

Paul Ma
University of Minnesota
Carlson School of Management

Jee-Eun Shin
University of Toronto

Charles C.Y. Wang
Harvard Business School

May 2021

Abstract

We examine the properties, determinants, and consequences of peer benchmarks chosen by firms to evaluate relative total shareholder returns (rTSR) in executive relative performance contracts. Among firms that explicitly use rTSR, 60% select specific peer firms while 40% select a stock index as benchmarks. Consistent with the intent to filter out the systematic component of TSR, firms' chosen benchmarks exhibit a return-regression slope coefficient of 1 and remove a significant amount of systematic noise in TSR. However, index-based benchmarks are considerably noisier compared to those based on specific peers. Inconsistent with standard contracting models, firms using index-based benchmarks do not exhibit relatively lower pay-to-performance sensitivities, nor do they face lower gains from filtering precision. Instead, index-benchmark use is associated with weaker corporate governance and compensation consultants' preferences, which are uncorrelated with observable firm attributes. The use of index-based benchmarks is also associated with lower ROA, even after controlling for benchmarks' noisiness and companies' governance attributes. Our analyses suggest the salience of peer comparisons as an important attribute of relative performance benchmarks.

JEL: G30, J33, M12, M52
Keywords: Relative TSR; Measurement error; Systematic risk; Compensation consultants; Style effects; Benchmark salience

*The authors can be contacted at paulma@umn.edu, jee-eun.shin@rotman.utoronto.ca, and charles.cy.wang@hbs.edu.
We thank the associate editor and two anonymous referees for their valuable suggestions. We have benefited from the feedback of Ana Albuquerque, Brian Cadman, Mark Lang, David Larcker, Chandra Kanodia, Michael Minnis, Pervin Shroff, and seminar participants at Harvard Business School, National University of Singapore, University of Chicago, and the Minnesota Empirical Accounting Conference. 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 market participants judge companies and their executives. For example, since 2006, the SEC has required all firms to disclose rTSR in their annual reports. 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.[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 the 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 windfalls that can result from general market movements, such as in the case of stock options. Consequently, as noted by Hugessen Consulting, such a metric "satisfies motivation and retention

[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

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, this paper analyzes the properties, determinants, and consequences of firms' rTSR benchmark choice. The proportion of firms with *explicit* relative performance (RP) incentives increased from 20% to 48% in 2014; among these firms, the use of rTSR in these contracts has increased from 70% in 2006 to 87% in 2014 (see Figure 1). Our analyses focus on the sample of firms from 2006 and 2014 that explicitly tie executive compensation to rTSR, for whom we expect the selection of benchmarks to construct the performance measures to be more important (Gibbons and Murphy, 1990; Gong *et al.*, 2011). We find significant variation in how firms determine rTSR benchmarks: about 60% of the rTSR firms choose a customized set of peer firms ("specific peers") while 40% select a stock index. Moreover, rTSR-based incentives (whether using index-based or specific-peer benchmarks) represent economically significant parts of executives' overall compensation contracts. Thus, the properties of the rTSR metric (e.g., the quality of the peers selected) could have a significant impact on managerial incentives.

We begin by examining the properties of rTSR benchmarks. First, we evaluate whether the selection of rTSR benchmarks is consistent with the intent to filter out the systematic component of TSR. To address this question, we derive a necessary condition for capturing systematic performance: peer benchmarks must exhibit a return-regression slope coefficient (benchmark-return-beta) of 1. We find that firms' chosen RP benchmarks, both specific-peer and index-based, exhibit an average benchmark-return-beta of 1.

Second, we examine how well firms' chosen benchmarks perform in filtering the systematic component of TSR. To answer this question, we develop an evaluative framework that, under fairly general assumptions, allows for the analysis of the distributional properties of the measurement errors of a chosen set of peers for the systematic component of a firm's TSR. In particular, the framework facilitates the assessment of rTSR benchmark quality: 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 (MEV). Using search-based peers (SBPs) (Lee *et al.*, 2015, 2016) as a normative benchmark, we find that firms' chosen rTSR benchmarks significantly reduce the systematic noise in TSR, again consistent with the intent to filter out the systematic component of TSR. However, the performance of peer benchmarks varies across benchmark types. Specific peers perform well in capturing the systematic components of firms' TSRs: compared to an rTSR constructed using randomly selected peers, the firm's chosen specific peers remove about 93% of the noise from filtering. Index-based benchmarks do not fare as well in contrast: they only remove about 52% of the noise compared to an rTSR constructed using randomly selected peers.

Our second set of analyses explore the reasons why a significant proportion of firms that tie executive compensation to rTSR choose index-based peers. Under standard contracting models (Holmström, 1979), if the board understands the noisier nature of index-based benchmarks, we would expect compensation contracts based on them to exhibit lower pay-to-performance sensitivities. However, our tests show no differences in the pay-to-performance sensitivities between the firms that utilize specific-peer and index-based benchmarks. Nor do we find evidence that index-using firms face lower gains from filtering precision. We also do not find evidence for other efficiency explanations offered by the literature for why firms might select less precise rTSR benchmarks: managerial efforts being correlated with peer performance (Janakiraman *et al.*, 1992; Aggarwal and Samwick, 1999a) or managerial ability to self-insure against the systematic factor (Garvey and Milbourn, 2003).

Instead, we find that the observed selection of index-based benchmarks is systematically associated with proxies for governance weaknesses, such as having abnormally high executive compensation, a large board, or a heavy director workload. In addition, we find that compensation consultants exhibit systematic preferences towards either index- or specific-peer benchmarks and that these "styles" influence firms' rTRS benchmark choices.[2] Furthermore, firms do not appear to screen compensation consultants based on *these* preferences: While firms hire index- and specific-peer-preferring consultants in roughly equal proportions, these benchmark-type preferences are orthogonal to observable firm characteristics. Together, our analysis suggests that the use of index-based rTSR benchmarks could be partly an unintended and overlooked consequence of compensation consultant selection.

Finally, we examine the performance consequences of the use of index-based rTSR benchmarks. Our identification strategy stems from the observation that compensation consultants' preferences

[2]This finding contrasts with firms' decisions to tie executive compensation to rTSR (as opposed to TSR), for which we do not find any compensation consultant style effects.

for indexes are unrelated to observable firm characteristics but they significantly explain firms' choice of an index. Using compensation consultant preferences for index-based benchmarks as an instrumental variable, we show that both intent-to-treat and two-stage-least squares estimates suggest that the choice of index-based benchmarks leads to lower ROA.

Interestingly, this effect is not fully explained by benchmarks' abilities to filter the systematic components of TSR. We explore an alternative explanation motivated by the greater salience of specific-peer benchmarks, in which peers are explicitly enumerated instead of being grouped in an index and the average number of firms is much lower than index benchmarks. For example, social comparison theory (Festinger, 1954) in the psychology literature argues that workers are intrinsically motivated to compare themselves to others and that effort provision increases because of the need to maintain a positive self-image. A related literature in economics and finance highlights the importance of relative status incentives (Zajonc, 1965; Luttmer, 2005; Clark *et al.*, 2008), which can explain the behavior of corporate executives (e.g., Avery *et al.*, 1998; Dyck *et al.*, 2008; Malmendier and Tate, 2009; Masulis and Mobbs, 2014; Dai *et al.*, 2015; Raff and Siming, 2016; Focke *et al.*, 2017; Chattopadhyay *et al.*, 2020). We argue that the salience of specific-peer benchmarks can help to activate social comparisons or concerns for relative status. Consistent with this idea, prior studies show that: a) relative performance information feedback can motivate workers, even when such information is not explicitly tied to compensation (Falk and Ichino, 2006; Hannan *et al.*, 2008; Tafkov, 2013); and b) salience of information matters for motivation (Hossain and List, 2012; Englmaier *et al.*, 2016).

We find empirical support for the salience hypothesis of specific-peer benchmarks' performance effects. For example, consistent with the idea that an explicit list of a large number of peers is unlikely to make the benchmarks more salient relative to an index that consists of a large number of peers, we find that the choice of index-based benchmarks has no performance effect among the subsample of firms with an above-median number of peers. Instead, our performance effects concentrate in the subsample of firms with a below-median number of peers. Together, our analyses point to an alternative channel, above and beyond their abilities to filter for systematic noise, through which RP benchmarks can impact firm performance, and they suggest the salience of peer comparison as an important (perhaps overlooked) attribute of relative performance benchmarks.

Our work contributes novel evidence on the design, properties, and consequences of rTSR benchmark selection within executive compensation contracts. First, we show that, while there are two predominant ways for selecting relative performance benchmarks (using an off-the-shelf index or a more thoughtful selection of specific peers), our findings overall suggest that a tailored approach to selecting relative performance benchmarks pays off. This approach yields lower filtering noise, produces a greater degree of peer-comparison salience to managers, and leads to higher performance. Thus, we add to the understanding of the value of improving the quality of RP benchmarks and highlight the salience of peer comparisons as an important attribute to consider in benchmarking design.

Second, our findings on the noisiness of index-based benchmarks also adds to the 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). 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.

Third, our results speak to the unresolved debate about *how* compensation consultants influence the executive compensation design process (e.g., Conyon *et al.*, 2009; Cadman *et al.*, 2010; Armstrong *et al.*, 2012; Murphy and Sandino, 2010). Prior literature offers two views on the role of compensation consultants: one view is that compensation consultants have distinct styles that particular firms seek out. Another view is that compensation consultants are substitutes who respond to their economic incentives or the economic circumstances and the incentives of the firm (Cai *et al.*, 2016). Extending the managerial styles literature (Bertrand and Schoar, 2003), our findings suggest a third view: compensation consultants can exhibit distinct styles, orthogonal to the economic circumstances of the firm, that firms do not screen for but are consequential.

Finally, we contribute to the literature by developing, under fairly general assumptions, empirically implementable measures for quantifying whether, and to what extent, a relative performance benchmark captures the systematic component of performance. Our empirical findings on the properties of firms' chosen rTSR benchmarks are broadly consistent with the prior literature, which 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). Our results are novel in providing a quantification of *how well* firms' chosen rTSR benchmarks capture systematic noise within a measurement error framework.

The remainder of the paper proceeds 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 market capitalization. Our analysis focuses on the sample from 2006 onward, since 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. The relative performance criteria include information on the specific peer firms or indexes selected to award grants based on relative performance. Finally, ISS Incentive Lab provides an enumeration of the identities of firms' outside compensation consultants.[3]

[3]For example, in 2008, Consolidated Edison selected as its peers for determining rTSR the following companies: 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.

2.1 Growing Importance 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 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 contained a relative performance component (Abs/Rel or Rel) more than doubled, from 8% in 2006 to 17% in 2014.

Table 1, Panel B and C, suggests that, at the firm level, usage of RP and rTSR incentives have increased dramatically since 2006. The number (Panel B) of firms in our sample that use RP incentives has more than doubled since 2006; similarly, the proportion of firms in our sample with explicit RP incentives increased from 20% in 2006 to 48% in 2014 (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). Together, the summary statistics in Table 1 and Figure 1 illustrate the increasing pervasiveness of explicit RP-based incentives and the prominence of rTSR in such incentive plans.

Our main analyses focus on the sample of firms from 2006 and 2014 that explicitly tie executive compensation to rTSR, for whom the selection of benchmarks to construct the performance measures are expected to be more important (Gibbons and Murphy, 1990; Gong *et al.*, 2011). Table 2, Panel A, reports summary statistics, at the grant level, about firms' benchmark choices for constructing rTSR. We find significant variation in the types of rTSR benchmarks chosen: about 56% of rTSR-based grants select a customized set of peers (column 1) while 40% select index-based benchmarks (column 2). (The remaining grants either do not have information about peer types in the data or use both types.) On average, specific-peer benchmarks consist of 18 peer firms (column 3); in contrast, index-based benchmarks consist of more than 360 peer firms on average (column 4).[4] These differences between the benchmark types, both in terms of their relevance (i.e., ability to capture common noise) and their salience to the executive, could have significant implications on incentives.

[4]ISS Incentive Lab provides the number of peers for specific-peer benchmarking firms. For those firms that use index-based benchmarks, we manually collected information on the number of peers that comprised the index.

We further assess how meaningful rTSR incentives are in executive compensation contracts. Table 2, Panel B, provides back-of-the-envelope estimates of the relative importance of meeting rTSR targets. Column 1 estimates the expected total plan-based compensation when all incentives are earned, including meeting all RP-based targets.[5] Column 2 estimates the allocated expected compensation stemming from meeting rTSR-based targets.[6] Overall, rTSR-based incentives comprise a significant proportion (26% on average) of the total expected plan-based compensation. Additionally, we find that rTSR-based incentives are similarly significant between firms that select specific-peer benchmarks and firms that select index-based benchmarks (columns 3 and 4).[7] Given the economic importance of these rTSR payouts, the significant differences between the benchmark types highlighted in Panel A—for example, in terms of their relevance and their salience to the executive—could have significant implications on their incentive effects.

3 Evaluating Properties of RP Benchmarks

In this section, we evaluate whether the selection of rTSR benchmarks is consistent with the intent to filter out the systematic component of TSR and how well the chosen benchmarks do so. To perform these analyses, we derive two empirical tests through a measurement-error framework.

3.1 Theoretical Foundations

Our starting point is a factor structure for a firm's performance,

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

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

[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.

[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]Our back-of-the-envelope estimates are consistent with the existing evidence on 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.

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 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. Importantly, the structure of Eqn. (1) yields the following *necessary* condition for a performance measure to be consistent with the common component of a firm's performance.

Proposition 1 *The performance metric p should exhibit a regression slope of 1 with respect to the systematic component of performance (c_t).*

In practice, boards wishing to filter out systematic noise cannot perfectly do so, since the common component of performance is unobservable. However, they can estimate it using the contemporaneous performance of a select group of peer firms:

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

where the measurement error (ω_b) is assumed to have finite variance σ_b^2. In this framework, better peers should exhibit lower MEV (lower σ_b^2), and perfect measurement of the common risk component of performance is the special case where $\sigma_b^2 = 0$.

In principle, boards wishing to filter out common noise should select benchmarks that exhibit a performance-regression slope of 1, following Proposition 1. Among the alternative peer groups that could satisfy such a property, the MEVs ultimately determine peer groups' relative quality.

We derive an empirical metric that allows for a *relative* comparison of MEVs between different estimates of the common component of performance (i.e., peer formulations). 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 that while the measurement error can have *any* statistical structure (i.e., need not be "classical"), ϵ_t is always uncorrelated with the measurement error by the decomposition property. Thus, for two peer benchmarks $\hat{c}_1$ and $\hat{c}_2$, the variances of the firm's relative performance identify the ordering of MEVs:

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

Because the $Var(\epsilon_t)$ 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 property.

Proposition 2 *No other measurements of the common risk component of performance, such as alternative peer benchmarks, can produce lower MEV. Equivalently, no other peer formulations should produce a lower variance in relative performance than* $Var(p_t - {c^*}_t)$.

Proposition 2 suggests that, to empirically show 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 $Var(p_t - \hat{c}_t)$. Considering all possible peer formulations would be intractable. Thus, our approach is to rely on the literature and consider the peer formulation that best explains contemporaneous firm performance. If firms' chosen relative peers produce similar, or lower, MEVs than this benchmark, we consider the chosen peers to contain little measurement error.

3.2 Empirical Analyses on Benchmark Properties

Our empirical analysis begins with evaluating whether the observed selection of rTSR benchmarks is consistent with the intent to filter out the systematic component of TSR. Proposition 1 suggests that, if so, we should obtain a benchmark-return beta of 1. We obtain estimates of benchmark-return betas from the following time-series returns regression for each firm:

$$R_{it} = \alpha_i + \beta_{ip} R_{p_{it}} + \epsilon_{it} \tag{7}$$

where R_{it} is firm i's monthly cum-dividend returns in period t, $R_{p_{it}}$ is the benchmark peers' returns, and β_{ip} is the benchmark-return beta. In estimating peer returns, we use the median of the peer set's returns for firms that select a set of specific RP peer firms. Although the choice of the order statistic from the peer-return distribution can be arbitrary, the 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, suppose firm i reports its fiscal-year-end date as December 2000. In that case, we obtain monthly stock-return data for the calendar window January 2001 to December 2001 for it and its performance peers, disclosed in that proxy statement, to calculate returns. Our methodology reflects how the selected peers are used in RP contracts and how they relate to realized firm performance *ex-post*.[8]

Our empirical analysis focuses on those firms that tie their CEOs' performance-based incentives to rTSR, as the quality of the RP metric should be especially important to them. Therefore, we restrict 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) (required for the MEV analysis). In total, our

[8]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.

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. We obtain stock returns data from CRSP monthly files and exclude firms with fewer than ten months of valid monthly returns in total. Detailed construction of our sample is described in Table A.I.

Table 3 reports the results from estimating Eqn. (7). 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 at the 10% level. Moreover, we find that the average slope is close to (and statistically not different from) 1 for both specific peers and index-based peers. We interpret these findings to suggests that firms' rTSR benchmark choices are consistent with the intent to capture systematic noise in TSR.[9]

As discussed above, multiple peer sets could exhibit benchmark-return betas of 1 for a given firm. Thus, we next evaluate the extent to which firms' chosen rTSR benchmarks capture the systematic components of their TSRs by examining their MEV properties. Following Proposition 2, we compare whether firms' chosen peers produce significantly greater MEVs compared to their search-based peer firms (SBPs). We utilize SBPs—representing a firm's 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. Lee *et al.* (2015) and Lee *et al.* (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.[10] We also compare the MEVs 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. Finally, we include the S&P500 as a benchmark and evaluate the degree of MEVs under this standard alternative.

[9] Consistent with this interpretation, those firms that tie incentives to TSR but do *not* use any peers yield a benchmark-return beta of 0. For these firms, the implicit performance benchmark is a fixed constant of 0.

[10] 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. In untabulated results, we also examine how results differ by using another normative peer benchmark—the peers most commonly co-covered by sell-side analysts ("ACPs" of Lee *et al.*, 2016)—and find results very similar to those using SBPs.

Table 4 reports our estimates of MEVs for firms' chosen peers and alternative benchmarks. The first row begins with a validation of the MEV test proposed in the paper. We examine the MEV properties of firms that tie incentives to TSR (i.e., do not use any peers as benchmarks or any relative performance metrics), for whom the implicit "chosen" performance benchmark is a fixed constant of 0. Using a peer set such as SBPs to filter for systematic noise should lead to a significant decline in our MEV measure for these firms. Indeed, by comparing columns 1 and 2 of the first row, we find the MEVs for these firms' chosen benchmarks are significantly higher—at least 86%—than that of SBPs (column 7).[11] These differences are not only economically significant but also statistically significant at the 1% level (column 5).

The second row of Table 4 reports our estimates of MEVs for our primary sample of firms that tie incentives to rTSR. Our first observation is that, in comparing column 1 of the first two rows, the MEVs of TSR firms' chosen benchmarks are significantly higher than those of rTSR firms. Similarly, in untabulated results, we also find that, among the rTSR firms, their benchmark MEVs would have been substantially higher (by at least 67%) if they had used TSR (i.e., chosen a benchmark of 0). Consistent with the results of Table 3, these findings are consistent with these firms selecting rTSR peers to capture systematic noise in TSR.

Table 4 also provides evidence on the extent to which rTSR firms' chosen peers filter out systematic noise in TSR. In row 2, we find that firms' chosen peers produce MEVs that remain significantly—at least 14%—higher than SBPs'. As an alternative benchmark, we also report $Var(p_t - \hat{c}_t)$ for randomly selected firms (column 3) and the S&P500 index (column 4).[12] Not surprisingly, the results suggest that random peers produce significantly greater MEVs than both firms' chosen peers (at least 49% greater) and SBPs (at least 70% greater). Similarly, using the S&P500 index returns as the rTSR benchmark produces significantly greater MEVs than both firms' chosen peers (at least 35% greater) and SBPs (at least 53% greater).[13]

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

[12] For the analysis of random peers, we compute $\hat{c}$ for each firm-benchmark in the sample based on the median of ten randomly drawn CRSP peers that existed during the base firm's sampling period. We report on random peer benchmarks based on the average across 1,000 random peer draws (with replacement) per firm. In untabulated analysis, we also draw random indexes (there are a total of 77 unique indexes in our sample) instead of random sets of peers and experiment with alternative peer set sizes (e.g., 100 peers) and obtain similar results.

[13] The percentage values are derived from the ratio of column 3 (column 4) to column 1 and the ratio of column 3 (column 4) to column 2 for random peers (S&P 500).

In the last column of Table 4, we report a summary performance metric for 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:

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

Applying Eqn. (5), this ratio simplifies to:

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

Assuming that the MEVs are bounded above by random peers (i.e., if the board gave little effort to the peer selection problem) and bounded below by SBPs, then Eqn. (9) can be interpreted as the amount of noise (i.e., that would be generated by random peers) that are resolved due to the boards' peer selection efforts.[14] Alternatively, we interpret $Var(p-\hat{c}_{random})-Var(p-\hat{c}_{sbp}$ as the total gain from filtering precision, and column 8 represents the portion of the total *filtering-precision gain* that is achieved by firms' chosen peers.

Across all firms, we find that boards' choice of peers achieves about 80% of the total filtering-precision gain. Row 2, Table 4, suggests that rTSR firms' chosen peers remove a substantial amount of systematic noise and that firms' choices are on average better than randomly selecting peers. Nevertheless, there remains significant room for improvement, at least on average.

In rows 3 and 4, Table 4, we examine the MEV properties by rTSR benchmark types and show that the underperformance mainly concentrates in index-based benchmarks. We find that firms' chosen specific-peer and index-based benchmarks generate greater MEVs than SBPs and lower MEVs than randomly selected peers. However, there is heterogeneity in how well these benchmark types perform. Specific peers chosen by firms perform well in capturing the common component of TSR: they produce MEVs of similar magnitude to firms' SBPs (column 5), and they achieve about 93% of the total filtering-precision gain. On the other hand, index-based benchmarks perform relatively poorly: they produce relatively higher MEVs, and they achieve only 52% of the total

[14]Relative to the S&P500 index, for example, random peers generate MEVs that are about 10% higher among rTSR firms. Thus, we interpret random peers as generating the upper-bound MEVs from attempting to filter systematic noise.

filtering-precision gain, leaving nearly 50% of the potential gain on the table.

We point out that the finding that index-based peers exhibit a return slope of 1 is not inconsistent with the observation that they contain a significant amount of measurement errors. A benchmark-return beta of 1 is a *necessary*, but not sufficient, condition for perfectly measuring the common component of firm performance.[15] Our interpretation of the slopes in Table 3 is that boards' choices of rTSR benchmarks are consistent with the desire to filter out the common component of their firms' TSR. However, Table 4 suggests that some boards achieve these objectives more effectively than others. Given the prevalence of index-based benchmarks and their relatively poor performance in the presence of available superior peer sets, our results raise the following question: why do some boards select index-based benchmarks?[16]

4 Understanding Benchmark Choice

We now turn to analyze 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.

4.1 Pay-to-Performance Sensitivities

We begin by assessing whether the choice of index-based benchmarks could be consistent with a rational choice under standard contracting models (Holmström, 1979). In particular, if the board understands the noisier nature of index-based benchmarks, we would expect compensation contracts based on them to exhibit lower pay-to-performance sensitivities.

To test this hypothesis, we estimate the pay-to-performance sensitivities for the sample of firms that tie compensation to rTSR. Based on (Gibbons and Murphy, 1990), we estimate the following

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

[16] In fact, in untabulated results, we find that firms' chosen compensation benchmark peers generate lower MEVs than index-based benchmarks. On the other hand, firms choosing specific peer-based benchmarks produce lower MEVs than their compensation benchmark peers.

specification:

$$\begin{aligned}\Delta ln(\mathit{CEO\ Total\ Pay}_{it}) &= \alpha + \beta_1(\mathit{Firm\ Return}_{it}) + \beta_2(\mathit{Chosen\ Peer\ Return}_{it}) \\ &+\beta_3(\mathit{Index\ Benchmark}_{it}) + \beta_4(\mathit{Firm\ Return}_{it} \times \mathit{Index\ Benchmark}_{it}) \\ &+\beta_5(\mathit{Chosen\ Peer\ Return}_{it} \times \mathit{Index\ Benchmark}_{it}) \\ &+\gamma'\mathbf{X_{it}} + \eta_t + \epsilon_{it}, \qquad (10)\end{aligned}$$

where $\Delta ln(\mathit{CEO\ Total\ Pay}_{it})$ is the change in the log of firm i CEO's total compensation (measured by Execucomp's *tdc*1), $\mathit{Firm\ Return}_{it}$ is firm i's stock returns over its fiscal year, $\mathit{Chosen\ Peer\ Return}_{it}$ is the contemporaneous returns of firm i's rTSR peers over its fiscal year, and $\mathit{Index\ Benchmark}_{it}$ indicates whether firm i uses an index-based rTSR benchmark in year t. We include several controls to account for changes in other performance attributes in firm i ($\mathbf{X_{it}}$): change in log of gross ROA, change in log of gross ROE, and change in log total assets. We also include year-fixed effects (η_t) and industry-fixed effects.

In this specification, a firm's selection of peers to capture systematic noise in TSR would imply a positive and significant β_1 and a negative and significant β_2. Further, to the extent boards appreciate the relatively noisy nature of index-based benchmarks and lower the sensitivity of their CEOs' compensation, we would expect to see a positive and significant β_5.

Table 5 reports OLS estimates of Eqn. (10). Columns 1 and 2 show that β_1 is positive and statistically significant at the 1% level while β_2 is negative and statistically significant at the 1% level, with the two coefficients being very similar in magnitudes. This result is unsurprising given that this sample of firms explicitly ties CEO compensation to the firm's stock returns relative to peers. However, our tests show no differences in the pay-to-performance sensitivities between the firms that utilize specific-peer and index-based benchmarks: neither β_4 nor β_5 is statistically significant at the 10% level. Columns 3 and 4 report nearly identical estimates, after controlling for industry-fixed effects (using 2-digit GICS industry groupings).

Overall, these findings suggest that the choice of index benchmarks is inconsistent with the standard contracting model. That is, boards nor their compensation consultants do not appear to account for the noisier nature of index-based benchmarks. Instead, the results point to the possibility that firms' selection of index-based rTSR benchmarks could be an inefficient outcome or result from alternative efficiency considerations. We examine these possibilities below.

4.2 Empirical Drivers of rTSR Benchmark Choice

Several alternative reasons could explain why firms may have chosen less precise RP benchmarks. One possibility is that more precise benchmarks are desirable, but some economic frictions led to the choice of less precise benchmarks. One set of frictions could come from the cost of and the differential gains from precision. For instance, firms that select index-benchmarks may have lower benefits of selecting specific peers, perhaps because a comparable set of peers is more difficult to identify or because they face greater market or idiosyncratic risk in their performance. Another 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 based on systematic risk (Bertrand and Mullainathan, 2001; Garvey and Milbourn, 2006).

Beyond governance, some fundamental economic attributes of the firm may also render less precise RP 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 benchmarks (or the filtration of systematic shocks more generally) less advantageous (Gibbons and Murphy, 1990; Albuquerque, 2013). Additionally, Janakiraman *et al.* (1992) and Aggarwal and Samwick (1999a) suggest that in oligopolistic industries, where managers' efforts are correlated with the performance of its peer benchmarks, precise benchmarks would lead managers to sabotage their industry competitors rather than improve their performance. As a result, it may be optimal to partially reward CEOs for the systematic shock 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 to eliminate market-level volatility from their performance. Another theory, offered by Garvey and Milbourn (2003), is that managers who are more able to self-insure against systematic noise benefit less from more precise benchmarks. If so, the selection of index-based benchmarks may reflect lower benefits from risk-sharing motives.

We also examine the role of compensation consultants. Prior literature suggests that compensation consultants 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). Anecdotally, consultants 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). We conducted interviews with eight compensation consultants and three compensation experts involved in determining CEO compensation packages at their respective corporations. While these interviewees 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 rTSR peer benchmarks. Certain consultants have built capabilities to identify ideal specific-peer benchmarks better; others choose indexes by default.

To explore which of these forces serve to explain firms' choice of index-based benchmarks, we investigate the empirical drivers of index-benchmark selection in Table 6. Our main dependent variable of interest is the indicator *Index Benchmark*. To test whether there are differential benefits to precision, we include *Filtering-Precision Gain* (as in Table 4 columns 6) as an explanatory variable. We also examine a number of explanatory variables relating to CEO, board, firm, and industry characteristics: we include four CEO characteristics—*CEO Expected Pay*, *CEO Abnormal 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*.[17] 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).[18] Finally, we include compensation-consultant-fixed effects.[19] The specifics of variable construction are explained in Table A.II; Panel

[17]We orthongalize *CEO Total Pay* into *CEO Expected Pay* and *CEO Abnormal Pay* following Core *et al.* (2008). *CEO Expected Pay* is the pay that is predicted by a cross-sectional regression model trained on a set of standard economic determinants of executive compensation; and *CEO Abnormal Pay* is the difference between *CEO Total Pay* and the estimated *CEO Expected Pay*. Our board characteristics are motivated from 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.

[18]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`.

[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. Moreover, in the sample, 4.7% of firm-year observations do not have a compensation consultant. There is no tendency between having a compensation consultant and the choice of an index or specific peer group benchmark (Fisher's exact p-value = 0.45). For simplicity, we group these observations as an additional consultant fixed effect, but excluding them does not qualitatively change our inferences in Tables 6, 7, and 8.

A reports the pooled summary statistics for each variable, and Panel B reports and compares the mean value in each variable between index- and specific-peer-choosing firms.

Column 1, Table 6, reports the marginal effects from a probit regression of the index selection indicator *Index Benchmark* on these characteristics; year- and industry-fixed effects are also included. We find that, all else equal, firms with higher *CEO Abnormal Pay*, larger *Board Size*, greater *Director Workload*, and higher % *Age 65+ Directors* are associated with a higher likelihood of index selection. We interpret these results as suggesting that governance weaknesses are associated with a greater likelihood of index selection.

On the other hand, we do not find evidence that the choice of indexes with associated with lower gains from filtering precision: while the coefficient on *Filtering-Precision Gain* is negative, it is not statistically significant at the 10% level. Similarly, we do not find evidence that greater performance volatility, which could be captured by having higher *Return Volatility* or *Book-to-Market*, result in greater likelihood of index selection. Neither of these variables is statistically significant at the 10% level. Nor do we find support for the hypothesis that the choice of index-based benchmarks is driven by oligopolistic industries or firms with outsized market power, which could be captured by having higher *Log Market Cap* or *Census-based HHI Index*. Neither of these variables is statistically significant at the 10% level. Finally, we do not find support for the hypothesis that the choice of an index is driven by managers who have greater abilities to self-insure. The coefficient on *CEO Age*, a common proxy for wealth and the ability to self-insure (Garvey and Milbourn, 2003), is not statistically significant at the 10% level.

Our finding that attributes related to governance weaknesses are associated with index selection is consistent with inadequate monitoring: boards exert insufficient effort to design and optimize compensation contracts. However, an alternative explanation is that the selection of indexes reflects the boards' deliberate trade-off between informativeness and opportunism. Suggestive of rent extraction, Bizjak *et al.* (2016) reports that selected peer firms experience lower stock returns than the focal firm. Similarly, Gong *et al.* (2011), using analyst price forecasts, reaches a similar conclusion. Our empirical results appear consistent with Bizjak *et al.* (2016): the constant term (α) in Table 3 suggests that, on average, firms' TSR relative to their chosen peers are positive and statistically significant, particularly for firms that select index-based benchmarks. (Since the slopes are equal to 1, the constant term can be interpreted as the difference between the base firm's TSR and its peer group's TSR.)

However, we offer two explanations for why we hesitate to attribute the association between index selection and governance weaknesses to opportunism. First, innocuously choosing noisier peers could result in more positive rTSR. Consistent with this idea, in untabulated results, we find that selecting random rTSR peers leads to positive and significant average rTSR values, which are larger in magnitude than the mean rTSR values using firms' chosen peers. Second, even if we interpret a significant constant term in Table 3 to be consistent with opportunism, we do not find the constant term to be significantly different between specific peers and indexes (p-value = 0.58). That is, index-choosing firms do not exhibit a significantly greater degree of opportunism. In our view, a more likely alternative explanation is 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.

Another key result in column 1, Table 6, is that compensation consultants play an important role in explaining chosen benchmark types. Relative to a probit specification without them, the inclusion of compensation-consultant-fixed effects increases the pseudo R^2 of the regression specification by over 40% proportionally. The χ^2 test in column 1 shows 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. As an alternative, we also assess the joint significance of compensation consultants using permutation tests, which Fee *et al.* (2013) argues is a more robust approach. In particular, we simulate a placebo distribution of the χ^2 statistic by randomly scrambling the assignment of compensation consultants (without replacement) to firms each draw. After each draw, we estimate a regression of firms' index selection choice on year, industry, and compensation-consultant-fixed effects, then obtaining the χ^2 statistic from a test of the joint significance of the resultant compensation-consultant-fixed effects. We perform this procedure 1,000 times and plot the resulting null distribution along with the actual test statistic in Figure 2, Panel A. The figure shows that the test statistic, based on the actual assignment of compensation consultants to firms, is entirely outside of the simulated null distribution. Overall, these empirical results corroborate the qualitative evidence from our interviews, which point to different compensation consultant "styles" for rTSR benchmark types.

As a means of comparison, in column 2, we examine the role of compensation consultants in the choice to link compensation to rTSR instead of TSR. Thus, our analysis sample adds to the sample reported in column 1 those firms that link their compensation to TSR (but not rTSR or any other type of relative performance metric). In general, the regression results suggest that this selection decision is quite different in nature compared to the decision among rTSR firms to choose index-based benchmarks: different variables load or variables load in different directions. Of particular interest is the finding that compensation-consultant-fixed effects are not significant in explaining the choice to tie compensation to rTSR: the χ^2 test in column 2 shows that these fixed effects are jointly insignificant (at the 10% level). We infer that the choice of rTSR versus TSR is likely a more salient decision, for which firms' particular characteristics are likely to matter more. On the other hand, the benchmark choice for specifying a firm's rTSR is likely to be a more subtle, albeit important, decision, for which compensation consultant styles are more likely to manifest.

Finally, we examine whether firms screen compensation consultants based on their index preference. We re-estimate consultant-fixed effects without any of the covariates in column 1 except for industry- and year-fixed effects; the χ^2 statistic testing the joint significance of consultant effects remains significant at the 1% level. We then partition the sample by the index preference of 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).

Column 3, Table 6, reports estimates of a probit regression for the probability of a firm using an index-preferring compensation consultant. We use the same set of covariates as column 1 but exclude compensation-consultant-fixed effects. Remarkably, the results suggest that *none* of the covariates are significantly associated with the compensation consultant's default tendencies towards index or specific peer benchmarks. Moreover, in a χ^2 test for the joint significance of CEO, board, firm, and industry characteristics for explaining a firm's choice of an index-preferring consultant, we fail to reject the null that all these covariates have zero slopes. 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 and the actual test statistic, based on the actual assignment of compensation consultants to firms, in Figure 2 Panel B. The figure shows that test statistic 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 are unrelated to firms' characteristics.

Our empirical evidence supports the view that boards, in hiring compensation consultants, do not appear to take into consideration the consultants' preferences for RP 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. For example, Ogden and Watson (2012) suggests that the 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). Consistent with this notion, we find that industry-fixed effects are jointly significant in column 1 of Table 6.

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 evidence in the health economics literature, which suggests that there are substantial differences in physicians' diagnostic testing practices that are unrelated to patient characteristics but impact patients' experiences or outcomes (Song *et al.*, 2010; Finkelstein *et al.*, 2016; Molitor, 2018; Cutler *et al.*, 2019). The findings in this section suggest that compensation consultants' index preferences are subtle to firms but important in explaining the choices of rTSR benchmark types, which can differ significantly in their quality. In the following section, we examine whether the choice of an index-benchmark *due to* compensation consultant preferences impacts firm performance.

5 Index Benchmark Selection and Firm Performance

Our final analysis explores the potential firm performance consequences of evaluating managers based on index-based benchmarks. Given our finding (in Table 5) that pay-to-performance sensitivities between index-based and specific-peer benchmarking firms are similar, our null hypothesis is that the choice of an index-based benchmark should not lead to differences in firm performance.

Our empirical causal identification strategy leverages the observations (from Table 6) that: compensation consultants exhibit preferences for benchmark types; these preferences significantly explain firms' choice of index-based benchmarks; and the assignment of these preferences appears to be as good as random, as they are not associated with firm-level attributes. These observations are consistent with the possibility that, conditional on firm and CEO characteristics, consultants' index preferences could influence firm performance *only through* boards' choice of an index-based benchmark. Based on this intuition, we employ consultants' preferences for indexes as an instrumental variable in identifying the causal effect of index-based benchmarks on firm performance.

Table 7 analyzes the effect of index selection on a firm's return on assets (ROA). We begin with a linear regression of ROA on an indicator variable for the assignment to an index-preferring compensation consultant (*Index Preferring*). We control for the same CEO, board, firm, and industry characteristics as in Table 6; we also include year- and industry-fixed effects. This regression, reported in column 1, shows that having an index-preferring consultant is associated with 60 basis point lower ROA, which is statistically significant at the 10% level. Assuming *Index Preferring* is a valid instrument, this is an "intent-to-treat" (ITT) estimate that represents a lower bound on the average performance effect of index selection.

Column 4, Table 7, reports the estimates from a two-stage-least-squares (2SLS) specification, in which we instrument for firms' actual choice of benchmark types (*Index Benchmark*) using *Index Preferring*. The instrument is strong—the first-stage partial F-statistic is over 28. Consistent with the ITT estimates, the 2SLS estimate suggests that the choice of an index is associated with 2.6 percentage point lower ROA, which is statistically significant at the 10% level.

As expected, the 2SLS estimates, at about half of a standard deviation in ROA, are larger than the ITT estimates. 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 in which board monitoring is a weak control on executive actions and where other control mechanisms—such as explicit incentives—are relatively more important. In this way, the average complier treatment effect could be interpreted as an upper bound of the average performance effect of index selection.

5.1 Effect Mechanisms

Together, the ITT and 2SLS estimates reported in columns 1 and 4, Table 7, are inconsistent with the null hypothesis of no performance effects from index selection. We thus examine alternative explanations for why the choice of index-based rTSR benchmarks could impact firm performance.

One possibility is that there *is* a pay-to-performance sensitivity difference in reality, but our empirical test in Table 5 is not sufficiently powerful. In particular, Table 5 report lower pay-to-performance sensitivity point estimates for firms with index-based benchmarks (i.e., the point estimates of the coefficients on *Firm Return* × *Index Benchmark* and *Chosen Peer Return* × *Index Benchmark* are negative and positive), albeit not statistically significantly so. Under this explanation, the lower performance among index-choosing firms would result from managers' responses to the lower pay-to-performance sensitivities they are perceived to face.

Because this explanation ultimately relates to benchmarks' abilities to capture systematic noise in TSR, to test such a possibility, we leverage the empirical estimates derived in Section 3 and analyzed in Table 4. Because we can only measure the MEV of a benchmark up to a constant, to capture a benchmark's ability to measure systematic noise we rely on its MEV in excess of SBPs, $\sigma^2_{b,chosen} - \sigma^2_{b,sbp}$ (e.g., column 5 of Table 4), which we can precisely identify for each firm. We denote this measure *Benchmark Noise* and use it as a control variable in our analyses of performance consequences of benchmark types. Note that if one assumes SBPs have 0 MEVs, *Benchmark Noise* measure would reflect the MEV of a given benchmark.

Table 7, columns 2 and 5, report the ITT and 2SLS estimates after controlling for *Benchmark Noise*.[20] In both cases, we find that the coefficients on *Index Preferring* and *Index Benchmark* are very similar, both in terms of economic magnitude and statistical significance, compared to the initial estimates. The coefficients on *Benchmark Noise* are negative, consistent with noisier

[20]We do not include *Filtering-Precision Gain* in the tests of Table 7 because it is highly correlated with *Benchmark Noise* and because there is no theoretical reason for doing so. Nevertheless, our results and inferences remain nearly identical if we included *Filtering-Precision Gain* in both tables.

benchmarks being associated with lower performance; however, they are not statistically significant at the 10% level. These results suggest that the ROA effect of index selection is not fully accounted for by explanations that relate to benchmarks' abilities to capture systematic noise in TSR.

An alternative explanation is that, all else equal, the salience of the peer benchmark information could have an incentive impact on executives. For example, social comparison theory (Festinger, 1954) in the psychology literature argues that workers are intrinsically motivated to compare themselves to others and that effort provision increases because of the need to maintain a positive self-image. A related literature in economics and finance suggest the importance of relative status incentives (Zajonc, 1965; Luttmer, 2005; Clark *et al.*, 2008), which can explain the behavior of corporate executives (e.g., Avery *et al.*, 1998; Dyck *et al.*, 2008; Malmendier and Tate, 2009; Masulis and Mobbs, 2014; Dai *et al.*, 2015; Raff and Siming, 2016; Focke *et al.*, 2017; Chattopadhyay *et al.*, 2020). We argue that specific-peer-based benchmarks are more salient—peers are explicitly enumerated instead of being grouped in an index and the average number of peers is about 5% of index benchmarks—and this salience helps to activate social comparisons or concerns for relative status. Consistent with this idea, prior literature shows that the disclosure of relative performance information *per se* can motivate workers' effort provision (Falk and Ichino, 2006; Hannan *et al.*, 2008; Tafkov, 2013; Blanes i Vidal and Nossol, 2011).

To test the salience hypothesis of specific-peer-based benchmarks' performance effects, we note that specific peers are more salient in two ways. First, specific peers are more salient because they contain, on average, significantly fewer peers (see, e.g., columns 3 and 4 of Panel A in Table 2). Second, even controlling for the number of peers, indexes are less salient because they mask the identities of the constituent firms, unlike the alternative that explicitly identifies each peer firm. Based on this intuition, we provide two tests for the salience explanation. First, we control for the number of peers in firms' chosen benchmarks. Under the salience explanation, we expect the significant coefficients on *Index Preferring* and *Index Benchmark* to weaken. Indeed, columns 3 and 6 of Table 7 show that, after controlling for the number of peers, both coefficients are no longer statistically significant at the 10% level.

Table 8 reports the results of a second test for the salience explanation. Here, we split the sample into those firms with above-median (Above-Median # Peers) versus below-median number of peers (Below-Median # Peers), and re-estimate the ITT and 2SLS specifications in columns 3 and 6 of Table 7. If indexes are less salient because they mask the identities of the constituent firms, we expect such an effect to be less important among those firms with a relatively large set of peers. Our intuition is that an explicit list of a large number of peers is unlikely to make them more salient relative to an index that consists of a large number of peers. Thus, the salience explanation predicts that the ROA effect of indexes should be concentrated in the Below-Median # Peers sample. Consistent with our expectations, Table 8 shows that the coefficients on *Index Preferring* and *Index Benchmark* are statistically significant only in the Below-Median # Peers sample. Analogous estimates for the Above-Median # Peers sample are statistically insignificant, and their point estimates are substantially attenuated. Together, the results of Tables 7 and 8 point to an alternative channel, above and beyond their abilities to filter for systematic noise, through which RP benchmarks can impact firm performance.

5.2 Robustness

Our empirical analyses of the performance consequences of index selection rely on the assumption that compensation consultants' index preference satisfies the exclusion restriction that validates their use as instrumental variables. Although our finding (in column 3 of Table 6 and Panel B of Figure 2) that these preferences are uncorrelated with observable firm characteristics is consistent with such an assumption, these index preferences may be correlated with unobservable firm characteristics that are important in explaining their ROAs.

To evaluate the extent to which such concerns could confound our performance analyses, we implement the framework proposed by Altonji *et al.* (2005) and Oster (2017), which facilitates an assessment of the extent to which omitted variables could influence our estimates. As these methods apply to OLS, our robustness test focuses on the role of unobservables in the ITT estimates.

Specifically, 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 for bounding the magnitudes of potential omitted variable bias in OLS estimates. As applied to our research setting, Oster (2017) suggests that the sensitivity of our ITT estimate of the performance-effects of index selection depends on: i) the degree to which omitted variables are correlated with having an index-preferring consultant, and ii) the extent to which the omitted variables contribute to explaining firm performance.

Table 9 reports the results from the Oster (2017) approach in estimating bias-adjusted performance effects of index selection using the specification of column 1 in Table 7. We provide a range of bias-adjusted ITT 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.[21] 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). In addition, we consider bias-adjusted ITT 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%.[22]

Table 9, 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 to -90 basis points. These results suggest that the baseline ITT estimates (e.g., Table 7 column 1) 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 again remain significant and range from -80 to -170 basis points, suggesting that the ITT estimates remain robust even when the inclusion of omitted variables can double the amount of variation in firm performance explained. Finally, Panel C reports estimates when $R^2_{max} = 1$, 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 all but one case (when $\delta = 2$). Overall, these robustness tests re-assure our inference that the selection of index-based benchmarks (due to compensation consultants' preferences) leads to lower

[21] 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.

[22] 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 $R^2_{max} = 1$.

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. Central to the design of this metric is benchmark selection.

We analyze the properties, determinants, and consequences of firms' rTSR benchmark choices. In general, firms' rTSR benchmark choices are consistent with the desire to filter for the systematic noise in TSR. However, there is substantial variation in the type of rTSR benchmarks used by firms, and these differences are not innocuous. A substantial percentage of rTSR firms choose index-based benchmarks, which are significantly noisier than specific-peer benchmarks, and these choices lead to worse firm performance. Our analyses point to the salience of peer comparisons as an important (and perhaps overlooked) attribute of benchmarking. They also highlight the role of compensation consultants' preferences in selecting index benchmarks. Although beyond the scope of our paper, we believe a fruitful venue for future research is to understand the dynamics of the interactions between boards and compensation consultants and how they operate to determine the attributes of executive compensation contracts.

References

Aggarwal, R. K. and Samwick, A. A. (1999a). Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence. *Journal of Finance*, **54** (6), 1999–2043. 3, 5, 17

— and — (1999b). The other side of the trade-off: The impact of risk on executive compensation. *Journal of Political Economy*, **107** (1), 65–105. 18

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. 17

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. 18

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. 26, 46

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. 24

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. 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. 5

Avery, C., Chevalier, J. A. and Schaefer, S. (1998). Why do managers undertake acquisitions? An analysis of internal and external rewards for acquisitiveness. *Journal of Law, Economics, & Organization*, **14** (1), 24–43. 4, 25

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

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, 17

— and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics*, **118** (4), 1169–1208. 5

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

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*. 19

Blanes i Vidal, J. and Nossol, M. (2011). Tournaments Without Prizes: Evidence from Personnel Records. *Management Science*, **57** (10), 1721–1736. 25

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

Cadman, B., Carter, M. E. and Hillegeist, S. (2010). The incentives of compensation consultants and ceo pay. *Journal of Accounting and Economics*, **49** (3), 263–280. 5

Cai, C., Kini, O. and Williams, R. (2016). Do compensation consultants have distinct styles? *SSRN Working Paper 2724072*. 5, 18

Chattopadhyay, A., Shaffer, M. D. and Wang, C. C. (2020). Governance through shame and aspiration: Index creation and corporate behavior. *Journal of Financial Economics*, **135** (3), 704–724. 4, 25

Cheng, S. (2008). Board size and the variability of corporate performance. *Journal of Financial Economics*, **87** (1), 157–176. 18

Clark, A. E., Frijters, P. and Shields, M. A. (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, **46** (1), 95–144. 4, 25

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. 5, 17

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. 18, 34

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*. 22

Dai, L., Parwada, J. T. and Zhang, B. (2015). The governance effect of the media's news dissemination role: Evidence from insider trading. *Journal of Accounting Research*, **53** (2), 331–366. 4, 25

De Angelis, D. and Grinstein, Y. (2016). Relative performance evaluation in CEO compensation: A non-agency explanation. *SSRN Working Paper 2432473*. 8

Dyck, A., Volchkova, N. and Zingales, L. (2008). The corporate governance role of the media: Evidence from Russia. *Journal of Finance*, **63** (3), 1093–1135. 4, 25

Englmaier, F., Roider, A. and Sunde, U. (2016). The Role of Communication of Performance Schemes: Evidence from a Field Experiment. *Management Science*, **63** (12), 4061–4080. 4

Falk, A. and Ichino, A. (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics*, **24** (1), 39–57. 4, 25

Fee, C. E., Hadlock, C. J. and Pierce, J. R. (2013). Managers with and without style: Evidence using exogenous variation. *The Review of Financial Studies*, **26** (3), 567–601. 20

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, **7** (2), 117–140. 4, 25

Fich, E. M. and Shivdasani, A. (2006). Are busy boards effective monitors? *Journal of Finance*, **61** (2), 689–724. 18

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. 22

Focke, F., Maug, E. and Niessen-Ruenzi, A. (2017). The impact of firm prestige on executive compensation. *Journal of Financial Economics*, **123** (2), 313–336. 4, 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. 22

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. 3, 17, 19

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. 17

Gibbons, R. and Murphy, K. J. (1990). Relative performance evaluation for chief executive officers. *Industrial & Labor Relations Review*, **43** (3), 30S–51S. 2, 7, 9, 15, 17

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, 7, 19

Hannan, R. L., Krishnan, R. and Newman, A. H. (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review*, **83** (4), 893–913. 4, 25

Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, **10** (1), 74–91. 1, 3, 15

Holmström, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, **55** (2), 303–328. 9

Hossain, T. and List, J. A. (2012). The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. *Management Science*, **58** (12), 2151–2167. 4

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. 3, 17

Jensen, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *the Journal of Finance*, **48** (3), 831–880. 18

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. 18, 34

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

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. 2, 11, 12, 41

—, — and — (2016). The search for peer firms: When do crowds provide wisdom? *Harvard Business School Working Paper*. 2, 12

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*. 5

Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being*. *The Quarterly Journal of Economics*, **120** (3), 963–1002. 4, 25

Malmendier, U. and Tate, G. (2009). Superstar CEOs. *Quarterly Journal of Economics*, **124** (4), 1593–1638. 4, 25

Masulis, R. W. and Mobbs, S. (2014). Independent director incentives: Where do talented directors spend their limited time and energy? *Journal of Financial Economics*, **111** (2), 406–429. 4, 25

—, Wang, C., Xie, F. and Zhang, S. (2018). Directors: Older and wiser, or too old to govern? *SSRN Working Paper*. 18

Molitor, D. (2018). The evolution of physician practice styles: evidence from cardiologist migration. *American Economic Journal: Economic Policy*, **10** (1), 326–56. 22

Murphy, K. J. (2002). Explaining executive compensation: Managerial power versus the perceived cost of stock options. *University of Chicago Law Review*, **69**, 847. 20

— and Sandino, T. (2010). Executive pay and "independent" compensation consultants. *Journal of Accounting and Economics*, **49** (3), 247–262. 5, 18

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. 22

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, pp. 1–18. 26, 27, 46

Pfizer Inc. (2016). *Criteria for the Selection of a Compensation Committee Consultant*. Tech. rep., Pfizer Corporation. 22

Raff, K. and Siming, L. (2016). Knighthoods, damehoods, and CEO behaviour. *Journal of Corporate Finance*, p. forthcoming. 4, 25

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

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. 22

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

Tafkov, I. D. (2013). Private and Public Relative Performance Information under Different Compensation Contracts. *Accounting Review*, **88** (1), 327–350. 4, 25

Yermack, D. (1996). Higher market valuation of companies with a small board of directors. *Journal of Financial Economics*, **40** (2), 185–211. 18

Zajonc, R. B. (1965). Social Facilitation. *Science*, **149** (3681), 269–274. 4, 25

Table A.I.
Sample Selection

Panel A of this table reports the selection criterion used to generate the final samples used in Tables 3 and 4. Panel B reports the selection criterion used to generate the final samples used in Tables 5, 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 3 and 4	1,444	330
(i) Sample after merging in firm characteristics used in Table 5	1,025	274
(ii) Sample after merging in firm characteristics used in Tables 6, 7, 8, and 9	1,070	291

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 same summary statistics by the type of rTSR benchmark. Observations are at the annual fiscalyear-firm level. Significance levels of the correlation coefficients in Panel B are indicated by *, **, *** for 10%, 5%, and 1% respectively.

Variables are defined as follows (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]. *Index Benchmark* is a dummy variable that equals 1 if the firm uses an index as its relative performance benchmark in a given fiscal year. *Index Preferring* is a dummy that equals 1 if the firm uses an index preferring consultant as defined in the text. *Filtering-Precision Gain* ($\sigma^2_{b,random} - \sigma^2_{b,sbp}$) is the potential gain in filtering precision relative to randomly chosen peers. *Benchmark Noise* is the excess measurement error variance ($\sigma^2_{b,chosen} - \sigma^2_{b,sbp}$) as defined in Table 4. *Number of Peers* is the log number of peers of a firm's chosen rTSR benchmark. *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.

Panel A: Distributional Statistics

	Obs	*Mean*	*Std.Dev.*	*P25*	*Median*	*P75*
ROA	1070	0.049	0.053	0.024	0.044	0.077
Index Benchmark	1070	0.337	0.473	0.000	0.000	1.000
Index Preferring	1070	0.597	0.491	0.000	1.000	1.000
Filtering-Precision Gain	1070	0.003	0.004	0.001	0.002	0.003
Benchmark Noise	1070	0.000	0.003	−0.000	0.000	0.001
Number of Peers	1070	3.651	1.512	2.566	3.045	4.625
CEO Expected Pay	1070	7.476	4.714	4.096	6.296	9.069
CEO Abnormal Pay	1070	1.704	5.212	−0.620	0.905	2.891
CEO Tenure	1070	5.823	4.523	3.000	5.000	8.000
CEO Age	1070	56.760	5.086	54.000	57.000	60.000
% Busy Directors	1070	0.021	0.046	0.000	0.000	0.000
Board Size	1070	10.590	2.059	9.000	10.000	12.000
Director Workload	1070	0.802	0.330	0.583	0.727	0.923
% Age 65+ Directors	1070	0.318	0.317	0.222	0.333	0.500
Log Market Cap	1070	9.037	1.275	8.122	8.891	9.759
Census-based HHI Index	1070	0.072	0.038	0.051	0.060	0.082
Return Volatility	1070	0.079	0.048	0.047	0.067	0.098
Book-to-Market	1070	0.524	0.314	0.306	0.486	0.684

Table A.II.
Continued

Panel B: Firm Characteristics by rTSR Benchmark Type

	Specific Peers (1)	Index (2)	(1) - (2) (3)
Number of Peers	2.80860	5.30653	-2.49792***
	(0.56342)	(1.42112)	(-40.92395)
Index Preferring	0.50635	0.77562	-0.26928***
	(0.50031)	(0.41775)	(-8.78457)
Filtering-Precision Gain	0.00309	0.00218	0.00091***
	(0.00380)	(0.00412)	(3.60439)
Benchmark Noise	0.00005	0.00101	-0.00096***
	(0.00201)	(0.00422)	(-5.05408)
ROA	0.05082	0.04667	0.00414
	(0.05376)	(0.05005)	(1.22011)
CEO Expected Pay	7.41200	7.60302	-0.19102
	(4.87196)	(4.39167)	(-0.62652)
CEO Abnormal Pay	1.00892	3.07004	-2.06112***
	(3.92602)	(6.89486)	(-6.22279)
CEO Tenure	5.62764	6.20776	-0.58011**
	(4.50482)	(4.54038)	(-1.98638)
CEO Age	56.64457	56.98615	-0.34158
	(5.11567)	(5.02741)	(-1.03871)
% Busy Directors	0.01978	0.02255	-0.00278
	(0.04569)	(0.04782)	(-0.92510)
Board Size	10.47955	10.80609	-0.32655**
	(1.90930)	(2.31206)	(-2.45894)
Director Workload	0.78513	0.83593	-0.05080**
	(0.31708)	(0.35131)	(-2.38804)
% Age 65+ Directors	0.29306	0.36645	-0.07339***
	(0.34546)	(0.24479)	(-3.60167)
Log Market Cap	8.98207	9.14367	-0.16160**
	(1.25136)	(1.31543)	(-1.96287)
Census-based HHI Index	0.07066	0.07572	-0.00506**
	(0.03464)	(0.04290)	(-2.07973)
Return Volatility	0.07904	0.07872	0.00032
	(0.04849)	(0.04689)	(0.10264)
Book-to-Market	0.53866	0.49532	0.04334**
	(0.32314)	(0.29260)	(2.14019)
Observations	709	361	1070

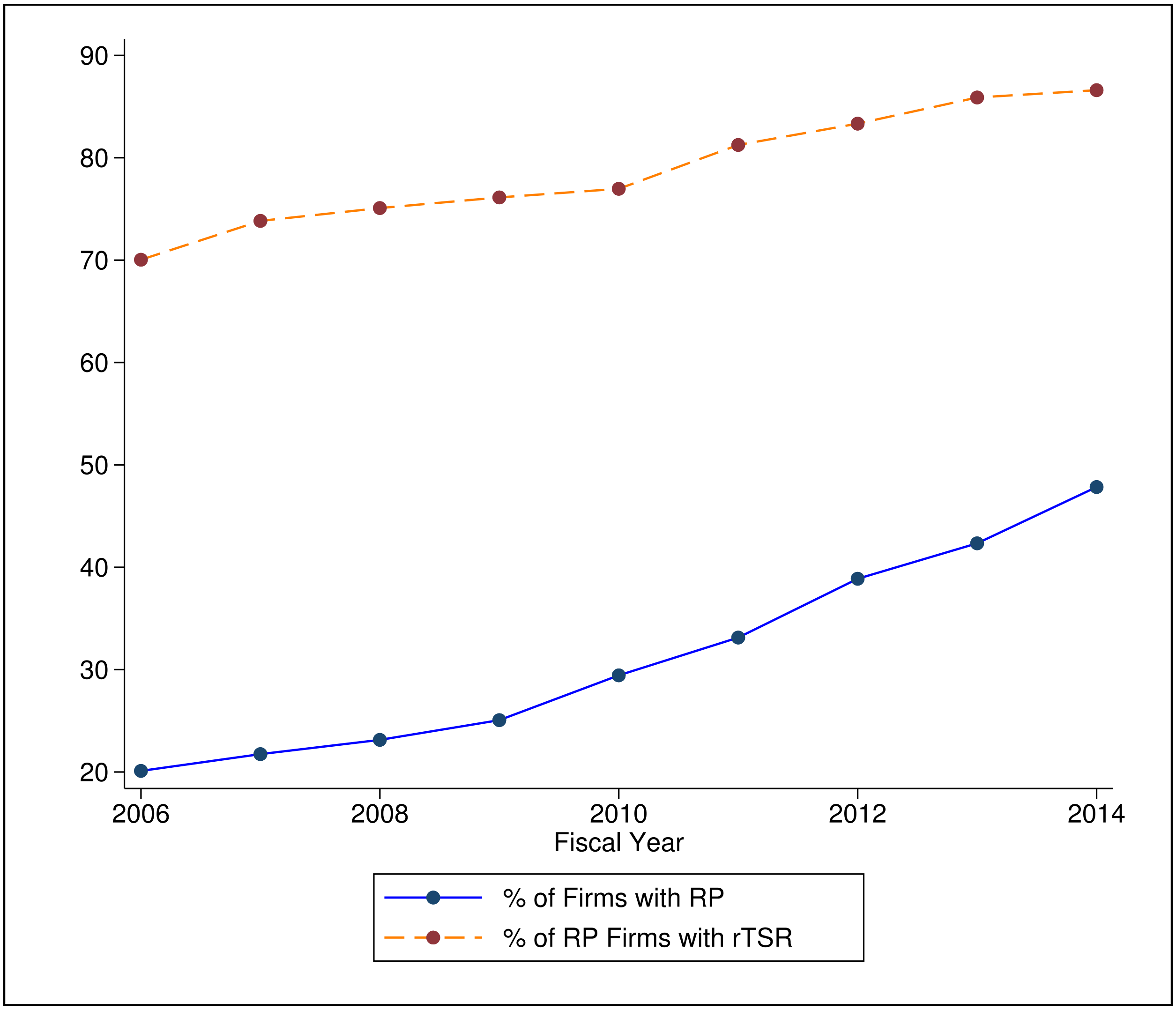
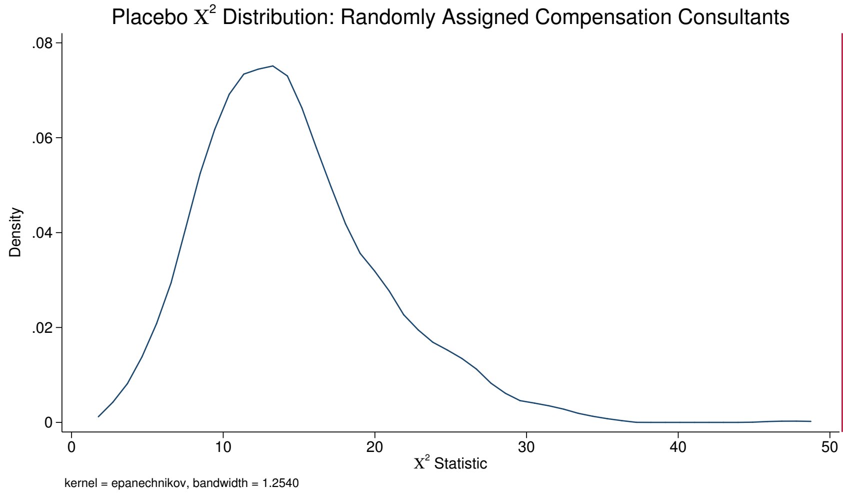
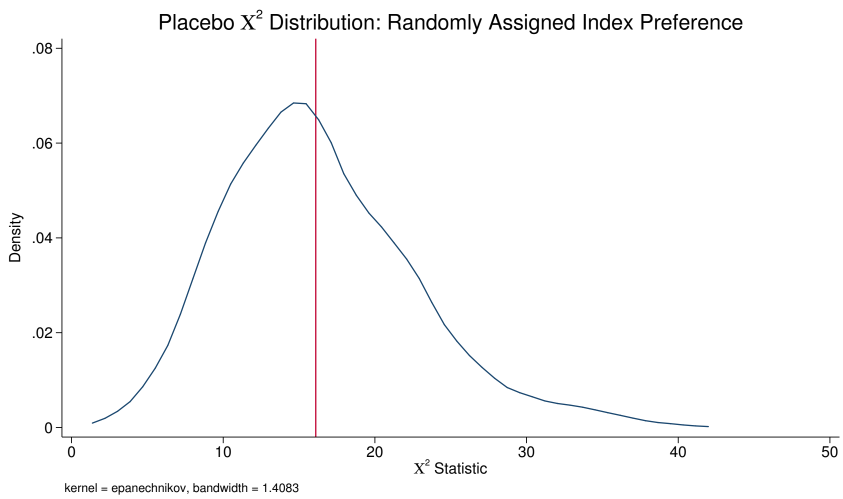


Figure 1. Fraction of Firms Using Relative Performance Contracts 2006-2014 The solid line plots the fraction of firms in the ISS Incentive Labs sample prior to any sample selection restrictions that disclose the award of at least one performance grant based on relative performance (RP) in a given fiscal 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 of index selection only on 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 6. 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	Mean # 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.
Summary Statistics on Types of Relative Performance Benchmarks
2006-2014

Columns 1 and 2 of Panel A of this table summarize the percentages of rTSR-based *grants* associated with either specific peer or index-based benchmarks for fiscal years 2006-2014 using the ISS Incentive Labs data prior to any sample selection restrictions. The fractions do not sum to 1 between columns 1 and 2 because firms that either have unknown benchmarks or use both index and specific peer benchmarks are excluded. Columns 3 (4) reports the average number of peer firms chosen as benchmarks for RP grants associated with specific peers (indexes). Panel B reports the fraction of expected compensation that is based on rTSR at the annual-firm level based on the sample of firms that exclusively use specific peer or index benchmarks grants (firms using both types of grants are dropped). Column 1 reports the expected incentive-plan-based compensation in $ millions and is calculated as the sum of the target dollar value of the Estimated Future Payouts Under Non-Equity Incentive Plan Awards and Grant Date Fair Value of Stock and Option Awards from the values reported in the Grants of Plan-Based Awards Table, both of which assumes that performance targets are met. To compute the proportion of the expected incentive-plan-based compensation attributable to rTSR performance metrics, we assume that each performance metric associated with a grant is weighted equally in the calculation of the grant's value. Column 2 reports the fraction of expected-compensation in column 1 that is attributable to rTSR. The $ amount attributable to rTSR is computed by multiplying the weight on rTSR-based targets with the total expected incentive-plan-based compensation. Columns 3 and 4 estimate the same fraction based on subsamples of firms that either use only specific peers or index benchmarks.

Panel A: rTSR Benchmark Types

	Benchmark Type			
Fiscal Year	Specific-Peer Benchmarks (1)	Index-Based Benchmarks (2)	# of Peers (Specific) (3)	# of Peers (Indexes) (4)
2006	0.55	0.41	16.4	289.8
2007	0.57	0.40	16.1	315.4
2008	0.54	0.40	18.0	338.5
2009	0.56	0.39	18.4	386.5
2010	0.61	0.35	19.0	372.7
2011	0.59	0.37	19.1	330.1
2012	0.58	0.37	18.8	356.9
2013	0.54	0.42	18.9	394.1
2014	0.52	0.44	18.0	395.4
Total	0.56	0.40	18.3	360.9

Panel B: rTSR's Contribution to Grant-Based Expected Compensation

Fiscal Year	Expected Comp ($ Millions) (1)	Fraction rTSR-based (2)	Fraction rTSR-based (Specific) (3)	Fraction rTSR-based (Indexes) (4)
2006	6.88	0.30	0.30	0.29
2007	7.40	0.26	0.26	0.26
2008	7.58	0.28	0.28	0.28
2009	6.16	0.26	0.25	0.29
2010	7.03	0.27	0.27	0.27
2011	7.23	0.26	0.26	0.27
2012	7.56	0.27	0.27	0.26
2013	7.88	0.25	0.26	0.25
2014	7.92	0.24	0.25	0.23
Total	7.42	0.26	0.27	0.26

Table 3.
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 report the across-firm average constant and slope coefficient from time-series regressions, regressing base firm i's returns (R_{it}) on the contemporaneous returns of a portfolio of peers ($R_{p_{it}}$), respectively. 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 and corresponding SBP returns. 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 an index-based 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)
rTSR (N=356)	0.0022^{**}	1.0255^{***}	0.3272
	[0.0009]	[0.0258]	
Specific Peers (N=201)	0.0018	1.0052^{***}	0.8765
	[0.0011]	[0.0329]	
Index (N=155)	0.0026^{*}	1.0520^{***}	0.1864
	[0.0013]	[0.0387]	

Table 4.
Assessing Firms' Chosen RP Benchmarks: Measurement Error Variances

This table reports the distributional properties of the measurement errors for four 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), iii) peers randomly selected from CRSP, and iv) the S&P500 index benchmark. 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 1–4 report this variance, which estimates the MEV up to a constant (Eqn. 5), for the four benchmarks. Columns 5 and 6 report the mean difference in MEVs between chosen benchmarks and their SBPs and between firms' chosen benchmarks and random peers, respectively. Column 7 reports the reduction in variance of $(p - \hat{c})$ from selecting SBPs instead of the firm's chosen performance benchmark. Column 8 reports the fraction of total noise embedded in random peers that chosen peers remove (or total *filtering-precision gain* that is achieved by firms' chosen 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. Row 1 is based on the sample of firms that only use absolute performance grants with TSR being one of the performance metrics. Row 2 is based in our main sample of rTSR firms that satisfy the filters above; the third (fourth) 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.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\sigma^2_{b,chosen} + \sigma^2_e$	$\sigma^2_{b,sbp} + \sigma^2_e$	$\sigma^2_{b,random} + \sigma^2_e$	$\sigma^2_{b,sp500} + \sigma^2_e$	$\sigma^2_{b,chosen} - \sigma^2_{b,sbp}$	$\sigma^2_{b,chosen} - \sigma^2_{b,random}$	$\frac{\sigma^2_{b,chosen} - \sigma^2_{b,sbp}}{\sigma^2_{sbp} + \sigma^2_e}$	$\frac{\sigma^2_{b,random} - \sigma^2_{b,chosen}}{\sigma^2_{b,random} - \sigma^2_{b,sbp}}$
TSR (N=123)	0.0162***	0.0087***	0.0123***	0.0129***	0.0075***	0.0039	0.8704	—
	[0.0033]	[0.0013]	[0.0013]	[0.0030]	[0.0022]	[0.0025]		
rTSR (N=356)	0.0049***	0.0043***	0.0073***	0.0066***	0.0006***	-0.0024***	0.1355	0.8007
	[0.0004]	[0.0003]	[0.0005]	[0.0005]	[0.0002]	[0.0002]		
Specific Peers (N=201)	0.0045***	0.0043***	0.0079***	0.0072***	0.0002	-0.0033***	0.0557	0.9324
	[0.0004]	[0.0004]	[0.0006]	[0.0006]	[0.0002]	[0.0003]		
Index (N=155)	0.0054***	0.0044***	0.0065***	0.0059***	0.0010**	-0.0011***	0.2375	0.5210
	[0.0007]	[0.0004]	[0.0007]	[0.0007]	[0.0004]	[0.0001]		

Table 5.
Assessing Pay for Performance Sensitivity

This table reports the change in CEO pay on both firm's own TSR and the firm's chosen benchmark's TSR. If a firm selects specific peers, the benchmark TSR is the median of the chosen peers' annual returns; if a firm selects index-based benchmarks, the benchmark TSR is the annual index return. All columns include controls for Δ total assets, Δ ROA, and Δ ROE as well as year-fixed effects. Columns 3 and 4 include Industry-fixed effects using the 2-digit Global Industry Classification definitions. Columns 2 and 4 include an interaction of the index benchmark choice with firms' own and chosen benchmark's TSR. Observations are at the annual firm-year level. 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.

	Δ Log CEO Pay			
	(1)	(2)	(3)	(4)
Firm Return	0.328***	0.368***	0.330***	0.371***
	[0.057]	[0.071]	[0.056]	[0.072]
Chosen Peer Return	-0.357***	-0.395***	-0.342***	-0.381***
	[0.122]	[0.135]	[0.123]	[0.137]
Index Benchmark		0.007		0.019
		[0.027]		[0.029]
Firm Return × Index Benchmark		-0.118		-0.118
		[0.113]		[0.115]
Chosen Peer Return × Index Benchmark		0.167		0.170
		[0.188]		[0.188]
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Observations	1,025	1,025	1,025	1,025
Adj R^2	0.052	0.051	0.048	0.046

Table 6.
Explaining Benchmark Selection and Compensation Consultant Styles

This table reports the marginal effects from a probit regression, evaluated at the sample mean for continuous variables and at zero for different indicator dependent variables (dv), of the firm's choice in benchmark selection. Column 1's dv equals 1 if the firm selects a rTSR index-based benchmark and zero if it selects a specific peer benchmark. Column 2's dv equals 1 for firms that tie compensation to rTSR and zero for firms that tie compensation to TSR, and to no other relative performance metrics. Column 1's dv equals one if the firm selects an index-preferring consultant and zero if it selects a specific-peer-preferring consultant. All columns include CEO, board of directors, firm, and industry characteristics controls as defined in Table A.II as well as year and industry-fixed effects using the 2-digit Global Industry Classification definitions. Columns 1 and 2 include compensation-consultant-fixed effects and report the corresponding p-values of the joint F-tests of the significance of the compensation-consultant-fixed effects. The corresponding p-values of the joint F-tests of industry-fixed effects and firm characteristics are also separately reported. Observations are at the annual firm-year level. 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.

	$Pr(Index\ Benchmark) = 1$	$Pr(rTSR) = 1$	$Pr(Index\ Preferring) = 1$
	rTSR Sample	rTSR + TSR Sample	rTSR Sample
	(1)	(2)	(3)
Filtering-Precision Gain	-8.446	-1.300	-7.032
	[7.554]	[1.434]	[6.925]
CEO Characteristics			
CEO Expected Pay	-0.004	-0.002	0.006
	[0.012]	[0.003]	[0.012]
CEO Abnormal Pay	0.015***	-0.004**	0.002
	[0.005]	[0.002]	[0.004]
CEO Tenure	0.008	-0.003*	0.003
	[0.005]	[0.002]	[0.006]
CEO Age	0.004	0.001	-0.004
	[0.006]	[0.002]	[0.005]
Board and Firm Characteristics			
% Busy Directors	0.398	0.249*	-0.424
	[0.480]	[0.138]	[0.539]
Board Size	0.033**	-0.009*	-0.001
	[0.013]	[0.005]	[0.015]
Director Workload	0.236***	-0.040*	-0.085
	[0.076]	[0.024]	[0.063]
% Age 65+ Directors	0.263*	-0.021	0.191
	[0.156]	[0.060]	[0.148]
Log Market Cap	0.013	0.050***	0.037
	[0.045]	[0.016]	[0.046]
Return Volatility	0.817	-0.294	-0.119
	[0.530]	[0.187]	[0.513]
Book-to-Market	0.026	0.005	0.022
	[0.093]	[0.031]	[0.094]
Industry Characteristics			
Census-based HHI Index	0.113	0.575*	-0.073
	[0.590]	[0.321]	[0.797]
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Comp Consultant FE	Yes	Yes	No
p-value of χ^2 (Comp FE)	0.0000	0.2909	.
p-value of χ^2 (Industry FE)	0.0000	0.0000	0.0046
p-value of χ^2 (Controls)	0.0005	0.0000	0.4450
Observations	1,070	1,185	1,070
Pseudo R^2	0.2736	0.3041	0.0662
Mean of Dep. Var.	0.35	0.84	0.60

Table 7.
Impact of Indexes on ROA

Columns 1 to 3 estimate the intent to treat (ITT) effect of being assigned to an index-preferring consultant as defined in Table 6 on firms' ROA. Using the index-preferring consultant as an instrument, columns 4 to 6 report the 2SLS estimates of an indicator for having chosen an index as the rTSR benchmark. The first stage F-stat of the instrument is reported at the bottom of the table for columns 4-6. All columns include CEO, board of directors, firm, and industry characteristics controls as defined in Table A.II as well as year and industry-fixed effects using the 2-digit Global Industry Classification definitions. Observations are at the annual firm-year level. 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.

	ITT			**2SLS**		
	(1)	(2)	(3)	(4)	(5)	(6)
Index Preferring	-0.006*	-0.006*	-0.005			
	[0.003]	[0.003]	[0.003]			
Index Benchmark				-0.026*	-0.025*	-0.035
				[0.014]	[0.014]	[0.026]
Benchmark Noise		-0.876	-0.689		-0.623	-0.850
		[0.667]	[0.675]		[1.028]	[0.971]
Number of Peers			-0.002*			0.004
			[0.001]			[0.006]
CEO Characteristics						
CEO Expected Pay	-0.001	-0.001	-0.001	-0.000	-0.000	-0.000
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
CEO Abnormal Pay	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
CEO Tenure	0.000	0.000	0.000	-0.000	-0.000	-0.000
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
CEO Age	0.001*	0.001	0.001	0.001*	0.001*	0.001*
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
Board and Firm Characteristics						
% Busy Directors	-0.014	-0.014	-0.013	0.030	0.032	0.034
	[0.036]	[0.036]	[0.035]	[0.033]	[0.033]	[0.035]
Board Size	-0.003***	-0.003***	-0.003**	-0.002*	-0.002*	-0.002*
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Director Workload	-0.009	-0.009	-0.008	-0.007	-0.007	-0.007
	[0.005]	[0.005]	[0.005]	[0.006]	[0.005]	[0.006]
% Age 65+ Directors	0.007	0.007	0.009	0.010	0.010	0.007
	[0.011]	[0.011]	[0.011]	[0.013]	[0.013]	[0.013]
Log Market Cap	0.011***	0.011***	0.011***	0.009*	0.008*	0.009*
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.005]
Return Volatility	-0.258***	-0.241***	-0.238***	-0.189***	-0.180***	-0.186***
	[0.059]	[0.060]	[0.059]	[0.056]	[0.056]	[0.056]
Book-to-Market	-0.046***	-0.045***	-0.047***	-0.047***	-0.047***	-0.045***
	[0.009]	[0.009]	[0.009]	[0.010]	[0.010]	[0.011]
Industry Characteristics						
Census-based HHI Index	-0.029	-0.025	-0.029	-0.003	-0.002	0.009
	[0.037]	[0.036]	[0.036]	[0.040]	[0.040]	[0.043]
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
First-Stage F				28.693	27.406	9.741
Observations	1,070	1,070	1,070	1,070	1,070	1,070
Adj R^2	0.346	0.347	0.349	0.309	0.311	0.297

Table 8.
Index Effect and Number of Peers

This table reports subsample results of columns 3 and 6 in Table 7 stratified by the number of peers. The Below (Above) Median # Peers subsamples are determined by whether the firm's number of peers is below or above the sample median. Columns 1 and 2 estimate the intent to treat (ITT) effect of being assigned to an index-preferring consultant as defined in Table 6 on firms' ROA. Using the index-preferring consultant as an instrument, columns 3 and 4 report the 2SLS estimates of an indicator for having chosen an index as the rTSR benchmark. The first stage F-stat of the instrument is reported at the bottom of the table for columns 3 and 4. Observations are at the annual firm-year level. 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.

	ITT		**2SLS**	
	Below Median # Peers	**Above Median # Peers**	**Below Median # Peers**	**Above Median # Peers**
	(1)	(2)	(3)	(4)
Index Preferring	-0.008*	-0.004		
	[0.004]	[0.005]		
Index Benchmark			-0.120*	-0.042
			[0.070]	[0.068]
Benchmark Noise	0.708	-1.104	1.343	-1.363
	[1.165]	[0.787]	[1.500]	[1.246]
Number of Peers	-0.002	-0.002	-0.031	0.010
	[0.005]	[0.002]	[0.020]	[0.020]
CEO Characteristics				
CEO Expected Pay	-0.001	-0.001	-0.001	-0.001
	[0.001]	[0.001]	[0.002]	[0.001]
CEO Abnormal Pay	-0.001	0.000	-0.001	0.000
	[0.001]	[0.000]	[0.001]	[0.001]
CEO Tenure	0.000	-0.000	-0.000	-0.001
	[0.001]	[0.001]	[0.001]	[0.001]
CEO Age	0.000	0.001	0.001	0.002**
	[0.001]	[0.001]	[0.001]	[0.001]
Board and Firm Characteristics				
% Busy Directors	0.004	-0.038	0.016	0.019
	[0.039]	[0.061]	[0.048]	[0.054]
Board Size	-0.001	-0.003**	-0.002	-0.004*
	[0.002]	[0.002]	[0.002]	[0.002]
Director Workload	-0.004	-0.012	-0.014*	0.001
	[0.007]	[0.009]	[0.008]	[0.016]
% Age 65+ Directors	0.014	0.008	0.011	0.011
	[0.017]	[0.015]	[0.025]	[0.015]
Log Market Cap	0.010*	0.013**	0.006	0.011*
	[0.005]	[0.006]	[0.006]	[0.006]
Return Volatility	-0.177**	-0.245***	-0.120	-0.202***
	[0.084]	[0.077]	[0.083]	[0.073]
Book-to-Market	-0.057***	-0.037**	-0.048***	-0.034*
	[0.011]	[0.015]	[0.012]	[0.018]
Industry Characteristics				
Census-based HHI Index	0.033	-0.015	0.113	0.017
	[0.070]	[0.047]	[0.128]	[0.049]
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
First-Stage F			6.951	2.453
Observations	544	526	544	526
Adj R^2	0.343	0.370	0.124	0.302

Table 9.
Sensitivity of ITT Estimates

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 by:

$$\delta \frac{\sigma_{index,observable}}{\sigma^2_{observables}} = \frac{\sigma_{index,unobservables}}{\sigma^2_{unobservables}}.$$

Zero selection ($\delta = 0$) corresponds to column 1 of Table 7. Following Oster (2017), Panel A of this table assumes that R^2_{max} equals 1.3X the $\tilde{R}^2$ of column 1 in Table 7 (≈ 0.47). Panel B assumes that R^2_{max} equals 2X $\tilde{R}^2$ (≈ 0.73). Panel C assumes that $R^2_{max} = 1$. All columns include compensation consultant-, year-, 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)			
	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$
	(1)	(2)	(3)	(4)
Panel A: $R^2_{max} = 1.3\tilde{R}^2$				
Index Preferring	-0.007**	-0.007**	-0.008**	-0.009**
	[0.003]	[0.003]	[0.003]	[0.003]
Panel B: $R^2_{max} = 2\tilde{R}^2$				
Index Preferring	-0.008***	-0.011***	-0.014**	-0.017**
	[0.003]	[0.004]	[0.005]	[0.007]
Panel C: $R^2_{max} = 1$				
Index Preferring	-0.010***	-0.015**	-0.021**	-0.029
	[0.004]	[0.006]	[0.010]	[0.041]
Year FE	Yes	Yes	Yes	Yes
Industry 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	1,070	1,070	1,070	1,070