Relative Performance Benchmarks: Do Boards Get It Right?*

Paul Ma
University of Minnesota
Carlson School of Management

Jee Eun Shin
Harvard Business School

Charles C.Y. Wang
Harvard Business School

February 7^{th}, 2017

Abstract

Standard principal-agent models suggest that boards design incentive contracts that filter out common shocks in performance to motivate costly effort from the CEO—a process entailing the judicious selection of benchmarks for relative performance evaluation (RPE). We evaluate the efficacy of firms' chosen RPE benchmarks and document that, relative to a normative benchmark, index-based benchmarks perform 14% worse in their time-series return-regression R^2 and 16% worse in measurement error variance; firms choosing specific peers only modestly under-perform. Structural estimates suggest that, absent frictions, the underperformance of index-based benchmarks imply a counterfactual performance penalty of 106-277 basis points in annual returns. Consistent with these estimates, firms choosing index-based benchmarks exhibit lower annual returns and ROA. Finally, reduced-form analyses suggest that the inefficient benchmarking is associated with governance-related frictions. Collectively, these findings provide new evidence on the explicit practice of RPE and its implications for corporate governance and firm performance.

JEL: D86, G30, J33, J41, M12
Keywords: Empirical contract theory; executive compensation; relative performance evaluation; common shock filtration; search-based peers; board of directors; corporate governance

*The authors can be contacted at paulma@umn.edu, jeshin@hbs.edu, and charles.cy.wang@hbs.edu. We have benefited from advice and suggestions from Hengjie Ai, Santiago Bazdresch, Alan Benson, Gerardo Perez Cavazos, Frank Gigler, Guojin Gong, Thomas Hemmer, Sudarshan Jayaraman, George John, Chandra Kanodia, Dave Larcker, Eva Liang, Melissa Martin, Xiaoxia Peng, Jonathan Nam, Ruidi Shang, Matt Shaffer, John Shoven, Pervin Shroff, Martin Szydlowski, Chao Tang, Akhmed Umyarov, Felipe Varas, Martin Wu, Ivy Zhang and seminar participants at Georgia State University, Hong Kong University Accounting Research Symposium, and the Singapore Management University Accounting Research Symposium. We are also grateful to discussants Gaizka Ormazabal (SMU), Eric Yeung (HKUST), and Brian Cadman (FARS). We 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 Grant. All errors remain our own.

1 Introduction

The provisioning of performance-based incentives to managers of publicly traded companies is a quintessential application of the standard principal-agent framework. A central prediction of these models is that the principal—in practice the board—will design contracts that filter out common shocks to performance, i.e., that are outside the CEO's control, to optimally induce costly effort from the risk-averse manager, in order to maximize firm performance (Holmström, 1979, 1982).

This central prediction has spawned a large body of literature in economics, finance, and accounting seeking to detect the existence of relative performance evaluation (RPE) or common shock filtration in executive compensation or turnover decisions—albeit with mixed success.[2] A major empirical challenge facing the prior literature had been that firms did not publicly disclose if and how they employed RPE. Under these limitations, researchers relied on implicit tests of RPE, requiring them to define the board's choice of the peer group—typically chosen from broad industry groupings (Antle and Smith, 1986) augmented by similarities in certain firm characteristics (Albuquerque, 2009). Highlighting the joint hypothesis problem, the major takeaway from this body of work is that the presence of RPE in CEO compensation or turnover is highly sensitive to the researcher's selection of the firm's relative performance benchmarks (e.g., Albuquerque, 2009; Lewellen, 2015; Jayaraman, Milbourn and Seo, 2015; Nam, 2016; Drake and Martin, 2016).

Since 2006, the identity of firms that *explicitly* use performance benchmarking to incentivize executives were revealed due to SEC's reforms in mandatory compensation disclosure.[3] These disclosures revealed that, since 2006, an increasing number of companies provide CEOs with incentive contracts that have an explicit RPE component—typically referred to as a performance share contract. Among the sample of companies—approximately 1,500 of the largest firms in the stock market—collected by ISS Incentive Lab, reported in Figure 1, the proportion of firms that employ explicit RPE in CEO contracts increased from around 20% in 2004 to nearly 50% in 2014.

Given the increasing prevalence and importance of RPE incentives in CEO contracts, a more pressing and open problem for researchers is an assessment of the efficacy of firms' RPE practices, as described in their performance share contracts, rather than whether or to what extent firms engage

[2]For example Table 1 of Albuquerque (2009) provides an useful enumeration of the mixed evidence.

[3]See http://www.sec.gov/rules/final/2006/33-8732a.pdf for additional details regarding the nature of these reforms.

in RPE. Indeed, a recent WSJ article covering the 2016 Nobel Prize in Economics for contract theory voiced concerns over whether observed RPE practices adhere to the theoretical principles: "That we do not see [RPE contracts] more in practice is often a failure of governance" (Creighton, 2016).

This paper is the first to assess the efficacy of firms' selected RPE peer firm benchmarks in performance share contracts. Our starting point is the classic principal-agent model of Holmström (1979) and Holmström (1982), which predict that RPE peers are optimally chosen to filter out the common component of performance. Based on this framework, we first examine the extent to which firms' chosen RPE peers capture common shocks in performance and analyze their theoretical implications for firm performance, absent frictions in peer selection. Then, we examine potential sources of frictions or alternative theories and economic factors that could explain why certain firms' RPE peers do not appear to adequately capture common shocks in firm performance.

We document four main findings. Our first set of analyses compares the performance of firms' disclosed peer benchmarks to that of a normative RPE benchmark. In particular, we rely on Search-Based Peers (SBPs), a set of economically-related peer firms identified based on investor perceptions as reflected in their searches on the SEC EDGAR platform. Lee, Ma and Wang (2015) and Lee, Ma and Wang (2016) show that SBPs exhibit superior ability relative to other state-of-the-art peer identification schemes in explaining the variations in firms' stock returns, valuation multiples, and fundamental performance characteristics.

Overall, we find that firms' disclosed benchmarks perform significantly worse in explaining the co-movement of stock returns. This underperformance is concentrated in the set of firms (nearly half of the sample) that use index-based benchmarks, while firms using specific peers perform only modestly worse. Compared to a portfolio of top 10 SBPs, firms' disclosed index-based benchmarks explain 14% less of the time-series variation in the firms' monthly stock returns (i.e., time-series regression R^2), whereas the underperformance is merely 2.3% for firms using specific peers. We stress that these findings are not unique to the choice of SBPs as the normative benchmark—an alternative benchmark based on firms which commonly share analyst coverage (i.e. analyst co-covered peers or ACPs) as the base firm yields similar results (Lee *et al.*, 2016). Together, these findings raise questions concerning the appropriateness of choosing indexes—effectively benchmarking to a large number of peers—in lieu of a narrower peer set, a concern that compensation consultants have also echoed.[4]

To provide more structure in assessing the "efficacy" of RPE benchmarks, we extend the standard principal-agent framework by introducing errors from measuring the common component of performance. We show that poorer selection of RPE benchmarks (larger measurement error variances) implies lower efficacy in filtering out common shocks and leads to poorer incentives for managers. Critically, this framework enables us to structurally estimate the variance of the measurement error up to a scalar constant for any set of RPE peers. Consistent with the time-series R^2 results, our estimates suggest that firms' chosen index-based benchmarks yield measurement error variances that are *at least* 16% higher relative to SBPs, whereas specific-peer-based benchmarks exhibit measurement error variances that are *at least* 4.4% higher.

Our second set of findings suggests that, under the principal-agent framework and in the absence of frictions from selecting a precise set of peers, the observed underperformance of firms' RPE benchmarks has significant performance consequences due to the manager's choice of effort. Our structural estimates suggest, for plausible ranges of risk aversion, that the measurement error variances of firms' chosen RPE benchmarks through their effect on effort, imply an on-average performance penalty of 60 to 153 basis points in annual stock returns, and 92 basis points at the midpoint of the risk aversion range. These magnitudes are economically significant, particularly in light of the relatively large firms that constitute our sample. Consistent with our findings in R^2s, these effects are concentrated and largest in the subset of firms that use index-based benchmarks, for whom we find an on-average performance penalty of 106 to 277 bps in annual stock returns across the range of risk aversion.

Consistent with these theoretical performance implications, our third set of findings shows that the choice of an index-based benchmark is cross-sectionally associated with lower *realized* firm performance, as measured by annual ROA and stock returns. Specifically, firms with index-based benchmarks perform 80 basis points lower in ROA and 320 basis points lower in returns relative to firms with specific peers as benchmarks. We caveat that this reduced-form analysis

[4] For example, while praising the idea of relative total shareholder returns as "seductive in theory," Barry Sullivan, managing director of the compensation consulting firm Semley Brossy, said during the 2014 National Association of Corporate Directors Leading Minds of Compensation Conference that "the challenge is in practice, do we really have a peer group that we feel good measuring? If not, do we then have to extend to a broader market index, and if we do that, are we stepping too far away from our core business or are we introducing some noise around the relative TSR construct?"

treats firms that use specific-peer-based benchmarks as counterfactuals to those firms that use index-based-benchmarks. To the extent that our control variables do not fully capture the differences in underlying characteristics that could be associated with firm performance between these two types of firms, it is possible that the estimated performance differences are overstated. Overall, however, these reduced-form estimates provide empirical support to the conclusions of the structural analyses, that poorer RPE benchmark selection implies economically significant performance consequences.

Having documented the adequacy of firms' selected RPE benchmarks in capturing common shocks and their performance consequences, our final analyses examine the sources of economic frictions or alternative theories that can explain the observed underperformance of firms' chosen RPE benchmarks. We first endogenize the board's choice of RPE benchmarking efficacy, and the extended model predicts less efficacious benchmarking when the firm has greater idiosyncratic volatility, and when the manager or the board is of lower quality or ability (due to lower marginal incentive effects from improving benchmarks). Our empirical evidence does not find support for firm-level volatility or managerial quality driving benchmarking inefficacy; instead, we find evidence that the lower monitoring quality of the board, or governance weaknesses, are associated with poorer benchmarks and the decision to choose an index-based benchmark. Second, we argue that the observed poor performance is not the expected outcome of plausible alternative theories, i.e., outside of our model, for how boards may select peers in performance contracts: for example, if the manager is capable of self-insuring against the common factor (Garvey and Milbourn, 2003); or if the actions of firms influence peer performance (Janakiraman, Lambert and Larcker, 1992; Aggarwal and Samwick, 1999a).

Together, our results can be interpreted in one of two ways. A stronger view is that they suggest that the selection of RPE peers are important for the inducement of executive effort, and the inability to capture common shocks has significant performance implications for firms. Alternatively, a weaker interpretation of our findings is that the quality of RPE peers provides incremental information about the quality of a firm's board and governance.

We contribute to the literature on relative performance evaluation, as related to CEO compensation or turnover.[5] In contrast to this body of work which primarily seeks to test the comparative

[5]See Warner, Watts and Wruck (1988); Gibbons and Murphy (1990); Janakiraman, Lambert and Larcker (1992); Murphy and Zimmerman (1993); Parrino (1997); Aggarwal and Samwick (1999b); Bertrand and Mullainathan (2001); Engel, Hayes and Wang (2003); Kaplan and Minton (2012); Jenter and Kanaan (2015); Lewellen (2015); Jayaraman,

statics of RPE and the selection of RPE benchmarks under a joint hypothesis problem, we provide first evidence on the efficacy of firms' chosen RPE benchmarks in capturing the common component of performance, for those firms that explicitly provide RPE incentives to their CEOs.

Our findings—that the observed variations in RPE benchmark adequacy are associated with governance-related frictions and also with worse firm performance—contribute to the literature examining the consequences of board monitoring (e.g., Core, Holthausen and Larcker, 1999; Bertrand and Mullainathan, 2001; Fich and Shivdasani, 2006; Morse, Nanda and Seru, 2011) by highlighting a novel channel through which poor monitoring can affect executive incentives and firm performance. Moreover, our results inform the design and practice of CEO compensation contracts. In particular, our analyses show that a substantial amount of incremental stock return co-movement can be captured by a judicious selection of RPE peers, and that a normative set of peers (SBPs or ACPs) suggested by Lee, Ma and Wang (2015, 2016) can provide such an improvement.

This study also relates to the growing body of work that has emerged since the 2006 mandate to disclose details of compensation benchmarking practices. In general, our work differs in the method of analysis, both analytically and empirically, of the board's choice of RPE benchmark quality. While prior work by Gong, Li and Shin (2011) suggests that RPE peers remove common shocks better than random firms (a lower bound), we instead compare these RPE peers to a normative upper bound (SBPs), as well as an alternative lower bound (S&P500 index), and examine the potential explanations for benchmarking inadequacy. The contemporaneous work of Bizjak, Kalpathy, Li and Young (2016) also examines the properties of firms' disclosed RPE benchmarks. In particular, it analyzes how CEO compensation would have changed had the RPE peer group been different from what was chosen, and documents, using simulations, that CEO compensation is not on average influenced through RPE peer selection. In contrast to this paper, we study the efficacy of common shock filtration and their performance implications through costly effort—an effect through the second moment, rather than a first moment, of compensation.

The rest of the paper is organized as follows. Section 2 relates our results and empirical approach to the existing literature. Section 3 lays out our data and descriptive statistics illustrating the rise of explicit grant-based relative performance benchmarking and provides empirical evidence on the efficacy of the board's choice of benchmarks. Section 4 maps our empirical test to the

Milbourn and Seo (2015); Nam (2016); Drake and Martin (2016).

principal-agent framework in order to recover primitives which describe the efficacy of relative performance benchmark and its performance implications. Section 5 investigates the potential sources and alternative theories for the observed ineffective benchmarking. Section 6 concludes.

2 Related Literature and Background

The standard principal-agent framework aims to solve the problem of eliciting costly hidden effort from a risk-averse agent by balancing incentives with risk sharing. Specifically, the informativeness principle in Holmström (1979) states that any contractible sufficient statistic informative of the agent's effort choice should be incorporated into the compensation contract in order to improve the incentive-risk sharing trade-off.

A particular application of this principle is the usage of RPE, whereby the agent's performance is filtered to remove a common shock to performance which is unrelated to the agent's effort choice. The clear benefit of this approach is that it increases the agent's optimal choice of effort and performance, which Ghosh and John (2000) and Rubin and Sheremeta (2015) have experimentally verified.

This powerful prediction has motivated academics to seek empirical evidence of RPE in the settings of either CEO compensation or turnover. The standard reduced-form test of the model associates measures of the CEO's compensation/turnover as the dependent variable regressed on measures of the firm's performance and the common component of performance (e.g., Gibbons and Murphy, 1990). The reduced-form predictions for RPE is a positive coefficient on the firm's own performance and a negative coefficient on the measure of the shock to common performance. Intuitively, these regressions test whether CEO compensation or turnover responds to noise (the common shock) or the agent's effort (the idiosyncratic component of performance).

In testing variations of the reduced-form compensation regression above, researchers have generally turned in mixed results.[6] An empirical challenge acknowledged in this literature is that

[6] The mixed results in turn have motivated additional theoretical reasonings into why boards might not filter for common shocks: e.g., product market competition and more broadly the effect of firm's own effort on the common factor of performance or mis-measurement of the common factor (Janakiraman *et al.*, 1992; Joh, 1999; Schmidt, 1997; Aggarwal and Samwick, 1999a; Vrettos, 2013; Antón *et al.*, 2016); alternative CEO preferences (Lazear, 2000; Garvey and Milbourn, 2003; Gopalan *et al.*, 2010; Feriozzi, 2011; Hemmer, 2015; DeMarzo and Kaniel, 2016); alternative functional forms in the production technology (Himmelberg and Hubbard, 2000; Celentani and Loveira, 2006; DeMarzo *et al.*, 2012; Hoffmann and Pfeil, 2010); endogenizing the outside labor option (Oyer, 2004; Rajgopal *et al.*, 2006; De Angelis and Grinstein, 2016); or multi-tasking (Holmström and Milgrom, 1991; Feltham and Xie, 1994; Baker,

the board's choice of peers used to estimate the common component of performance as well as the actual method of RPE, i.e., the contract form mapping idiosyncratic performance to payout, is often unobservable to the econometrician. This opacity in turn forced researchers to make strong assumptions about the peer selection process. While traditional approaches identified peer firms based on broad industry groupings, such as the 2-digit standard industry classification (SIC) scheme (Antle and Smith, 1986; Barro and Barro, 1990; Janakiraman *et al.*, 1992; Jenter and Kanaan, 2015), a growing literature has found that identifying the "right" set of RPE benchmarks is critical to finding evidence of RPE in the context of the compensation or turnover regressions. More recent approaches include augmenting industry groupings by matching on size (Albuquerque, 2009), using product market competitors, either directly identified (Lewellen, 2015) or inferred through textual analysis (Jayaraman *et al.*, 2015), or using firms whose financial statements are most comparable (Nam, 2016), or matching firms based on their life cycle (Drake and Martin, 2016). The general finding from this recent literature is that the presence of RPE in CEO compensation or turnover hinges on the usage of these more sophisticated identification schemes for the peer firms that the board of directors are likely to have chosen.

Since 2006, firms have been required to disclose additional details of their executive compensation practices. These disclosures include which peer firms or benchmarks the board uses to evaluate relative performance-based grants. Using the first post-regulation year of data, Gong *et al.* (2011) re-investigates the relationship between CEO compensation and peer performance using the disclosed performance peers. It confirms that the CEO's compensation is decreasing in RPE peers' performance, suggesting that peer group selection matters. Furthermore, it finds that firms' chosen RPE peers exhibit greater fundamental similarity compared to a set of randomly chosen firms, consistent with the objective of common shock filtering. However, lacking a normative benchmark, it is unable to determine how much of the common shock firms' chosen RPE peers remove (e.g., relative to an upper bound) (Kleinberg, Liang and Mullainathan, 2015) and it does not assess the economic implications of the quality of RPE peer selection for firm performance.

A related set of papers test for whether a firm's choice to engage in RPE (i.e., on the extensive margin) is consistent with its costs and benefits as predicted in (Gibbons and Murphy, 1990). Carter, Ittner and Zechman (2009) studies a sample of UK firms and find that the propensity to engage

2002).

in RPE—identified via explicit disclosures—is not associated with the degree of a firm's exposure to common shocks.[7] In contrast to these papers, both Gong *et al.* (2011) and Li and Wang (2016) find the opposite result in the propensity to engage in RPE for those US firms with explicit RPE contracts post 2006. Tice (2015) finds a positive association between the firm's decision to engage in explicit RPE on the extensive margin with firm performance and investment efficiency.

Francis, Hasan, Mani and Ye (2016) documents a positive association between the quality of a firm's compensation peers and the firm's performance, using managerial ability scores. Their argument is based on a peer effects channel where firms work harder in order to mimic their successful compensation peers, but they do not find any such effects with performance peers.

Because compensation peers tend to be larger in market capitalization than their focal firms, there is also a literature that studies whether these peers could be a result of managerial rent-seeking. For example, whereas Bizjak, Lemmon and Naveen (2008) and Faulkender and Yang (2010) find evidence of opportunistic benchmarking with compensation peers, Cadman and Carter (2013) does not. Note that the performance benchmarks we study are distinct from the compensation peers which serve to estimate the manager's outside option. In our framework, there is no role for rent-seeking within the selection of performance benchmarks because they do not affect the outside option of the manager.

The principal-agent model we estimate follows Holmström and Milgrom (1987), who show that the solution to dynamic moral hazard problems can be reduced to a series of spot linear contracts. Our structural estimation strategy is similar in principle to Schaefer (1998) who estimates the compensation contract's sensitivity to performance and its relationship to firm size. However, we differ in the focus towards relative performance evaluation in providing estimates of the measurement error in the common component of performance. Similar to our approach, Dikolli, Hofmann and Pfeiffer (2012) also model measurement error in the common shock of performance; however, they focus on understanding how various forms of the error structure can bias implicit tests for RPE, which we do not perform because the relative performance contracts are directly observed.

Recent work which seeks to quantify the loss from moral hazard in the CEO setting include Gayle and Miller (2009) and Ai, Kiku and Li (2016). Gayle and Miller (2009) estimates another version of

[7] More recently, Bell and Van Reenen (2016) finds evidence of pay for luck among UK firms that use explicit relative performance contracts. Like Bertrand and Mullainathan (2001), they find the phenomenon to be concentrated in the set of firms with poor governance characteristics.

the Holmström and Milgrom (1987)model to quantify how much of the growth in compensation in the last quarter century can be attributed to the moral hazard channel. Ai *et al.* (2016) estimates another dynamic agency model and finds that eliminating moral hazard would result in a 3.4% increase in aggregate output. Whereas their analysis focuses on quantifying moral hazard through the ratio of observable to unobservable noise in performance, we focus on the extent of the incentive misalignment through ineffective benchmarking.

3 Data and Descriptive Evidence of Benchmarking Behavior

3.1 Data Description

Our data comes from ISS Incentive Lab, which collected details of compensation contracts and incentive awards for named executive officers at the individual grant level from firms' proxy statements (DEF 14A).[8] Data coverage begins in 2004 and includes every firm which was ever ranked in the top 750 in terms of their market capitalization in any given year. Due to backward- and forward-filling, each year our ISS Incentive Lab sample (2004-2014) encompasses the entire S&P500, most of the S&P midcap 400, and a small proportion of the S&P small-cap 600. Roughly speaking, each annual cross section encompasses the largest 1,000 firms in the stock market in terms of market capitalization.

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

Panel A, Table 1 provides summary statistics on 13,324 CEO grants awarded by 1,673 unique firms across the 2004-2014 period. Over this period, the number of grants awarded to CEOs increased from 2 grants per year to over 3.5 grants. The proportion of incentive awards paid out in cash has remained stable since 2006, at roughly 35% of all CEO grants while in the same time period,

[8] Recent work leveraging this data to study CEO compensation include: Bettis, Bizjak, Coles and Young (2014); De Angelis and Grinstein (2016); Bizjak, Hayes and Kalpathy (2015); Tice (2015); Gallani (2015); Drake and Martin (2016); Bizjak, Kalpathy, Li and Young (2016).

stock-based payouts have increased from 36% to 49% with option-based payouts declining from 29% to 15%. Notably, the proportion of CEO grants which have a relative evaluation component ("Abs/Rel" or "Rel") has more than doubled, from 8% in 2006 to 17% in 2014.[9]

Panel B, Table 1 suggests that the number and proportion of companies that explicitly provide RPE incentives (i.e., issued RPE-based) grants have more than doubled since 2006. Moreover, Panel C suggests that among such firms, the use of stock returns as an evaluative metric in RPE-based grants have also risen dramatically. Specifically, Figure 1 shows that whereas 20% of the companies in our ISS Incentive Lab sample provide RPE-based grants to their CEOs in 2004, 48% of firms do so by 2014. Furthermore, whereas 70% of the companies providing RPE-based grants use stock returns as an evaluative metric in 2006, 87% do so by 2014.[10] Together, the summary statistics shown in Table 1 illustrate the rising importance and pervasiveness of explicit RPE contracts as well as the importance of stock returns in evaluating relative performance.

To further understand the economic magnitude of RPE-based contracts to CEO incentives, Table 2 provides estimates of the relative importance of meeting RPE performance thresholds. In particular, we seek to estimate how much incremental compensation would be earned by the CEO if she met the minimum RPE-based target performance thresholds, assuming all other incentives are earned. Column 3 provides an estimate of the expected total compensation when all incentives are earned, including (but not exclusive to) meeting all minimum RPE-based targets.[11] Columns 4 and 5 report an estimate of the amount of total expected compensation (in column 3) stemming from meeting RPE-based targets and from meeting stock-based RPE targets, respectively. [12] Our estimates suggest that the incremental increase in executive compensation arising from meeting all RPE targets and all stock price-based RPE targets—"Incentive Ratio" reported in columns 6

[9]The increasing trend in the explicit use of RPE-based grants is consistent with descriptive evidence from the prior literature. For example, our summary statistics are comparable to Bettis, Bizjak, Coles and Young (2014), who also use data from ISS Incentive Lab spanning the time period 1998-2012 (e.g., see their Table 1). Francis *et al.* (2016) documents summary statistics on a hand-collected sample of RPE usage based on S&P 1500 firms for the period 2006-2010. They report considerably smaller numbers of firms that use RPE than studies based on the ISS Incentive Lab data. However, these differences are likely due to their focus on RPE with explicitly-stated peers which excluded cases where indexes are used for RPE purposes.

[10]Gong *et al.* (2011) find a similar prevalence of the use of stock price in RPE contracts in a hand-collected sample of firms from the cross section of proxy statements from the first year of the 2006 SEC disclosure reforms.

[11]Expected compensation is calculated using the values as reported in the "Grants of Plan-Based Awards Table" by adding the target dollar value of the "Estimated Future Payouts Under Non-Equity Incentive Plan Awards" and the "Grant Date Fair Value of Stock and Option Awards."

[12]To gauge the expected compensation derived from different performance metrics, we calculate the equally-weighted portion of expected compensation that corresponds to each performance metric.

and 7—is economically significant.[13] Specifically, meeting all minimum RPE targets increases the CEO's compensation by 104-145% between 2004 to 2014; similarly, meeting stock-based RPE targets increases CEO compensation by between 41-75%.

Our back-of-the-envelope estimates are consistent with the existing and growing evidence on the importance of performance-based, and in particular RPE-based, incentives for CEOs. For example, Bettis *et al.* (2014) shows that the RPE-related components of RPE-grant-issuing firms between 1998 to 2012 consistently determine more than 30% of the realized total compensation amount. Similarly, De Angelis and Grinstein (2016) shows that, for a hand-collected sample of S&P500 firms in 2007, about a third of the firms explicitly mention that their performance-based awards are RPE-based, and that firms with RPE contracts attribute about half of the estimated total performance award value to RPE. In addition, they also document that about 75% of the performance metrics associated with RPE are market measures which is consistent with the notion that stock price-based measures are predominant for RPE purposes.

In Table 3, we provide information on the different types of benchmarks used for RPE purposes. The sample of RPE grants is the same as in Panel B of Table 1. In particular, we consider four different benchmark categories: the use of a specific peer set, the S&P500 index, the S&P 1500 index, and other indexes (generally industry-based). Columns 4-7 report the percentage of RPE grants that use each particular type of benchmark in a given fiscal year. Column 8 reports the percentage of RPE grants for which the benchmark cannot be specified (missing). Because each grant can be associated with multiples types of benchmarks, the sum of the percentages across columns 4 to 8 can exceed one. Finally, column 9 reports the average number of peer firms for firms which select a specific peer set. In summary, we observe that around half of all RPE grants choose specific peers as a benchmark, with an average number of peers between 15 to 18. For firms which choose an index benchmark, the most popular choice is the S&P500.

[13] We calculate grant-level incentive ratios which denote the incremental potential incentive when the CEO meets either all her RPE-based targets or all her stock price-based RPE targets in column 4 and 5, respectively. In particular, the incentive ratio in column 6 (7) is calculated as "expected compensation if all targets are met" divided by "expected compensation if all other targets excluding all RPE-based targets (stock price-based RPE targets) are met." For example, the RPE-based incentive ratio of 2.07 in 2014 implies that on average, CEOs meeting their RPE-based targets, potentially can earn more than double of the counter-factual in which they miss their RPE-based targets.

3.2 Explaining Common Shocks in Stock Returns: Firms' Chosen RPE Benchmarks vs. SBPs

Given the rising prevalence of explicit stock-return-based RPE contracts, and given the important role they play in CEO incentives, our paper seeks to assess the efficacy of firms' explicit RPE practices. In particular, we focus on a previously under-explored aspect of RPE: the ability of selected peer firms to filter out the common component of stock returns. To assess the performance of firms' chosen RPE benchmarks, we compare them to the Search-Based Peer firms (SBPs) of Lee *et al.* (2015) as a normative benchmark. SBPs represent firms' economic benchmarks as collectively perceived by investors and inferred from the co-search patterns on SEC's Electronic Data-Gathering, Analysis, and Retrieval (EDGAR) website. The findings of Lee *et al.* (2015, 2016) suggest that SBPs dominate other state-of-the-art methods for identifying economically-related firms in terms of explaining the co-movement in stock returns, valuation multiples, growth rates, R&D expenditures, leverage, and profitability ratios. Among the 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 compared to 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. Although common shocks affecting a firm's stock returns are unobservable, SBPs serve to provide a lower-bound estimate in the importance of such common shocks. For contrast, we also compare firms' chosen RPE benchmarks to the S&P500 index as a normative lower bound estimate of the common shock.

Since we are concerned with those firms that *explicitly* provide RPE incentives to its CEOs, we restrict attention to the subsample of firms covered by ISS Incentive Lab that issued RPE-based grants to their CEOs and who disclose the peers or indexes (market or industry-based) used in determining performance payouts (i.e., the sample described in Panel B of Table 1). Table 3 reports summary statistics on the type of peers used as the RPE benchmark for this subsample. Notably, in 2014, 49% of RPE grants to CEOs identify specific peer firms as the relative benchmark, while 22% use either the S&P500 or 1500 indexes, 17% use some other index (e.g., narrower or industry-specific indexes), and 15% do not specify the actual peer benchmark.[14] The distribution of RPE peer types

[14] Due to multiple metrics per grant, these percentages do not necessarily sum to 1.

has remained stable in the 8-year period from 2006 to 2014. Among the firms choosing an index, the distribution of index choices has also remained stable. In 2014, for example, 45% of the RPE grants using an index chose the S&P500, 12.5% chose the S&P1500, and the remaining 42.5% chose other indexes.

To simplify the analysis, we focus on those firms whose RPE contracts use stock returns as the relevant performance metric (i.e. the sample described in Panel C of Table 1), which, as discussed above, constitute the vast majority of the RPE grants in our sample. We also restrict the analysis to those firms for whom there is sufficient data to construct SBPs. In total, our sample contains 356 unique firm-benchmark-type (i.e., index vs. specific peers) observations between fiscal years 2006 to 2013. Detailed construction of our final sample is reported in Appendix Table A.I.

To investigate the efficacy of firms' chosen benchmarks in explaining common shocks, our first test estimates the average R^2 values from firm-specific time-series returns regressions of the form:

$$R_{it} = \alpha_{it} + \beta_i R_{ipt} + \epsilon_{it} \tag{1}$$

where R_{it} is firm i's monthly cum-dividend returns in period t and R_{ipt} is the benchmark return of a firm's benchmark peers' returns. For firms that select a set of specific RPE peer firms, we use the median of the peer set's returns, reflecting firms' most common choice of the performance target around which RPE payouts apply linearly (Reda and Tonello, 2015). For firms that select an index as the RPE benchmark, we use the corresponding index returns. For the RPE benchmarks disclosed in the proxy statement for a given fiscal year, we use returns from the following fiscal year in estimating R^2s. For example, if a firm reports its fiscal year end date as December 2000, we obtain monthly stock return data for the calendar window January 2001 to December 2001 for firm i and its corresponding performance peers as disclosed in that proxy statement to calculate R_{ipt}.

We compare the R^2s generated from firms' selected peer benchmark returns to those obtained from both the S&P500 index and the (search-traffic-weighted) returns of firms' SBPs. As shown in Lee *et al.* (2015) and Lee *et al.* (2016), weighting SBPs by the relative magnitude of EDGAR co-search fractions, interpreted as a measure of similarity or relevance between firms, performs best in explaining the variations in base-firm returns. Following Lee *et al.* (2015), to avoid look-ahead bias, SBPs are always identified using search traffic from the prior calendar year. Returns data are obtained from CRSP monthly files, and firms with fewer than ten months of valid monthly returns in total are excluded from the sample. To facilitate comparisons, all the regressions are conducted using the same underlying set of base firms. The average number of monthly returns per firm is 37.

Panel A, Table 4 reports the cross-sectional means of the resulting time-series R^2 values. Columns 1-2 in the first row show that across all 356 unique firm-benchmark observations in the sample, using the S&P500 as the benchmark yields an average R^2 of 32.8%. In contrast, firms' chosen RPE benchmarks produce an average R^2 of 48.3%, which is significantly higher (at the 1% level) than that produced by the S&P500 (as reported in column 4). Column 3 reports the cross-sectional mean of the time-series R^2s produced by the search-traffic-weighted portfolios of firms' top-10 SBPs. The average R^2 value of 51.8% is significantly higher (at the 1% level) than that produced by firms' chosen RPE benchmarks (as reported in column 2).[15] In summary, firms' chosen peers exhibit an economically significant advantage of a 15.5% higher R^2 in comparison to the S&P500 and a modest under-performance of 3.5% relative to SBPs.

Rows 2 and 3 of Table 4, Panel A examines the efficacy of firms' selected benchmarks for the subset of RPE firms that use specific peers (N=201), and an index (N=155) respectively. We find that the under-performance of firms' RPE benchmarks is concentrated among the set of firms using index-based benchmarks, among whom the average time-series R^2 is 40.0%. By comparison, the average R^2s produced by SBPs (46.4%) represent not only a statistically significant improvement but also an economically significant one (i.e., a 16% proportional improvement). However, firms choosing index-based benchmarks continue to outperform the 32.9% R^2 generated from the S&P500, albeit by a smaller amount—a 21.6% proportional improvement—relative to those firms choosing specific peers, whose average R^2 outperform those produced by the S&P500 proportionally by 67.1%.[16] Indeed, among the firms choosing specific RPE peers, their benchmarks produce an average R^2 of 54.8%, which is statistically no different from the average R^2 of 56.0% generated from firms' SBPs.

Panel B of Table 4 assesses the robustness of these findings using alternative peer performance measures: the mean and the 75^{th} percentile of peer portfolio returns. Since these variations do not affect index returns, these robustness tests only focus on firms that use specific RPE peers. Similar to the results found in Panel A, across the set of specific-peer-benchmarking firms, the mean

[15] An equal-weighted portfolio of the top-10 SBPs produces similar results, with 50% average R^2.

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

(75^{th} percentile) of chosen peers' returns yield an average time-series R^2 of 54.7% (51.7%) in return regressions. The under-performance of 4.4% relative to SBPs is statistically significant at the 1% level for the 75^{th} percentile of specific peer performance, but not significant for the mean of the portfolio of specific peers.

In summary, we find that the performance of firms' chosen RPE benchmarks, in terms of the time-series R^2s from time-series regressions, lies somewhere on the spectrum between those generated by the S&P500 at the lower bound and SBPs at the upper bound. Specific-peer-based benchmarks chosen by firms lie on the upper end of this range, as their performance in R^2s is statistically indistinguishable from those of SBPs. Index-based benchmarks chosen by firms, on the other hand, lie somewhere in the middle of the performance range, significantly outperforming the S&P500 index on average but also underperforming SBPs. While these findings rely on SBPs as the normative benchmark, we also find similar results (available upon request) using an alternative normative benchmark consisting of peer firms with whom the base firm shares analyst coverage.[17] Given that the inefficiency in benchmarking resides in firms that benchmark to a large number of firms, through the choice of broad market or industry-based indexes, it is unsurprising that our main findings are robust to the choice of alternative normative benchmarks.

These findings suggest that, among the subset of firms with explicit RPE contracts that select specific peers, their boards appear to be judiciously and carefully selecting peer firms whose performance filters out the common performance shocks. On the other hand, for the remaining firms that choose an index-based benchmark, representing 43.5% of the sample, there appears to be significant room for improvement relative to SBPs. An alternative possibility is that these choices reflect optimal contracting of heterogeneous firms, a point we return to in Section 5. In the next section, we investigate the ramifications of the underperformance in R^2 through the lens of the standard principal-agent framework.

[17] Lee *et al.* (2016) shows that, compared to a set of state-of-the-art peer identification schemes, peers defined through analyst co-coverage (ACPs) are among the best performing (second to SBPs) in explaining the out-of-sample variation of base firms' stock returns.

4 Interpreting the R^2 Differences

Although the differences in R^2s reported in the last section are suggestive, it is difficult to interpret the economic magnitudes and the implications of these R^2 differences for firm performance. To interpret and assess the ramifications of the underperformance in R^2s of firms' chosen RPE benchmarks, we adopt the principal-agent framework of Holmström and Milgrom (1987), which allows us to map the R^2 construct into the model's primitives.

4.1 Basic Setup

The starting point of our model follows Holmström and Milgrom (1987) and Gibbons and Murphy (1990). We assume a risk-neutral principal (the board) and a risk-averse agent (the CEO), and that firm performance follows a factor structure consisting of (i) unobserved managerial effort $[a]$, (ii) a common shock that is outside of the manager's control $[c \sim_{iid} N(0, \sigma_c^2)]$, and (iii) a firm-specific idiosyncratic shock $[\epsilon \sim_{iid} N(0, \sigma^2)]$:[18]

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

A potential concern is whether the factor structure above is a reasonable assumption under our setting. Fortunately, one testable implication of Eqn. 2 is that the coefficient on the benchmark portfolio in Eqn. 1 equals 1. In un-tabulated results, we find that the estimated slopes are approximately 1 which provides credence to our assumption.

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

$$\max_a e^{-\eta\left(w - \frac{\kappa}{2}a^2\right)} \text{ s.t. } w \geq \underline{w}, \tag{3}$$

where $\underline{w}$ represents the manager's reservation wage and η is her CARA risk aversion. The manager's

[18]The assumption that the common shock is outside of the manager's control is consistent with interpreting these shocks as macroeconomic events.

optimal effort choice (and expected firm performance) is given by

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

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

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

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

$$\mathbb{E}[-e^{-\eta[w-\frac{\kappa}{2}a^2]}] \geq u(\underline{w}), \text{ and} \tag{4-PC}$$

$$a \in \text{argmax } \mathbb{E}[-e^{-\eta[w-\frac{\kappa}{2}a^2]}], \tag{4-IC}$$

and the optimal relative performance contract is given by

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

The first component of the optimal contract, α^*, is the fixed portion of the manager's compensation and reflects her exogenously determined outside option.[19] The second component, $\beta^* = \frac{1}{1+\eta\kappa\sigma^2}$, represents the pay-performance sensitivity portion of the contract.[20]

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

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

Thus, to motivate optimal effort, the principal designs a contract that rewards managers for effort by perfectly removing the common component that drives firm performance and is outside of the manager's control.[21]

The key comparative static for our purposes is the negative effect of the variance in idiosyncratic

[19] $\alpha^* = \underline{w} + \frac{\kappa\eta\sigma^2-1}{2\kappa(1+\kappa\eta\sigma^2)^2}$. Although this component is typically determined through benchmarking to compensation peers to measure the outside option $\underline{w}$. Note that compensation peers are distinct from the RPE peers we study. See Gong *et al.* (2011) and De Angelis and Grinstein (2016) for further details on the comparison of the two types of peer groups.

[20] The empirical relative performance literature has been built around testing the predictions of Eqn. 6 by examining whether the coefficient on own performance p is positive and the common shock (peer performance) c is negative for CEO compensation or turnover.

[21] Note that this contract is strictly preferable to the one without relative performance evaluation i.e. when $w^* = \alpha^* + \beta^* p$ and $a^* = \frac{1}{\kappa+\eta\kappa^2(\sigma^2+\sigma_c^2)}$ as long as $\sigma_c^2 > 0$

shocks to performance through managerial effort: $\frac{\partial a^*}{\partial \sigma^2}$ from Eqn. 7. The intuition is that all else equal, higher σ^2 means that a greater proportion of the firm's performance is unpredictable—or explained by idiosyncratic shocks—even after filtering out the common component of performance. Thus, the manager's compensation, driven by inferred effort level $p - c = a + \epsilon$ is more likely driven by factors outside of the manager's control (i.e., noise), reducing the manager's incentives to exert effort.

One may be concerned whether real world contracts exhibit the linearities assumed in the model and thus whether the comparative statics are meaningful. The validity of the linear contract framework has garnered both theoretical and empirical support in the context of CEO compensation. On the empirical side, Haubrich (1994) calibrates the model using observed data and documents plausible values of the model's primitives, in particular values of β^* that are consistent with the empirical findings in Jensen and Murphy (1990). Aggarwal and Samwick (1999b) and Jin (2002) test the model's theoretical prediction, and finds empirical support, that firms with greater idiosyncratic volatility in performance (higher σ^2) exhibit lower pay-performance sensitivity (lower values of β^*). With respect to the assumption of linearity itself, the literature argues that such contracts can be rationalized by a principal's preference for robustness to uncertainties of the agent's action space (Hurwicz and Shapiro, 1978; Chassang, 2013; Carroll, 2015).[22]

To better understand whether explicit RPE contracts exhibit linearity in relative performance, we collect a random sample of 20 contracts from our main sample (across both specific peer and index benchmarking firms). We find two standard contract forms, both of which are effectively linear in relative performance. 3 out of 20 of the contracts are (piecewise) linear in relative performance of stock returns. One example of such a contract is illustrated in Figure 2. The remaining 17 contracts are (piecewise) linear in the firm's performance as a percentile of benchmark peers' stock-return distribution.[23] One example of such a contract is illustrated in Figure 3.

Figure 2 shows that Ansys has a payout multiplier which is linear in its stock returns relative to the returns of the benchmark, the NASDAQ composite index in this case. Two additional features are worth noting. First, the contract has a kink point around the target performance level where

[22] The Holmström and Milgrom (1987) model also derives optimal contracts as linear because incentives are history independent.

[23] We infer this from proxy statement disclosures of "linear interpolation" in performance between pre-specified target levels.

own firm returns equal the index returns. Second, there is floor to extreme under-performance and a cap to extreme out-performance—typically referred to as the threshold and max payouts respectively.

Figure 3 shows that Intersil has a payout multiplier which is linear in the firm's performance as a percentile of the peer stock-return performance. Again, there is a ceiling and a floor for extreme outperformance and underperformance, and there are kink points at the 75^{th} and 25^{th} percentiles.

Critically, although the majority of our sample of contracts are linear in the percentile of firm performance, we verify that these type of contracts are effectively linear in stock returns relative to peers. In Figure 4, we plot for the firms in our whole sample the relationship between stock returns and the percentile of performance relative to the peer distribution using monthly return data during our main sample period. This figure shows that the firms' stock returns is (piecewise) linear in returns percentiles (defined from the benchmark return distribution). In other words, performance contracts that are linear in performance percentiles are effectively linear in the performance relative to a fixed percentile of the benchmark distribution (e.g., the median). Numerically, the regression fit between firm returns and the percentile of the benchmark yields a R^2 of 96.7%. Therefore, the data supports the idea that performance contracts are (locally) linear in relative performance.[24]

4.2 Imperfect Common Shock Filtration

Having established the baseline case, we now depart from Gibbons and Murphy (1990) by introducing imperfect filtering and assume that the principal (i.e., the board) observes the common shock with error,

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

where $\omega_b \sim_{iid} N(0, \sigma_b^2)$.[25] Here, lower σ_b^2 represents the greater ability of performance peers or benchmarks to remove common shocks, and perfect common shock filtering reduces to the special case where $\sigma_b^2 = 0$.[26]

[24] Walker (2016) provides additional examples of local linearity in performance share plans.

[25] We assume that ω_b has mean zero which is without loss of generality in the manager's optimization choice in effort. A non-zero mean would enter into the level of the manager's expected compensation. A recent practitioner's report finds that firms beat the target payout 48% of the time in 2015 which suggests that ω_b is close to zero. http://paygovernance.com/wp-content/uploads/2016/12/Viewpoint-66.pdf

[26] Note that this formulation has the same analytical prediction as the original Holmström and Milgrom (1987) and Aggarwal and Samwick (1999b) framework of a second signal of firm performance (performance peers/benchmarks) with the two signals sharing a correlation ρ. One can think of picking better peers/benchmarks in the two models as

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

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

Notably, poorer measurement of the common shock (higher σ_b^2) reduces the equilibrium effort level and expected firm performance.[27]

This is because measurement errors introduce additional noise in the manager's compensation, in particular her inferred effort level $p - \hat{c} = a + \omega_b + \epsilon$. Thus, similar to the baseline case above, poorer measurement of the common shock induces the manager to choose a lower level of effort.

The R^2 results of Table 4 can be interpreted in the framework of this model, since there is a one-to-one mapping with the variance in measurement errors. In particular, the time-series return regression $p_t = \delta_1 + \delta_2 \hat{c}_t + \epsilon_t$ yields an R^2—the squared correlation coefficient ($\rho^2_{p,\hat{c}}$)—that can be expressed as a function of the primitives of the model:

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

For a given firm—i.e., fixing σ^2 and σ_c^2—lower σ_b^2 corresponds to higher R^2. Therefore, the results of Table 4 imply that SBPs produce lower measurement error variances in the common performance factor relative to firms' chosen RPE benchmarks.

In fact, the assessment of peer benchmarking adequacy reported in Table 4 can be re-cast in terms of measurement error variances, which is ultimately the economic quantity of interest to our analysis. Under the model assumptions, the following data moments—the variances in the prediction errors from peer benchmarks—can identify the measurement error variances up to a scalar constant:

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

Although we cannot identify the magnitude of the measurement error variances, their differences

either a decrease in σ_b^2 or an increase in the absolute value of ρ.

[27]Note that this RPE contract is preferable to the no RPE contract as long as $\sigma_c^2 > \sigma_b^2$ and defined as the boundary condition in Wu (2016). This condition can be empirically tested by checking whether $Var(p) > Var(p - \hat{c})$, which in un-tabulated results, we verify holds.

between one set of peer benchmarks from another can be identified, a refinement over the R^2 measures which, as shown in Eqn. 10, contains σ_c^2, but also other factors as well. Moreover, these sample moments allow us to obtain a lower bound estimate on the proportional improvement between two benchmark candidates.[28]

4.3 Structural Estimates of Measurement Error Variances and Performance Implications

Rows 1-3 of Panel A, Table 5 presents the simple method of moments parameter estimates of Eqn. 11 for the S&P500 (row 1), firms' chosen RPE benchmarks (row 2), and SBPs (row 3), where p represents monthly stock returns of the base firm. $\hat{c}_{sp500}$, $\hat{c}_{sbp}$, and $\hat{c}_{firm}$ are monthly stock returns of the S&P500 index, (traffic-weighted-average) returns of firms' SBPs, and the median returns of firms' chosen RPE benchmarks, respectively.[29] In column 1, the estimated $\sigma^2_{b,firm} + \sigma^2$ across the whole sample equals 44.489, whereas $\sigma^2_{b,sbp} + \sigma^2$ equals 40.742 for a statistically significant difference (at the 1% level) of 3.747. On a relative basis, firms' chosen RPE benchmarks produce *at least* 9.2% greater variance in measurement errors. Columns 2 and 3 of the same panel recover estimates for the subset of firms that selected specific peers and indexes, respectively. Similar to our findings in Table 4, SBPs' out-performance of firms' chosen RPE benchmarks is concentrated in the set of firms that use indexes. For index-based firms, their measurement error variances are at least 16.4% greater than SBPs, whereas for firms using specific peers, their variances are at least 4.4% greater, both statistically significant at the 1% level. In summary, our findings on the performance of firms' chosen RPEs in terms of measurement error variances, the ultimate construct of interest, are empirically and theoretically consistent with the earlier R^2 results of Table 4.[30]

Given the greater measurement error variance implicit in the benchmarks chosen by firms, our main goal is to quantify the economic implications in terms of managerial effort and expected firm performance, which can be estimated using the sample analogue of Eqn. 9. In particular, given the manager's risk aversion (η) and effort-cost parameter (κ), the impact of poorer RPE benchmarks on

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

[29]Under a similar framework, Schaefer (1998) provides estimates of the term $\eta\kappa\sigma^2$, but does not incorporate measurement error into the common factor of performance, hence our estimates are generally not comparable.

[30]Our procedure also yields an estimate for the cost of effort $\kappa \approx 0.17$ which can be interpreted as the unit of effort required to generate one basis point in returns.

expected performance is given by

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

This computation requires identification of the risk-aversion (η) and the effort-cost (κ) parameters; however, with 3 unknowns (κ, η, $\sigma^2 + \sigma_{b,firm}^2$) and 2 equations (9 and 11), the model is under-identified. Our calibration, therefore, borrows accepted ranges of the risk aversion parameter η from the prior literature.[31] We also restrict $\kappa > 0$, consistent with the model, since otherwise Eqn. 9 would yield two roots. Row 4 of Panel A, Table 5 reports method of moments estimates of κ, under the assumption that $\eta = 0.625$, the midpoint of the range considered in Haubrich (1994).[32]

Performance implications of SBPs' improvement over firms' chosen RPE benchmarks are obtained by applying the method of moments parameter estimates and the assumed risk-aversion parameter to Eqn. 12. Across all firms in the sample, we estimate the counter-factual performance gained (lost) under SBPs (S&P500) to be 91.73 (281.83) basis points in annual returns. In other words, the on-average under-performance of firms' selected benchmarks—in terms of explaining the common component of firm performance—implies an economically significant performance effect.[33] These performance implications are again driven through the set of firms that select index-based benchmarks. Interestingly, we find that firms that selected specific peers would have lost 385 basis points had they instead selected the S&P500.

The estimates in Table 5 come from a pooled sample, rather than the average of parameters across firms, in order to provide sufficient power to estimate κ. For robustness, Table 6 report the average parameter estimates across firms from re-estimating Eqn. 11 on a firm-by-firm basis. Table 6 also estimates the measurement error variances using alternative peer performance measures: the mean and 75^{th} percentile of peers' returns. In general, we find similar but slightly larger magnitudes in our estimate of error variance in comparison to the pooled estimates, suggesting that pooling or

[31] Under-identification of the risk aversion parameter is common (see, e.g., Gayle and Miller, 2015).

[32] Haubrich (1994) considers the range of η to lie between 0.125 and 1.125 which corresponds to CRRA values of 1.1 to 9.9 (the standard range in the macro literature) based on a CEO wealth of 8.8 million in order to rationalize the possibility of low performance pay sensitivity documented in Jensen and Murphy (1990).

[33] In un-tabulated results, we also estimate the performance consequences (in annual returns) of counter-factually switching to SBPs using the lower and upper bounds of the manager's risk-aversion parameter considered in Haubrich (1994). We find an effect size of 60 corresponding to the lower bound and 153 basis points corresponding to the lower bound for all firms and a range of 106 to 277 for index-based benchmarking firms. Relative to our sample median annual returns, this represents an economically significant 4.6% to 11.6% proportional decline overall and between 7.9% to 20.6% for the sub-sample of index-based benchmarking firms.

averaging across firm-specific estimates yield similar conclusions.

5 Understanding the Sources of Ineffective Benchmarking

The structural estimates of the prior Section suggest that, in the absence of frictions from selecting precise peer benchmarks, the on-average underperformance of firms' selected RPE benchmarks, in particular among firms selecting index-based benchmarks, imply performance penalties that are economically large. Our hypothesis is that these economic magnitudes are, at least in part, rationalized by certain economic frictions, to which we turn our attention below. We begin by extending the baseline model to endogenize the board's problem of benchmark selection. We then use the model's predictions to guide our empirical investigation into the potential plausible economic explanations for the observed underperformance in RPE benchmarks.

5.1 Comparative Statics of Benchmarking Efficacy σ_b^2

To analyze the potential sources of ineffective benchmarking more formally, we generalize the problems faced by the board and the manager introduced in Eqn. 5 and 8, which assume that the quality of the benchmarking technology available to the board (σ_b^2) is exogenously determined. Instead, we now assume that improving the benchmark peers' quality, in terms of filtering out common shocks (or lowering σ_b^2), entails costly effort on the part of the board, and model the cost function to be quadratic in peer quality.

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

$$\sigma_b^{2*} = \arg\max_{\sigma_b^2} f(\sigma_b^2;\theta,\kappa,\sigma^2) = \arg\max_{\sigma_b^2} \frac{1}{2\kappa + 2\kappa^2\eta(\sigma_b^2+\sigma^2)} - \underline{w} - \frac{1}{2}\theta\left(\frac{1}{\sigma_b^2}\right)^2, \tag{13}$$

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

Because the objective function exhibits increasing differences in σ_b^2 with respect to each of the state variables (i.e., $\frac{\partial^2 f}{\partial\sigma_b^2\partial\theta} > 0$, $\frac{\partial^2 f}{\partial\sigma_b^2\partial\kappa} > 0$, and $\frac{\partial^2 f}{\partial\sigma_b^2\partial\sigma^2} > 0$), by Topkis' Theorem, the model yields the following three predictions. First, the level of peer precision is decreasing in the board's cost of effort or monitoring ($\frac{\partial \sigma_b^{2*}}{\partial \theta} > 0$). In other words, boards will be more likely to exert the effort to search for better benchmarks when it is more skilled or higher-quality monitors (e.g., less distracted or captured). Second, the level of peer precision is increasing in the CEO's quality or ability ($\frac{\partial \sigma_b^{2*}}{\partial \kappa} > 0$). Third, the level of peer precision is decreasing in the level of volatility in firm performance ($\frac{\partial \sigma_b^{2*}}{\partial \sigma^2} > 0$). The intuition for the latter two predictions is that boards are more likely to exert the effort to produce better benchmarks when the marginal benefits are higher: i.e., when the managers are more likely to exert effort as a result of better filtering, either because their cost of effort is lower or they more talented (lower κ) or because their efforts contribute to firm performance more (lower σ^2).[34] Below we test these hypotheses by examining how the characteristics of the CEO, the board, and the firm may explain the observed variation in the performance of RPE benchmarks.

5.2 Empirical Drivers of Benchmarking Efficacy

To test these hypotheses empirically, we first construct measures of benchmarking adequacy that assesses the performance of firms' selected RPE benchmarks relative to the normative range represented by the S&P500 at the lower bound and SBPs at the upper bound. The first measure of benchmarking adequacy is based on each firm's time-series R^2, and is defined as $\left(R^2_{rpe} - R^2_{sp500}\right) / \left(R^2_{sbp} - R^2_{sp500}\right)$. Intuitively this proxy measures the percentage of the performance range—represented by the difference in R^2s from SBPs and from the S&P500—that is achieved by firms' chosen RPE benchmarks. The higher the value in this measure, the more adequate are firm's chosen benchmarks. To mitigate the effect of outliers, we winsorize this measure at the 1% level. Similarly, we define a second measure of benchmark adequacy based on the measurement-error-variance proxy, defined as $\left(\sigma^2_{sp500} - \sigma^2_{rpe}\right) / \left(R^2_{sp500} - R^2_{rpe}\right)$. The interpretation of this measure and our treatment of its outliers remain unchanged.[35]

Table 7 reports the results of OLS regressions of the benchmarking adequacy measures on a set

[34] This set up also yields the technical result that $\frac{\partial \sigma_b^{2*}}{\partial \eta} > 0$ when $\kappa\eta(\sigma^2 + \sigma_b^2) > 1$ and $\frac{\partial \sigma_b^{2*}}{\partial \eta} < 0$ when $\kappa\eta(\sigma^2 + \sigma_b^2) < 1$. That is, there is a non-linear relation in the marginal effect of the manager's risk aversion on the quality of peers selected. All else equal, boards choose lower-quality peers when managers become more risk averse among those managers with a "high" degree of risk aversion (i.e., $\eta > \frac{1}{\kappa(\sigma^2 + \sigma_b^2)}$); on the other hand, boards choose higher-quality peers when managers become more risk averse among those managers with a "low" degree of risk aversion (i.e., $\eta > \frac{1}{\kappa(\sigma^2 + \sigma_b^2)}$). This is because the marginal benefit of improving benchmark quality is non-linear in the manager's risk aversion. We do not emphasize this comparative static since we cannot empirically observe or proxy for variations in managers' risk aversion.

[35] Note that these two outcome variables are correlated but not collinear, as Table A.II reports a correlation of 0.51.

of CEO, board, and firm characteristics that measure the model's primitives as well as an indicator for choosing index-based benchmarks. We include three proxies for CEO characteristics—*Log of CEO Pay*, *CEO Tenure*, and *CEO Age*—that could capture managerial talent or cost of effort; we include four proxies for board characteristics—*% Busy Directors*, *Board Size*, and *Director Workload*—that could capture board quality or cost of monitoring; finally, we include five measures of firm-level characteristics—*Log Market Cap*, *Return Volatility*, *Book-to-Market*, *Staggered Board*, and *Dual-Class Shares*—that could capture the relative importance of idiosyncratic shocks in firm performance as well as additional governance characteristics.[36]

Under the model's predictions, we expect positive associations between RPE benchmark adequacy with *Log of CEO Pay*, *CEO Tenure*, or *CEO Age*, which can be interpreted as measures of CEO skill or ability. We expect negative associations with *% Busy Directors*, *Director Workload*, and *Board Size*, which can be interpreted as measures of board distractedness (in the case of busy directors and director workload) or incentives to free ride (in the case of board size), which approximate the board's cost of effort. Similar, we also expect to see negative associations with *Staggered Board* and *Dual-Class Shares*, director-entrenching governance mechanisms that can also increase the directors' cost of monitoring or the extent to which the board members are captured. We also expect to see positive (negative) associations with *Return Volatility* and *Book-to-Market* (*Log Market Cap*), since higher (lower) values in these variables reflect greater fundamental volatility.

Columns 1-3, Table 7 reports the OLS results using the R^2-based measure of benchmarking adequacy, where the columns vary by whether year- or industry-fixed effects are included and standard errors are clustered at the firm level. Since our dependent variable relies on the notion that the S&P500 index and SBPs represent the normative lower and upper bounds of performance, we restrict our analysis here to those subsets of firm-level observations for which SBPs indeed outperform S&P500 (about 80% of the sample). Interpreting the results of column 3, which includes both year- and industry-fixed effects, three types of characteristics are significantly associated (at the 10% level) with benchmarking adequacy. First, the choice of an index is associated with a significant decline (at the 1% level) in the relative performance of firms' chosen benchmarks, consistent with the findings of Tables 4-6. Moreover, the degree of adequacy is negative and significantly associated (at the 10% level) with the percentage of busy directors on the board, consistent with the model's

[36]The specifics of variable construction are detailed in Table A.II.

prediction, but also negative and significantly associated (at the 10% level) with CEO tenure, which is inconsistent with the model predictions. One interpretation of this result is that, instead of reflecting CEO ability or talent, CEO tenure could capture the extent of the CEO's power on the board.

Results reported in columns 4-6, Table 7, using the measurement-error-variance-based measure of benchmarking adequacy, provide similar results. Interpreting the results of column 6, which includes both year- and industry-fixed effects, we find that again the choice of index is negatively and significantly (at the 1% level) associated with benchmarking adequacy. Moreover, we find that the presence of dual-class shares, in which typically a small number of shareholders hold disproportionate voting rights, are also negatively and significantly (at the 1% level) associated with benchmarking adequacy, consistent with the model's predictions on the cost of board monitoring (which we expect to be higher for dual-class firms in which the board members are likely to be more captured).

Together, the firm characteristic that is most consistently and significantly associated with the quality of firms' chosen RPE benchmarks is the firm's choice of an index. The effects we document are not only statistically significant but economically large as well: the choice of an index is associated with in a decline in benchmarking efficacy of 0.54 standard deviations for the R^2 measure and 0.42 standard deviations for the measurement error variance measure.

Given this finding, we further test the model by investigating the economic determinants for selecting index-based benchmarks, an indicator for low-quality RPE peers. Table 8 reports the marginal effects from a probit regression of an indicator for having chosen an index-based benchmark on the same set of CEO, board, and firm characteristics of Table 7.

The results of Table 8 suggest that the choice of index-based benchmarks is associated with board-level governance weaknesses. In particular, interpreting the marginal effects of column 3, we find that the likelihood of choosing an index-based benchmark is positively and both economically and statistically significantly (at the 5% level) associated with the size of the board and directors' workload.[37]

We also do not find evidence that lower precision in RPE benchmarks is explained by lower CEO ability or effort. In particular, we do not find evidence of a negative and significant coefficient

[37] Column 3's point estimates suggest a marginal effect on the choice of index-based benchmarks of 5.9% and 6.6% for a one standard deviation increase in *Board Size* and *Director Workload* respectively.

on *Log CEO Pay*, *CEO Tenure*, or *CEO Age*. In fact, in all three specifications, we find a positive and significant coefficient on *Log CEO Pay*. One interpretation for this result is that, conditional on the controls, higher CEO pay reflects *excess* pay and thus an outcome of board-level governance weaknesses. If so, the observed positive and statistically and economically significant coefficient on *Log CEO Pay* is consistent with the model's prediction that lower precision in RPE peers could result from lower board monitoring quality.[38]

Moreover, we do not find consistent evidence that lower precision in RPE benchmarks is explained by higher volatility in firm performance. The coefficients on *Return Volatility* and *Book-to-Market* are not significantly positive in all three specifications. In fact, in columns 1 and 2 (with no fixed effects and with time-fixed effects), we find negative coefficients on *Book-to-Market* that are significant at the 10% level. Finally, we do find, in columns 1 and 2 (with no fixed effects and with time-fixed effects), negative coefficients on *Log Market Cap* that are significant at the 5% level. However, the significance does not survive the inclusion of industry-fixed effects in column 3.

Overall, we do not find consistent evidence that volatility in firm performance or CEO talent and skill explain the observed variation in benchmark adequacy. Instead, the evidence of Tables 7 and 8 suggest that the source of friction that may explain the observed poor performance of RPE benchmarks are associated with corporate governance and board monitoring quality issues, echoing the findings of Bertrand and Mullainathan (2001).

5.3 Discussion of Alternative Theories

We also consider several plausible alternative theories, outside of this model, that could explain the observed variation in benchmarking quality. One possibility is that choosing closely-related industry peers may not be appropriately capturing the exogenous component of performance. RPE contracts are predicated on the assumption that the firm's actions do not affect the performance of its benchmarks; actions of firms operating in oligopolistic industries may impact the performance of their industry competitors, violating the informativeness principle (Holmström, 1979) and invalidating their use as RPE benchmarks. In fact, under such conditions, RPE contracts may reward CEOs for the common shock (Janakiraman *et al.*, 1992; Aggarwal and Samwick, 1999a) and encourage firms

[38]Column 3's point estimates suggest a marginal effect on the choice of index-based benchmarks of 9.8% for a one standard deviation increase in *Log CEO Pay*.

to soften product market competition.[39] Thus, one prediction from such a theory is that larger firms wishing to remove market-level volatility in the firm's performance measures may be more likely to adopt broad indexes. However, our empirical results do not support such a theory. In Table 8, we find no statistically significant (at the 10% level) and positive relationship between firm size and the choice of an index-based benchmark.

It is also possible that perfect filtration of common shocks may not be optimal under certain circumstances. The dynamic agency models of Hoffmann and Pfeil (2010) and DeMarzo, Fishman, He and Wang (2012) predict that boards will imperfectly filter observable and *persistent* common shocks to performance. This prediction implies that the difference in efficacy between specific peers and index-benchmarking firms can be driven by greater common shock persistence in the latter group. However, in untabulated results, we do not find evidence consistent with this theory. Using SBPs to proxy for the common factor, we do not find a difference in the persistence of SBPs returns between firms which use specific peers versus those using index-based benchmarks.

One alternative explanation which Garvey and Milbourn (2003) offers is that RPE may not be necessary for managers with greater ability to self-insure against the common factor, as they benefit less from better benchmarks. If so, it is possible that the selection of index peers among certain CEOs may reflect non-RPE-based motivations. However, our empirical results are not consistent with these predictions. In particular, we do not find significantly positive associations (at the 10% level) between *CEO Age*, a common proxy for the ability to self-insure, and the choice of index-based benchmarks in Table 8; the coefficients are also weak in economic magnitudes.

Another possibility could be that the choice of performance peers could be due to aspirational reasons. Under such a theory, the selection of peer benchmarks would push managers of the firm to generate performance commensurate with and exceed those of well-performing firms.[40] However, such a theory is less likely to explain our empirical results, which are based on the analysis of firms

[39] A similar point is also made in Antón *et al.* (2016) which argues that the optimal RPE contract for a principal with common ownership across firms will reward instead of filter the common shock in order to soften competition and maximize common ownership weighted total profits. To the extent that this argument holds, we believe it must operate at the extensive margin of the decision to have a performance share contract and thus outside of our sample. For instance, based on a random sub-sample of performance share contracts, we found no instances of contracts with payouts increasing (rather than decreasing) in the performance of the benchmark.

[40] Other models such as Hayes and Schaefer (2009) argue that when there is asymmetric information about the match surplus between the manager and the firm and when boards care about short-run perceptions of firm value, then boards may inflate the CEO's wage as a costly signal of the match surplus. This effect suggest a rationalization for why boards may choose aspirational firms as compensation peers, but why it would also apply for the selection of performance peers is less obvious.

that rely on the relative performance of stock returns. Because stock returns combine the market's ex-ante discount rate with changes in market's expectations about future cash flows and market's discount rates (Campbell, 1991), they can be quite volatile relative to, for example, accounting-based performance measures. As such, motivating a high absolute level TSR performance through these contracts seem unnatural.

Finally, we note the following additional evidence which suggests that the implementation of performance share contracts and the selection of peers are consistent with the removal of common shocks. Comparing the solutions to the RPE problem (where we do not perfectly observe the common shock) to that of a contract that rewards based on absolute firm performance, one can deduce that RPE contracts are optimal when $\sigma_c^2 > \sigma_b^2$. The intuition is that relative-performance contracts are preferred when the relative peformance measure ($p - \hat{c} = a + \omega_b + \epsilon$) is a less noisy reflection of effort compared to the absolute performance measure ($p = a + c + \epsilon$). In other words, if the choice of peer benchmarks induces more noise in the inference of managerial effort, then contracting on the absolute performance measure would be better. In untabulated results, we find that on average firms that use explicit RPE contracts have on average 40% lower variance (statistically significant at the 1% level) in the realized relative performance ($Var(p - \hat{c})$) compared to the firm's realized performance ($Var(p)$). This is true for both firms that choose index-based peers (who exhibit 30% lower variance on average and significant at the 1% level) and firms that choose specific peers (who exhibit 48% lower variance on average and again significant at the 1% level).

Also consistent with the removal of common shocks, we find that during our sample period, among the 26 firms that switched benchmark types (from index to specific or vice versa), 19 (or 73%) switched from index to specific peer benchmarks. Moreover, we find in un-tabuluated results that prior to the switch to specific peers, firms performed worse in measurement error variances relative to SBPs and nearly equally after the switch—consistent with our main cross-sectional results. Together, these results suggest that firms' implementation of performance share contracts are consistent with the removal of common shocks.

5.4 Reduced Form Performance Implications In the Cross Section

We conclude our analysis by examining whether there are observed performance differences between firms that choose index-based benchmarks versus those who choose specific peers. Recall that our analysis in Section 4 suggest an economically significant improvement in counterfactual performance that results from improving the adequacy of RPE benchmarks, in particular among those firms that chose index-based benchmarks. Table 9 reports reduced-form regressions of firm performance, in terms of ROA (columns 1-3) and annual stock returns (columns 4-6), on the choice of index-based benchmarks and the set of CEO, board, and firm characteristics examined above. Consistent with our earlier calibration exercise, we find evidence in our sample that the choice of index-based benchmarks is associated with worse performance. Interpreting column 3, which includes both year- and industry-fixed effects, the choice of index-based RPE benchmarks is associated with an 80 basis point decline in ROA, which is statistically significant at the 5% level. Relative to our sample median ROA of 440 basis points, this is nearly a 20% proportional decline. Similarly, interpreting column 6, which includes both year- and industry-fixed effects, the choice of index-based RPE benchmarks is associated with a 320 basis point decline in annual returns, which is statistically significant at the 5% level. Relative to the median annual return of 1,320 basis points, this represents an economically significant 24% proportional decline.

We caveat that this reduced-form analysis uses firms that use specific-peer-based benchmarks as counterfactuals to those firms that use index-based-benchmarks. To the extent that our control variables do not fully capture the differences in underlying characteristics that could be associated with firm performance between these two types of firms, it is possible that the estimated performance differences are overstated. Indeed, our reduced-form estimates are larger compared to the structural estimates of Table 5.

Overall, our analysis suggests that governance-related frictions best explains the empirical patterns in RPE benchmark inadequacy and the choice of index-based benchmarks. Furthermore, we find empirical (reduced-form) evidence that the choice of index-based benchmarks is associated with poorer realized firm performance in our sample, consistent with our earlier structural estimation approach.

6 Conclusion

The explicit usage of relative performance based grants has steadily increased in recent years. The common wisdom behind such grants is that they help induce costly effort from the CEO. We document that firms which choose index-based benchmarks perform 14% worse relative to a normative state-of-the-art performance peer set in their ability to explain the time-series variation in returns, and at least 16.4% greater in the variance of measurement errors in the common component of performance. Firms which elect to use specific peers have at least 4.4% greater variance in the measurement errors.

Based on structural estimates from a principal-agent model, these poorer benchmarks imply, absent frictions, an average performance penalty in annual returns of 60 to 153 (106 to 277) basis points through their effect on the manager's effort across all (index-based benchmarking) firms. In the cross section, index-based benchmarking firms also have lower realized annual ROA (80 bps) and stock returns (320 bps).

Our conclusion is that boards do not perfectly filter out exogenous shocks to firm performance in setting compensation despite having explicit relative performance evaluation. The source of benchmarking inefficacy is rooted in the board's reliance on an index benchmark in lieu of specific peers and is inconsistent with the standard principal-agent framework and other alternative theories designed to rationalize inefficacious benchmarking. Instead, we provide evidence that governance weaknesses in the firm are associated with poorer benchmarks and the decision to choose an index-based benchmark.

We caveat that this set of findings is restricted to the sample of firms that disclose their relative TSR performance contract specifications. Nevertheless, as more firms adopt such practices, we are hopeful that the findings and techniques we develop here will help guide future research in this area.

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.

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

Ai, H., Kiku, D. and Li, R. (2016). Quantifying the impact of moral hazard: Evidence from a structural estimation. *SSRN Working Paper 2801116*.

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

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.

Antón, M., Ederer, F., Giné, M. and Schmalz, M. C. (2016). Common ownership, competition, and top management incentives. *SSRN Working Paper 2802332*.

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources*, **37** (2), 728–751.

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

Bell, B. and Van Reenen, J. (2016). CEO pay and the rise of relative performance contracts: A question of governance? *National Bureau of Economic Research Working Paper*.

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

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*.

Bizjak, J. M., Hayes, R. M. and Kalpathy, S. L. (2015). Performance-contingent executive compensation and managerial behavior. *SSRN Working Paper 2519246*.

—, 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*.

—, Lemmon, M. L. and Naveen, L. (2008). Does the use of peer groups contribute to higher pay and less efficient compensation? *Journal of Financial Economics*, **90** (2), 152–168.

Cadman, B. and Carter, M. E. (2013). Compensation peer groups and their relation with CEO pay. *Journal of Management Accounting Research*, **26** (1), 57–82.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The Economic Journal*, **101** (405), 157–179.

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, **105** (2), 536–563.

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

Celentani, M. and Loveira, R. (2006). A simple explanation of the relative performance evaluation puzzle. *Review of Economic Dynamics*, **9** (3), 525–540.

Chassang, S. (2013). Calibrated incentive contracts. *Econometrica*, **81** (5), 1935–1971.

Core, J. E., Holthausen, R. W. and Larcker, D. F. (1999). Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics*, **51** (3), 371–406.

Creighton, A. (2016). Nobel prize in economics: Who are oliver hart and bengt holmström. *Wall Street Journal*.

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

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

— and Kaniel, R. (2016). Relative pay for non-relative performance: Keeping up with the joneses with optimal contracts. *SSRN Working Paper 2755856*.

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

Drake, K. D. and Martin, M. (2016). Implementing relative performance evaluation: The role of life cycle peers. *SSRN Working Paper 2822388*.

Engel, E., Hayes, R. M. and Wang, X. (2003). CEO turnover and properties of accounting information. *Journal of Accounting and Economics*, **36** (1), 197–226.

Faulkender, M. and Yang, J. (2010). Inside the black box: The role and composition of compensation peer groups. *Journal of Financial Economics*, **96** (2), 257–270.

Feltham, G. A. and Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *Accounting Review*, **69** (3), 429–453.

Feriozzi, F. (2011). Paying for observable luck. *RAND Journal of Economics*, **42** (2), 387–415.

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

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

Gallani, S. (2015). Through the grapevine: Network effects on the design of executive compensation contracts. *SSRN Working Paper 2536477*.

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.

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

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

Ghosh, M. and John, G. (2000). Experimental evidence for agency models of salesforce compensation. *Marketing Science*, **19** (4), 348–365.

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

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.

Gopalan, R., Milbourn, T. and Song, F. (2010). Strategic flexibility and the optimality of pay for sector performance. *Review of Financial Studies*, **23** (5), 2060–2098.

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

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

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

Himmelberg, C. P. and Hubbard, R. G. (2000). Incentive pay and the market for CEOs: An analysis of pay-for-performance sensitivity. In *Tuck-JFE Contemporary Corporate Governance Conference*.

Hoffmann, F. and Pfeil, S. (2010). Reward for luck in a dynamic agency model. *Review of Financial Studies*, **23** (9), 3329–3345.

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

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

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

— and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, **7**, 24–52.

Hurwicz, L. and Shapiro, L. (1978). Incentive structures maximizing residual gain under incomplete information. *The Bell Journal of Economics*, **9** (1), 180–191.

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.

Jayaraman, S., Milbourn, T. T. and Seo, H. (2015). Product market peers and relative performance evaluation. *SSRN Working Paper 2632153*.

Jensen, M. C. and Murphy, K. J. (1990). Performance pay and top-management incentives. *Journal of Political Economy*, **98** (2), 225–264.

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

Jin, L. (2002). CEO compensation, diversification, and incentives. *Journal of Financial Economics*, **66** (1), 29–63.

Joh, S. W. (1999). Strategic managerial incentive compensation in japan: Relative performance evaluation and product market collusion. *Review of Economics and Statistics*, **81** (2), 303–313.

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

Kleinberg, J., Liang, A. and Mullainathan, S. (2015). The theory is predictive, but is it complete? an application to human perception of randomness. *Harvard University Working Paper*.

Lazear, E. (2000). Performance pay and productivity. *The American Economic Review*, **90** (5), 1346–1361.

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.

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

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

Li, Z. and Wang, L. (2016). Executive compensation incentives contingent on long-term accounting performance. *Review of Financial Studies*, **29** (6), 1586–1633.

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

Murphy, K. J. and Zimmerman, J. L. (1993). Financial performance surrounding CEO turnover. *Journal of Accounting and Economics*, **16** (1), 273–315.

Nam, J. (2016). Financial reporting comparability and relative performance evaluation. *Hong Kong Polytechnic University Working Paper*.

Oyer, P. (2004). Why do firms use incentives that have no incentive effects? *Journal of Finance*, **59** (4), 1619–1650.

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

Rajgopal, S., Shevlin, T. and Zamora, V. (2006). CEOs' outside employment opportunities and the lack of relative performance evaluation in compensation contracts. *Journal of Finance*, **61** (4), 1813–1844.

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

Rubin, J. and Sheremeta, R. (2015). Principal–agent settings with random shocks. *Management Science*, **62** (4), 985–999.

Schaefer, S. (1998). The dependence of pay-performance sensitivity on the size of the firm. *Review of Economics and Statistics*, **80** (3), 436–443.

Schmidt, K. M. (1997). Managerial incentives and product market competition. *Review of Economic Studies*, **64** (2), 191–213.

Tice, F. M. (2015). The effect of relative performance evaluation on investment efficiency and firm performance. *Working Paper*.

Vrettos, D. (2013). Are relative performance measures in CEO incentive contracts used for risk reduction and/or for strategic interaction? *Accounting Review*, **88** (6), 2179–2212.

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

Warner, J. B., Watts, R. L. and Wruck, K. H. (1988). Stock prices and top management changes. *Journal of Financial Economics*, **20**, 461–492.

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

Table A.I.
Sample Selection

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

Main Sample Selection	Firm-year Observations	Firm-year-month Observations	Unique Firms
(1) Firms in ISS Incentive Lab with CEO grant data for 2006 to 2013	9,615		1,542
(2) Less firms without CEO grants based on an RPE component	(6,829)		
	2,786		707
(3) Less firms for which the relative benchmark cannot be identified	(462)		
	2,324		621
(4) Less firms that do not use stock price as the relevant RPE performance measure	(446)		
	1,878		532
(5) Less firms without CIK-GVKEY matches	(57)		
	1,821		487
(6) Merged with monthly return data from CRSP		21,710	
(7) Less observations with missing SBP data	(685)	(7,418)	(83)
(8) Less observations that have either use index or specific peers in a given fiscal year	(85)	(1,107)	(35)
(9) Less observations with fewer than 10 monthly returns in the time-series regressions	(13)	(77)	(13)
Final Sample	1,038	13,108	356

Table A.II.
Descriptive Statistics

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

Variables are defined as follows, where the variable names from the relevant databases are reported in brackets. Using Compustat, we define the following variables on firm characteristics: *ROA* is the ratio of net income to total assets [ni/at]; *Log Market Cap* is the log of the firm's market capitalization ($Millions) as of the fiscal year end [mkvalt]; and *Book-to-Market* is the book value of common equity ($Millions) [ceq] divided by the market capitalization ($Millions) [mkvalt]. Using Execucomp, we define the following variables on CEO characteristics: *Log CEO Pay* is the log of the CEO's total compensation (in $Thousands) [tdc1]; *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 companies and directorships databases, we define the following variables on board characteristics: *% Busy Directors* is the percentage of the firm's directors with more than four board seats on public firms; *Board Size* is the number of directors on the board; and *Director Workload* is the number of full board meetings held over the prior fiscal year [BDMTGS] divided by the number of directors. Using the ISS Governance database, we define the following variables on a firm's governance characteristics: *Staggered Board* is a variable which equals 1 if the firm holds staggered director elections and 0 if the firm has a unitary board; *Dual-Class Shares* is an indicator variable which equals 1 if the firm has multiple classes of voting shares and 0 if it has a single class of voting shares. Using CRSP, we define the following variables on a firm's stock return characteristics: *Annual Returns* is a firm's annual CRSP cum-dividend returns [ret] over a given fiscal year; and *Return Volatility* is the standard deviation of monthly cum-dividend returns for a firm over the fiscal year. Finally, *Index* is a dummy variable which equals 1 if the firm uses an index as its RPE benchmark in a given fiscal year. Using search-based-peers (Lee *et al.*, 2015), $(R^2_{rpe} - R^2_{sp500})/(R^2_{sbp} - R^2_{sp500})$ is defined to be the performance of the RPE benchmarks in terms of times-series R^2 relative to that generated by the S&P500 index scaled by the difference in R^2 generated by SBPs and the S&P500. In the context of the model, $(\sigma^2_{sp500} - \sigma^2_{rpe})/(\sigma^2_{sp500} - \sigma^2_{sbp})$ is defined to be the performance of the RPE benchmarks in terms of times-series measurement error σ^2 relative to that generated by the S&P500 index scaled by the difference in σ^2 generated by the S&P500 and SBPs.

Panel A: Distributional Statistics

	Obs	Mean	Std. Dev.	P25	Median	P75
ROA	1,189	0.049	0.053	0.024	0.044	0.077
Annual Returns	1,189	0.143	0.327	-0.046	0.132	0.313
$(R^2_{rpe} - R^2_{sp500})/(R^2_{sbp} - R^2_{sp500})$	976	1.179	2.385	0.370	0.873	1.088
$(\sigma^2_{sp500} - \sigma^2_{rpe})/(\sigma^2_{sp500} - \sigma^2_{sbp})$	954	0.825	0.942	0.306	0.888	1.060
Index	1,189	0.346	0.476	0.000	0.000	1.000
Log CEO Pay	1,189	8.893	0.678	8.462	8.874	9.351
CEO Tenure	1,189	5.385	4.605	2.000	4.000	7.000
CEO Age	1,189	56.453	5.190	53.000	57.000	60.000
% Busy Directors	1,189	0.022	0.048	0.000	0.000	0.000
Board Size	1,189	10.611	2.057	9.000	11.000	12.000
Director Workload	1,189	0.805	0.349	0.583	0.727	0.929
Log Market Cap	1,189	9.032	1.267	8.123	8.876	9.751
Return Volatility	1,189	0.079	0.047	0.047	0.068	0.098
Book-to-Market	1,189	0.517	0.312	0.299	0.482	0.681
Staggered Board	1,189	0.349	0.477	0.000	0.000	1.000
Dual-Class Shares	1,189	0.030	0.171	0.000	0.000	0.000

Table A.II.
Continued

Panel B: Correlation Matrix

ROA	1.00							
Annual Returns	0.15***	1.00						
$(R^2_{rpe} - R^2_{sp500})/(R^2_{sbp} - R^2_{sp500})$	0.03	−0.00	1.00					
$(\sigma^2_{sp500} - \sigma^2_{rpe})/(\sigma^2_{sp500} - \sigma^2_{sbp})$	−0.03	0.01	0.51***	1.00				
Index	−0.04	−0.00	−0.17***	−0.24***	1.00			
Log CEO Pay	0.18***	0.06**	−0.03	−0.14***	0.13***	1.00		
CEO Tenure	0.02	0.02	−0.05*	−0.00	0.07**	0.01	1.00	
CEO Age	0.07**	0.02	0.05	0.02	0.03	0.06**	0.37***	1.00
% Busy Directors	0.01	0.03	−0.08**	0.00	0.06**	0.06**	−0.03	−0.06**
Board Size	−0.01	−0.02	−0.01	−0.01	0.07**	0.24***	−0.08***	0.05*
Director Workload	−0.08***	0.01	−0.02	−0.06*	0.08***	0.00	−0.06**	−0.09***
Log Market Cap	0.31***	0.12***	−0.00	−0.06*	0.04	0.68***	−0.05*	0.06*
Return Volatility	−0.26***	−0.09***	0.00	−0.09***	−0.01	−0.09***	0.04	−0.08***
Book-to-Market	−0.45***	−0.30***	0.06*	0.08**	−0.08***	−0.15***	−0.02	0.03
Staggered Board	0.03	−0.04	0.05	−0.06*	0.01	−0.13***	−0.00	0.02
Dual-Class Shares	0.13***	0.00	0.12***	−0.09***	0.01	0.11***	−0.04	0.01

% Busy Directors	1.00							
Board Size	0.05*	1.00						
Director Workload	0.05*	−0.35***	1.00					
Log Market Cap	0.06**	0.44***	−0.09***	1.00				
Return Volatility	0.02	−0.21***	0.12***	−0.34***	1.00			
Book-to-Market	−0.07**	0.00	0.09***	−0.25***	0.23***	1.00		
Staggered Board	−0.03	−0.12***	−0.01	−0.19***	0.08***	0.03	1.00	
Dual-Class Shares	−0.02	0.09***	−0.06**	0.07**	−0.03	−0.11***	−0.02	1.00

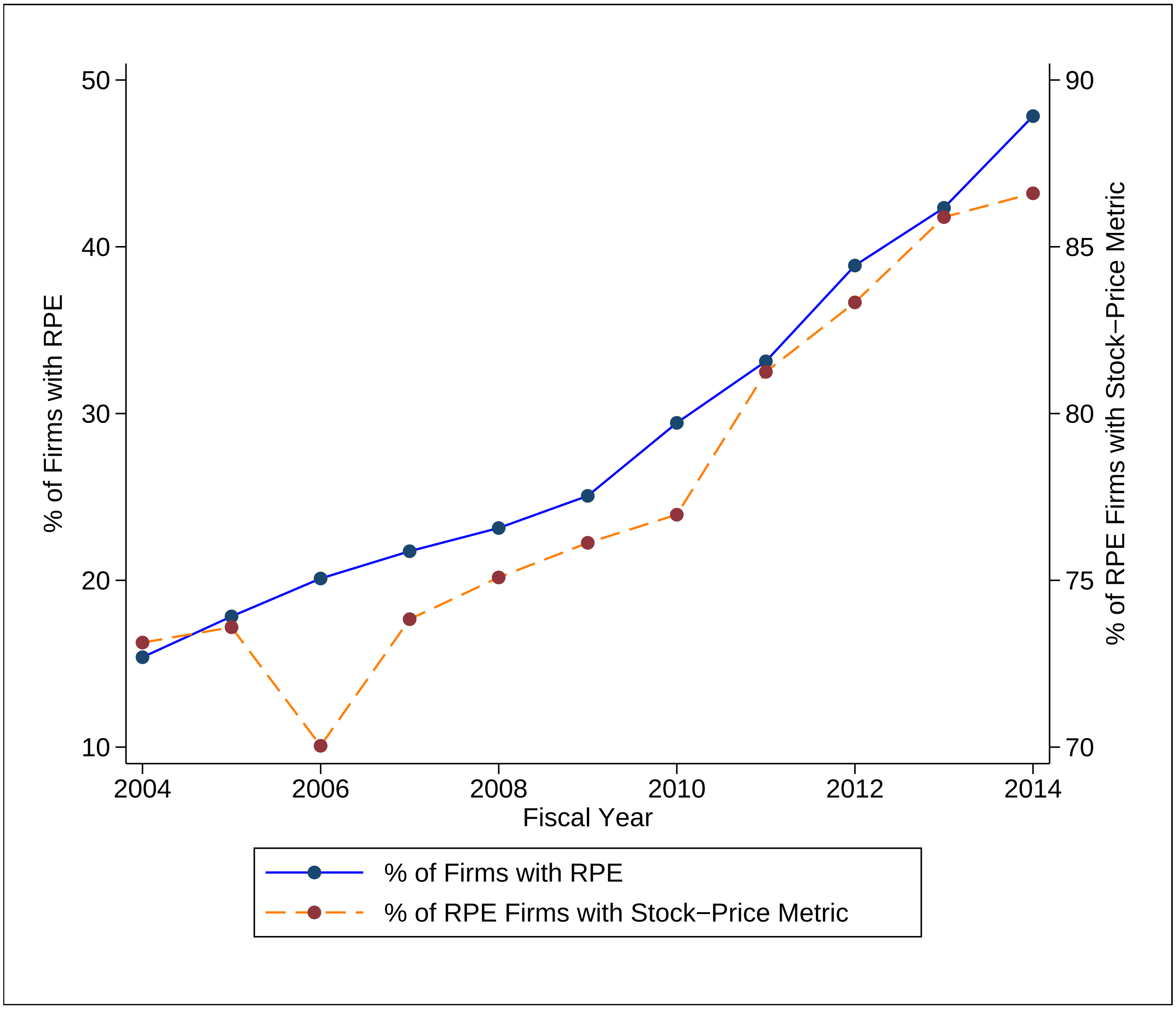

Figure 1. Fraction of Firms Using Relative Performance Contracts Between 2004-2014
This figure plots in the solid line a) the fraction of firms in the sample who disclose the awarding of relative performance grants (RPE firms) and in dotted lines b) the fraction of these RPE firms which use stock price as the metric of relative performance.

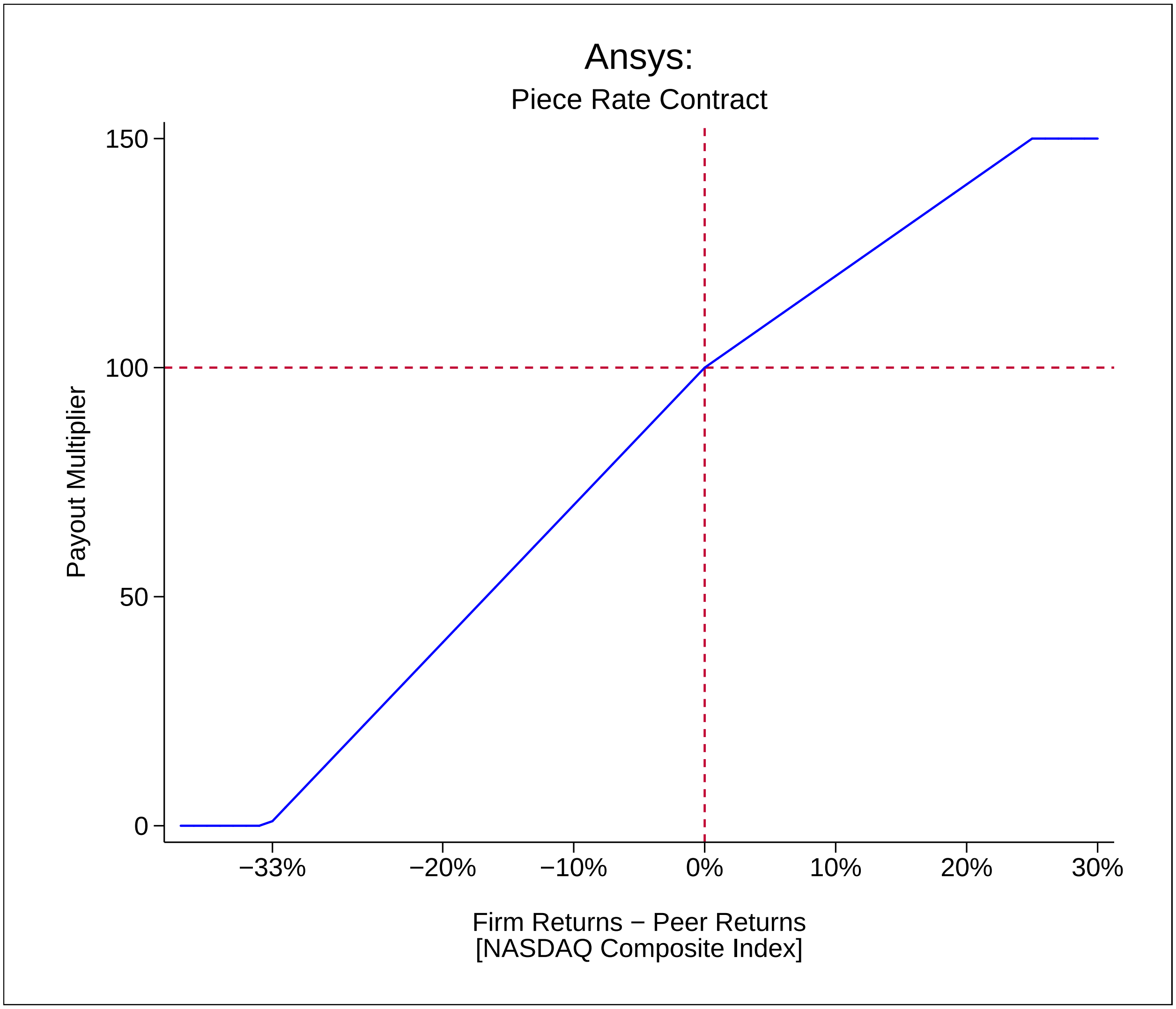

Figure 2. CEO Relative Performance Contract: Ansys Inc
This figure plots the contract payout multiplier for Ansys Corp in fiscal year 2014 as a function of own firm stock returns and the peer index (NASDAQ Composite) stock returns under the assumption that firm's stock returns are 40%. The payout multiplier X as a function of own performance p and peer performance c when $p \geq 0$ is:

$$X(p, c) = \begin{cases} \min\,[100 + 2(p - c), 150] & \text{if } p \geq c \\ \max\,[100 + 3(p - c), 0] & \text{if } p < c \end{cases}$$

Complete details are available in the proxy statement on page 42 of http://investors.ansys.com/~/media/Files/A/Ansys-IR/documents/ansys-2016-proxy-statement.pdf.

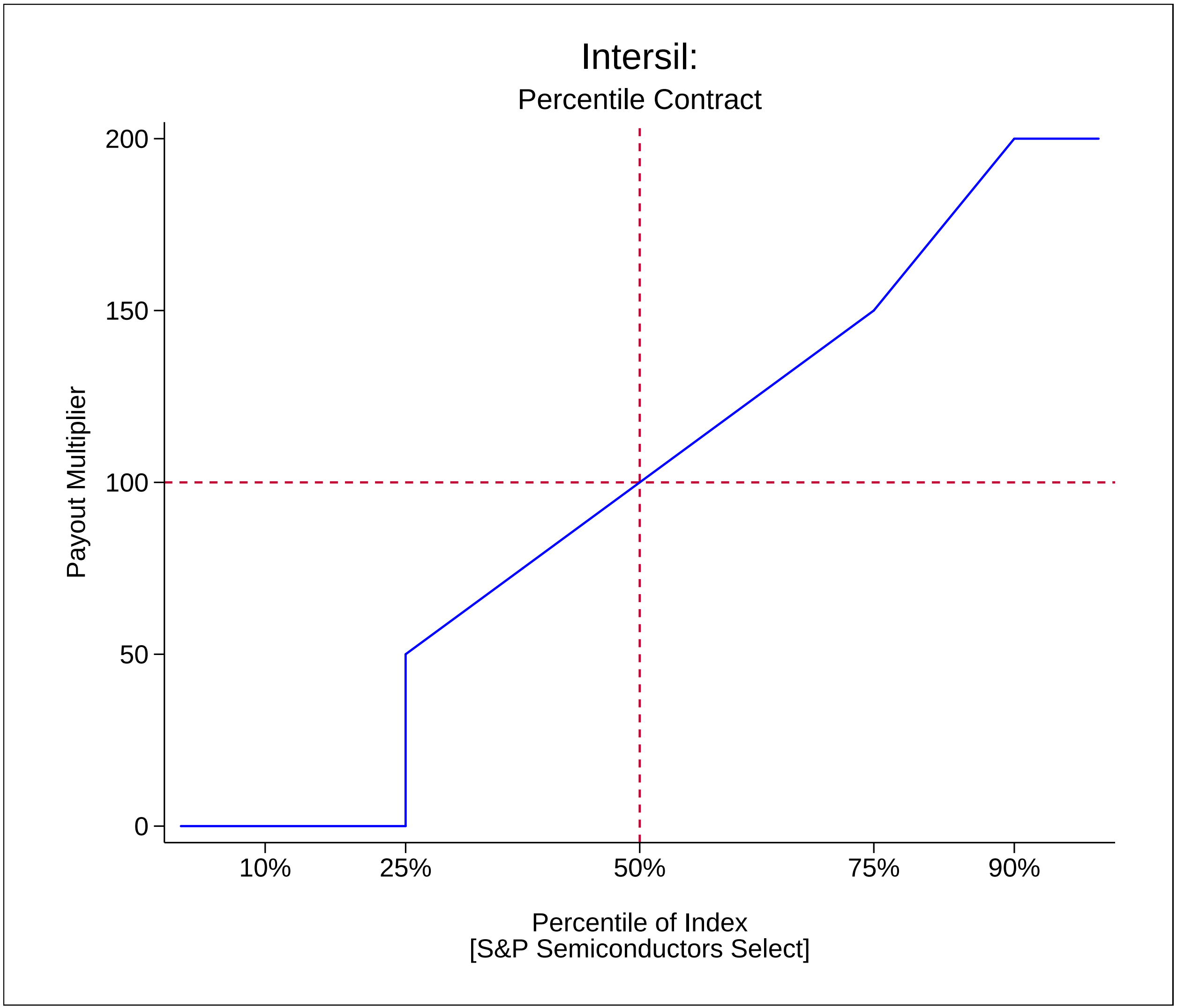

Figure 3. CEO Relative Performance Contract: Intersil Corp
This figure plots the contract payout multiplier for Intersil Corp in fiscal year 2014 as a function of own stock returns in percentiles relative to the constituents of the S&P Semiconductor Select Index. The payout multiplier X as a function of the firm's percentile in returns p relative to the index when $p \geq 0$ is:

$$X(p) = \begin{cases} 0 & \text{if } p < 25\% \\ 25 + 2p & \text{if } p \in [25\%, 75\%] \\ \min\left[150 + \frac{4}{3}p, 200\right] & \text{if } p > 75\% \end{cases}$$

Complete details are available in the proxy statement on page 24 of https://www.sec.gov/Archives/edgar/data/1096325/000109632515000012/isil-20150305xdef14a.htm.

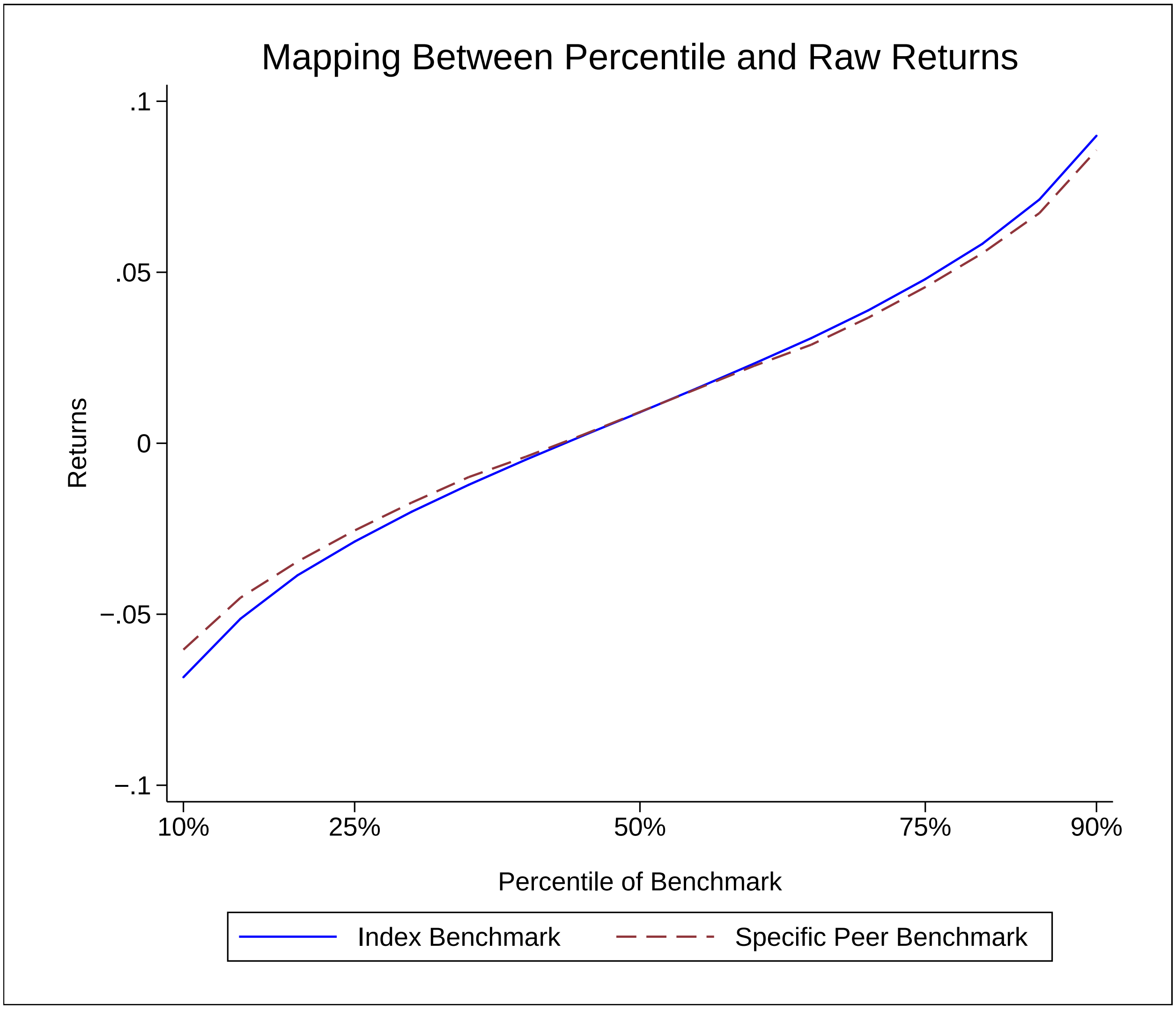

Figure 4. Mapping Between Percentile Performance and Raw Returns
This figure plot for the firms (index-based or specific peers based benchmarking) in our whole sample the relationship between monthly stock returns and the percentile of performance relative to the peer distribution using return data during our main sample period.

Table 1.
Summary Statistics on CEO Grants

Panel A of this table reports summary statistics for compensation grants awarded to the CEO in fiscal years 2004 to 2014. 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 (i.e., cash, option, or stock) out of 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) out of 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 an RPE component and a stock price-based RPE component respectively.

			Payout Type			**Evaluation Type**			
Fiscal Year	Unique # of Firms	Unique # of Grants	Cash	Option	Stock	Abs	Abs/Rel	Rel	Time
Panel A: All CEO Grants									
2004	1,306	2.28	0.23	0.48	0.29	0.25	0.04	0.03	0.68
2005	1,295	2.37	0.25	0.42	0.33	0.29	0.05	0.03	0.63
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 RPE Component									
2004	201	1.22	0.44	0.01	0.55	-	0.56	0.44	-
2005	231	1.18	0.39	0.01	0.60	-	0.56	0.44	-
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 Stock Price-Based RPE Component									
2004	147	1.12	0.33	0.00	0.67	-	0.49	0.51	-
2005	170	1.08	0.29	0.01	0.70	-	0.49	0.51	-
2006	180	1.18	0.24	0.02	0.73	-	0.49	0.51	-
2007	206	1.18	0.27	0.01	0.72	-	0.50	0.50	-
2008	217	1.18	0.20	0.01	0.79	-	0.49	0.51	-
2009	220	1.21	0.22	0.01	0.77	-	0.48	0.52	-
2010	264	1.18	0.19	0.00	0.81	-	0.47	0.53	-
2011	312	1.17	0.16	0.00	0.83	-	0.47	0.53	-
2012	380	1.17	0.15	0.01	0.84	-	0.53	0.47	-
2013	420	1.13	0.13	0.00	0.86	-	0.57	0.43	-
2014	459	1.18	0.12	0.00	0.88	-	0.62	0.38	-

Table 2.
Importance of CEO RPE Incentives

This table reports summary statistics on the RPE incentive ratio and the RPE-stock incentive ratio of compensation grants awarded to the CEO for fiscal years 2004 to 2014. The RPE incentive ratio measures the incremental potential incentive when the CEO meets all her RPE-based performance targets and is calculated as (expected compensation if all targets are met)/(expected compensation if all other targets excluding RPE metric-based targets are met). The RPE-stock incentive ratio measures the incremental potential incentive when the CEO meets all her RPE-based stock price targets and is calculated as (expected compensation if all targets are met)/(expected compensation if all other targets excluding RPE stock price-based targets are met). The amount of expected compensation is calculated using the values reported in the "Grants of Plan-Based Awards Table" in the proxy statement. Specifically, it is computed by adding the target dollar value of the "Estimated Future Payouts Under Non-Equity Incentive Plan Awards" and the "Grant Date Fair Value of Stock and Option Awards." For grants that use multiple performance metrics, we calculate the weighted portion of expected compensation that corresponds to each performance metric. We assume that each performance metric is weighted equally in the elicitation of the grant. Column 3 reports the average expected compensation. Columns 4 and 5 report the expected compensation amount that is attributable to RPE-based metrics and RPE-stock price-based metrics, respectively. Columns 6 and 7 report the accompanying incentive ratios.

		Expected Compensation			**Incentive Ratios**	
Fiscal Year	Unique # of Firms	Total	RPE	RPE & Stock-Based	RPE	RPE & Stock-Based
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2004	91	2,255,182	1,515,470	1,055,978	2.09	1.41
2005	113	2,300,872	1,602,096	1,198,129	2.04	1.45
2006	235	2,595,754	1,768,730	1,233,457	2.45	1.53
2007	265	2,804,220	1,962,250	1,369,168	2.21	1.49
2008	277	2,906,259	2,050,360	1,563,382	2.24	1.54
2009	280	2,480,505	1,787,294	1,336,082	2.32	1.54
2010	332	2,823,516	1,898,931	1,404,596	2.21	1.55
2011	377	2,751,953	1,820,329	1,389,227	2.13	1.65
2012	446	3,204,148	2,077,538	1,596,345	2.21	1.75
2013	478	3,244,364	2,082,448	1,650,666	2.08	1.72
2014	524	3,286,170	2,005,443	1,505,951	2.07	1.74

Table 3.
Summary Statistics of Relative Performance Benchmark Types

This table summarizes the percentage of RPE grants that are associated with different types of relative benchmarks for fiscal years 2006 to 2013. Columns 2 and 3 report the unique number of firms and grants respectively. Columns 4-8 report the percentage of RPE grants which use each particular type of benchmark: i.e., Specific Peers, the S&P500 index, the S&P1500 index, other indexes (generally industry-based), or unspecified. Because each grant can be associated with multiples types of benchmarks, the percentages across the columns 4 through 8 can exceed one. Finally, in column 9 we report the average number of chosen peer firms for RPE grants associated with specific peers as the benchmark.

			Relative Benchmark Type					
Fiscal Year	Unique # of Firms	Unique # of Grants	Specific Peer	S&P500	S&P1500	Other Index	Not Specified	# of Peers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2006	257	313	0.49	0.14	0.05	0.17	0.20	15.43
2007	279	355	0.55	0.16	0.05	0.14	0.14	16.20
2008	289	358	0.57	0.18	0.04	0.17	0.10	17.19
2009	289	373	0.55	0.17	0.05	0.13	0.15	17.16
2010	343	424	0.56	0.14	0.04	0.15	0.17	17.56
2011	384	471	0.55	0.15	0.05	0.15	0.16	18.74
2012	456	579	0.51	0.15	0.04	0.16	0.18	17.81
2013	489	596	0.49	0.18	0.05	0.17	0.15	18.26

Table 4.
Explaining Common Component of Returns: Firm Benchmarks vs. S&P500 and SBPs

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

$$R_t = \alpha_t + \beta_t R_{p_t} + \epsilon_t$$

using CRSP monthly returns data. In Panel A, Columns 1, 2, and 3 report across-firm average R^2s from time-series regressions, regressing base firm i's returns on the concurrent returns of a portfolio of peers. Column 1 uses the returns of the S&P500 index, column 2 uses the median returns of firms' chosen performance benchmarks, and column 3 uses the mean returns of search-based peers (Lee *et al.*, 2015). Column 4 and 5 report the differences in R^2 values reported in columns 2 and 1 (3 and 2), and column 6 reports the average number of observations per firm.

The results are reported for the sample of base firms for whom firms' chosen benchmarks are identifiable from ISS Incentive Lab. We use return data from 2006 to 2013 and include those firms for which there are at least 10 observations. The first row contains all firms in our sample that satisfy these filters; the second row estimates the same regressions on the subset of observations that select specific peers as benchmarks; the last row estimates the same regressions on the subset of observations that select a stock index as a benchmark.

In Panel B, we reproduce the results of the sub-sample for specific peers but use two alternative measures of peer performance in lieu of the median. The first is the mean of peer performance and the second is the 75^{th} percentile of peer performance. To facilitate comparisons, all the regressions are conducted using the same underlying set of firms. The reported N in parentheses represents the number of firms-benchmark combinations contained in each sample, and standard errors are reported in square brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

Sample	S&P500 (1)	RPE Benchmarks (2)	SBP (3)	(2)-(1) (4)	(3)-(2) (5)	Mean Obs per Firm (6)
Panel A: Baseline						
All	0.328***	0.483***	0.518***	0.155***	0.035***	36.820
(N= 356)	[0.010]	[0.012]	[0.012]	[0.011]	[0.008]	
Specific Peers	0.328***	0.548***	0.560***	0.220***	0.013	40.552
(N= 201)	[0.012]	[0.015]	[0.015]	[0.014]	[0.009]	
Index	0.329***	0.400***	0.464***	0.071***	0.064***	31.981
(N= 155)	[0.017]	[0.019]	[0.019]	[0.013]	[0.015]	
Panel B: Alternative Peer Performance Measures						
Specific Peers$_{mean}$	0.328***	0.547***	0.560***	0.219***	0.013	40.552
(N= 201)	[0.012]	[0.015]	[0.015]	[0.014]	[0.009]	
Specific Peers$_{75th}$	0.328***	0.517***	0.560***	0.189***	0.044***	40.552
(N= 201)	[0.012]	[0.015]	[0.015]	[0.014]	[0.010]	

Table 5.
Estimating Measurement Error Variances in Common Shock Filtration: Firm Benchmarks vs. S&P500 and SBPs

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

The first four rows in Panel A report the individual parameter estimates of the measurement error variances. Panel B reports the differences in the measurement error variances of the firms' chosen benchmark relative to two normative benchmarks 1) S&P500 and 2) SBPs. Finally, Panel C reports in annualized basis points the performance implications of using the firms' chosen benchmarks relative to the two normative benchmarks as described in Eqn. 12.

Column 1 reports estimates from the entire sample with the same sampling criterion as Table 4. Column 2 reports estimates for the sub-sample of observations where the base firm chooses specific firms as the RPE benchmark. Finally, column 3 reports estimates for the sub-sample of observations where the base firm chooses an index (e.g. S&P500, S&P1500, or others) as their RPE benchmark. Standard errors are reported in square brackets and calculated via the delta method where appropriate. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

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

Table 6.
Robustness: Estimating Measurement Error Variances

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

The first rows in Panel A report the individual parameter estimates of the measurement error variances. In Panel B, we re-produce the results for the specific peer sub-sample, but use two alternative measures of peer performance in lieu of the median. The first is the mean of peer performance and the second is the 75^{th} percentile of peer performance. Finally, Panel C reports the differences in the measurement error variances of the firms' chosen benchmark relative to SBPs for each measure of peer performance (mean, median, and 75^{th} percentile).

Column 1 reports estimates from the entire sample with the same sampling criterion as Table 4. Column 2 reports estimates for the sub-sample of observations where the base firm chooses specific firms as the RPE benchmark. Finally, column 3 reports estimates for the sub-sample of observations where the base firm chooses an index (e.g. S&P500, S&P1500, or others) as their RPE benchmark. The reported N at the bottom represents the number of firms-benchmark combinations contained in each sample. Standard errors are reported in square brackets and calculated via the delta method where appropriate. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	All Firms	Specific Peers	Index Peers
	(1)	(2)	(3)
Panel A: Average of Firm Level Estimates			
$\sigma^2_{b,sp500} + \sigma^2$	63.042***	68.567***	55.878***
	[4.503]	[5.900]	[6.940]
$\sigma^2_{b,firm,median} + \sigma^2$	46.878***	43.416***	51.367***
	[3.622]	[3.715]	[6.781]
$\sigma^2_{b,sbp} + \sigma^2$	41.344***	41.150***	41.596***
	[2.593]	[3.359]	[4.073]
Panel B: Alternative Peer Performance Measures			
$\sigma^2_{b,firm,mean} + \sigma^2$	47.192***	43.973***	51.367***
	[3.623]	[3.722]	[6.781]
$\sigma^2_{b,firm,p75} + \sigma^2$	50.187***	49.277***	51.367***
	[3.831]	[4.339]	[6.781]
Panel C: Δ Estimates			
$\sigma^2_{b,firm,med} - \sigma^2_{b,sbp}$	5.534***	2.266	9.771**
	[2.048]	[1.671]	[4.158]
$\sigma^2_{b,firm,mean} - \sigma^2_{b,sbp}$	5.848***	2.823**	9.771**
	[1.930]	[1.159]	[4.158]
$\sigma^2_{b,firm,p75} - \sigma^2_{b,sbp}$	8.843***	8.127***	9.771**
	[2.095]	[1.878]	[4.158]
N	356	201	155

Table 7.
Explaining the Adequacy of Benchmark Selection

This table reports results from OLS regressions of the adequacy of a firm's choice of RPE benchmarks on CEO, board of directors, and firm characteristics. We consider two measures of RPE benchmark adequacy. The first measure, reported in columns 1–3, is defined to be the performance of the RPE benchmarks in terms of times-series R^2 relative to that generated by the S&P500 index scaled by the difference in R^2 generated by SBPs and the S&P500. Since this measure treats the S&P500 and SBPs as the lower and upper bounds of R^2, our analysis focuses on the subset for which $R^2_{sbp} - R^2_{sp500} > 0$. The second measure, reported in columns 4–6, is defined to be the performance of the RPE benchmarks in terms of times-series σ^2 relative to that generated by the S&P500 index scaled by the difference in σ^2 generated by the S&P500 and SBPs. Since this measure treats the S&P500 and SBPs as the upper and lower bounds of σ^2, our analysis focuses on the subset for which $\sigma^2_{sp500} - R^2_{sbp} > 0$. Observations are at the annual firm-benchmark level and all variables are defined in Table A.II. Industry-fixed effects use 2-digit Global Industry Classification Standard definitions. Standard errors are clustered at the firm level and reported below the point estimates in brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	$(R^2_{rpe} - R^2_{sp500})/(R^2_{sbp} - R^2_{sp500})$			$(\sigma^2_{sp500} - \sigma^2_{rpe})/(\sigma^2_{sp500} - \sigma^2_{sbp})$		
	(1)	(2)	(3)	(4)	(5)	(6)
Index	-0.823***	-0.827***	-1.296***	-0.446**	-0.449**	-0.395**
	[0.296]	[0.296]	[0.437]	[0.187]	[0.186]	[0.197]
CEO Characteristics						
Log CEO Pay	-0.119	-0.085	-0.182	-0.163**	-0.153**	-0.106
	[0.277]	[0.288]	[0.237]	[0.069]	[0.068]	[0.071]
CEO Tenure	-0.034*	-0.033	-0.042*	0.001	0.001	0.003
	[0.021]	[0.021]	[0.024]	[0.014]	[0.013]	[0.012]
CEO Age	0.030	0.030	0.062**	0.006	0.006	0.004
	[0.020]	[0.020]	[0.029]	[0.010]	[0.010]	[0.008]
Board Characteristics						
% Busy Directors	-3.065*	-3.124**	-2.713*	0.553	0.595	0.414
	[1.608]	[1.561]	[1.468]	[0.850]	[0.815]	[0.800]
Board Size	-0.045	-0.049	-0.031	-0.004	-0.004	-0.022
	[0.085]	[0.086]	[0.061]	[0.016]	[0.017]	[0.023]
Director Workload	-0.073	-0.093	-0.028	-0.115	-0.114	-0.149
	[0.537]	[0.541]	[0.446]	[0.106]	[0.109]	[0.119]
Firm Characteristics						
Log Market Cap	0.073	0.062	0.228	-0.023	-0.028	-0.019
	[0.363]	[0.368]	[0.356]	[0.057]	[0.056]	[0.059]
Return Volatility	-0.042	-0.564	-0.847	-2.130**	-2.371**	-0.415
	[2.139]	[2.062]	[1.995]	[0.871]	[1.059]	[0.859]
Book-to-Market	0.447*	0.459*	0.518	0.195	0.195	0.181
	[0.244]	[0.237]	[0.316]	[0.143]	[0.150]	[0.133]
Staggered Board	0.288	0.262	0.492	-0.133	-0.133	-0.074
	[0.463]	[0.465]	[0.467]	[0.088]	[0.090]	[0.089]
Dual-Class Shares	2.041	2.048	2.025	-0.455	-0.443	-0.531**
	[2.239]	[2.238]	[1.846]	[0.353]	[0.356]	[0.262]
Time FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	Yes	No	No	Yes
Observations	980	980	980	958	958	958
Adjusted R^2	0.0488	0.0437	0.1672	0.0877	0.0843	0.1146

Table 8.
Explaining Selection of Benchmark Types

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

	$Pr(Index) = 1$		
	(1)	(2)	(3)
CEO Characteristics			
Log CEO Pay	0.146***	0.149***	0.101**
	[0.049]	[0.051]	[0.047]
CEO Tenure	0.007	0.007	0.007
	[0.005]	[0.005]	[0.005]
CEO Age	0.001	0.001	0.007
	[0.005]	[0.005]	[0.005]
Board Characteristics			
% Busy Directors	0.374	0.360	0.608
	[0.493]	[0.505]	[0.467]
Board Size	0.035**	0.035**	0.029**
	[0.014]	[0.014]	[0.014]
Director Workload	0.192***	0.193***	0.191***
	[0.068]	[0.068]	[0.073]
Firm Characteristics			
Log Market Cap	-0.065**	-0.066**	-0.019
	[0.032]	[0.033]	[0.033]
Return Volatility	-0.218	-0.207	0.182
	[0.471]	[0.557]	[0.584]
Book-to-Market	-0.155*	-0.152*	-0.051
	[0.088]	[0.089]	[0.093]
Staggered Board	0.043	0.044	0.053
	[0.059]	[0.060]	[0.062]
Dual-Class Shares	-0.048	-0.047	-0.023
	[0.115]	[0.115]	[0.117]
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
Observations	1,193	1,193	1,193
Pseudo R^2	0.0458	0.0468	0.1776

Table 9.
Performance Consequences of Inadequate Benchmarks

This table reports OLS regressions of firms' ROA (columns 1-3) and annual cum-dividend returns (columns 4-6) on an indicator of having picked an index as the RPE benchmark along with CEO, board of directors, and firm characteristics. Observations are at the annual firm-benchmark level and all variables are defined in Table A.II. Industry-fixed effects use 2-digit Global Industry Classification Standard definitions. Standard errors are clustered at the firm level and reported below the point estimates in brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	ROA			Annual Returns		
	(1)	(2)	(3)	(4)	(5)	(6)
Index	-0.007*	-0.007*	-0.008**	-0.020	-0.028*	-0.032**
	[0.004]	[0.004]	[0.004]	[0.018]	[0.015]	[0.016]
CEO Characteristics						
Log CEO Pay	-0.003	-0.002	-0.008*	-0.015	-0.012	-0.009
	[0.004]	[0.004]	[0.004]	[0.020]	[0.016]	[0.018]
CEO Tenure	0.000	0.000	0.000	0.001	-0.001	-0.001
	[0.000]	[0.000]	[0.000]	[0.002]	[0.002]	[0.002]
CEO Age	0.001	0.001*	0.001**	0.001	0.001	0.001
	[0.000]	[0.000]	[0.000]	[0.002]	[0.001]	[0.002]
Board Characteristics						
% Busy Directors	-0.000	-0.010	-0.016	0.047	0.098	0.132
	[0.035]	[0.034]	[0.033]	[0.236]	[0.209]	[0.210]
Board Size	-0.004***	-0.004***	-0.003***	-0.005	0.000	-0.000
	[0.001]	[0.001]	[0.001]	[0.007]	[0.005]	[0.006]
Director Workload	-0.007	-0.008	-0.004	0.031	0.054**	0.058**
	[0.005]	[0.005]	[0.004]	[0.030]	[0.025]	[0.026]
Firm Characteristics						
Log Market Cap	0.011***	0.011***	0.011***	0.022*	0.013	0.016
	[0.003]	[0.003]	[0.003]	[0.013]	[0.011]	[0.012]
Return Volatility	-0.121**	-0.124**	-0.232***	-0.045	0.535**	0.496
	[0.049]	[0.052]	[0.056]	[0.322]	[0.260]	[0.301]
Book-to-Market	-0.061***	-0.061***	-0.050***	-0.309***	-0.266***	-0.282***
	[0.008]	[0.008]	[0.008]	[0.040]	[0.036]	[0.037]
Staggered Board	0.007*	0.006	0.004	-0.010	0.004	0.005
	[0.004]	[0.004]	[0.003]	[0.019]	[0.016]	[0.016]
Dual-Class Shares	0.024***	0.023**	0.015	-0.051	-0.020	-0.017
	[0.009]	[0.010]	[0.010]	[0.062]	[0.040]	[0.039]
Time FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	Yes	No	No	Yes
Observations	1,193	1,193	1,193	1,193	1,193	1,193
Adj R^2	0.2866	0.2968	0.3486	0.0901	0.3762	0.3768