Product Market Peers and Relative Performance Evaluation

Sudarshan Jayaraman, Todd Milbourn, Florian Peters, and Hojun Seo

January 2018

Abstract

Relative Performance Evaluation (RPE) theory predicts that firms filter out common shocks (i.e., those affecting the firm and its peers) while evaluating CEO performance and that the extent of filtering increases with the number of firms in the peer group. Despite the intuitive appeal of the theory, previous tests of RPE find weak and inconsistent evidence. We hypothesize that one reason for the mixed evidence is the inaccurate classification of peers. Rather than using static, pre-defined Standard Industry Classifications (SIC), we exploit recent advances in textual analysis and define peers based on firms' product descriptions in their 10-K filings (Hoberg and Phillips, 2016). This alternative classification not only captures common shocks to firms' product markets more effectively but also tracks the evolving nature of these markets as 10-Ks are updated annually. Using product market peers, we find three pieces of evidence consistent with RPE in relation to CEO pay – (i) firms on average filter out common shocks to stock returns, (ii) the extent of filtering increases with the number of peers, and (iii) firms completely filter out common shocks in the presence of a large number of peers. We also examine forced CEO turnover decisions and find evidence consistent with RPE theory. Overall, our results suggest that a key identification strategy to testing RPE theory lies in accurately defining the peer group.

JEL codes: M40; M41; G30; J33
Keywords: Product market peers, Relative Performance Evaluation, CEO compensation, and Forced CEO Turnovers

---

Sudarshan Jayaraman: University of Rochester, e-mail: jayaraman@simon.rochester.edu. Todd Milbourn: Washington University in St. Louis, e-mail: milbourn@wustl.edu. Florian Peters: University of Amsterdam, e-mail: f.s.peters@uva.nl. Hojun Seo: National University of Singapore, e-mail:bizseoh@nus.edu.sg. We appreciate helpful comments from Dirk Black, Ettore Croci (discussant), Patricia Dechow, John Donovan, Richard Frankel, Radha Gopalan, Jerry Hoberg, Jared Jennings, Ranjani Krishnan, Ningzhong Li (discussant), Partha Mohanram, Kevin Murphy (discussant), Ram Natarajan, and workshop participants at 2015 Conference on Convergence of Financial and Managerial Accounting Research, Washington University in St. Louis, 2016 AFA annual meeting, 2016 MIT Asia Accounting Conference, and 2016 FMA conference. We thank Gerard Hoberg and Gordon Phillips for providing product market peers data. All errors remain our responsibility.

Product Market Peers and Relative Performance Evaluation

Abstract

Relative Performance Evaluation (RPE) theory predicts that firms filter out common shocks (i.e., those affecting the firm and its peers) while evaluating CEO performance and that the extent of filtering increases with the number of firms in the peer group. Despite the intuitive appeal of the theory, previous tests of RPE find weak and inconsistent evidence. We hypothesize that one reason for the mixed evidence is the inaccurate classification of peers. Rather than using static, pre-defined Standard Industry Classifications (SIC), we exploit recent advances in textual analysis and define peers based on firms' product descriptions in their 10-K filings (Hoberg and Phillips, 2016). This alternative classification not only captures common shocks to firms' product markets more effectively but also tracks the evolving nature of these markets as 10-Ks are updated annually. Using product market peers, we find three pieces of evidence consistent with RPE in relation to CEO pay – (i) firms on average filter out common shocks to stock returns, (ii) the extent of filtering increases with the number of peers, and (iii) firms completely filter out common shocks in the presence of a large number of peers. We also examine forced CEO turnover decisions and find evidence consistent with RPE theory. Overall, our results suggest that a key identification strategy to testing RPE theory lies in accurately defining the peer group.

JEL codes: M40; M41; G30; J33
Keywords: Product market peers, CEO compensation, Relative Performance Evaluation, and Forced CEO Turnovers

"…despite the obvious attractive features of relative performance evaluation, it is surprisingly absent from U.S. executive compensation practices. Why shareholders allow CEOs to ride bull markets to huge increases in their wealth is an open question…we view the weak evidence of relative performance evaluation as an important puzzle for executive compensation research."

Abowd and Kaplan (1999, pg. 157)
Journal of Economic Perspectives

One of the central tenets of agency theory is that increasing the "signal-to-noise" ratio of the performance measure reduces risk without compromising the level of incentive alignment (Holmstrom, 1979). In other words, eliminating sources of variation from firm performance, such as an industry-wide movement in stock returns that is beyond an individual manager's influence, results in a more efficient contract, i.e., one that achieves greater alignment without increasing risk. This is the idea behind Relative Performance Evaluation (RPE) where CEO compensation should not only be positively correlated with own firm performance, but also negatively correlated with industry-wide or market-wide performance to filter out uncontrollable common performance. RPE theory also predicts that this filtering should increase with the number of firms in the peer group, and in the limit, common performance should be completely filtered out from firm performance (Holmstrom, 1979, 1982; Gibbons and Murphy, 1990).

Despite the intuitive appeal of the theory, prior research documents weak and inconsistent empirical evidence on RPE and refers to this weak evidence as the RPE puzzle (e.g., Abowd and Kaplan, 1999; Frydman and Jenter, 2013; Jenter and Kanaan, 2015).[1] Motivated by the puzzle, prior studies attempt to document factors affecting the use of RPE in evaluating managerial performance. Albuquerque (2009) finds stronger RPE evidence in CEO pay when peers are composed of similar industry-size firms. Some prior studies attribute the RPE puzzle to managerial

[1] See also Antle and Smith (1986), Barro and Barro (1990), Gibbons and Murphy (1990), Janakiraman, Lambert and Larcker (1992), Prendergast (1999), DeFond and Park (1999), Bushman and Smith (2001), and Lambert (2001).

rent extraction (Bertrand and Mullainathan, 2001; Garvey and Milbourn, 2006), avoidance of self-destructive behavior (Gibbons and Murphy, 1990), and encouraging collusive behavior (Aggrawal and Samwick, 1999b). Also, studies such as Garvey and Milbourn (2003), Rajgopal, Shevlin, and Zamora (2006), Gopalan, Milbourn, and Song (2010), and Albuquerque (2014) attempt to find cross-sectional variations in the use of RPE.

In this study, we hypothesize that the prior lack of RPE evidence is (amongst other things) on account of how the peer group is defined. Most prior RPE studies use pre-defined industry classifications, such as the Standard Industry Classification (SIC), Global Industry Classification Standard (GICS) and North American Industry Classification System (NAICS) to identify peer firms.[2] However, we posit that those pre-defined industry groups are noisy proxies for the peer group (e.g., Clarke, 1989; Kahle and Walkling, 1996; Bhojraj et al., 2003; Dopuch et al., 2008; Brickley and Zimmerman, 2010; Guenther and Rosman, 1994).[3] Rather than using these static peer group definitions, we exploit recent advances in textual analysis and define similar peers based on firms' product descriptions in their 10-K filings (Hoberg and Phillips, 2016). We contend that this product-market based definition of peer groups better captures common shocks that affect the firm and its peers – which is what the theory purports to capture. An additional benefit is that these product-market based definitions reflect the changing nature of firms' business models as 10-Ks are updated annually (Hoberg and Phillips, 2016). Standard industry classifications, in contrast, fail to capture these dynamic aspects of product markets.

[2] See Table 1 of Albuquerque (2009) for a review of prior RPE studies and their use of peer group definitions. Among 15 empirical RPE studies conducted between 1986 and 2009, including Albuquerque (2009), 10 studies use SIC industries, 3 studies use market index, 1 study uses geographically close peers, and 1 study uses banking industry to identify peer firm performance.

[3] These pre-defined industry classifications focus on whether firms' *production processes* are similar rather than whether firms produce *similar products*. For example, "NAICS will be erected on a production-oriented, or supply-based, conceptual framework. This means that producing units that use identical or similar production processes will be grouped together in NAICS." http://www.naics.com/info.htm

To capture product-market peers, we use Hoberg and Phillips' (2016) Text-based Network Industry Classifications (TNIC) that are constructed based on firms' product similarities. Hoberg and Phillips (2016) calculate product similarity scores of all possible pairs of firms in each year by parsing firms' product descriptions in annual 10-K filings. If the similarity score between a firm and its potential peer firm is above the pre-determined similarity threshold, the latter is identified as a product market peer. Thus, each firm has its own distinct set of product market peers under this scheme. Also, the composition of the TNIC-based peer group varies over time because TNIC is based on firms' product descriptions in 10-K filings, which are updated annually. This reflects the changing nature of the firm's product markets as business strategies change. To form peer groups, we form quartile portfolios based on size and book-to-market (as in Albuquerque, 2009; 2014) within each focal firm's TNIC industry and define RPE peers as firms closest to the focal firm within the same quartile portfolio.[4] Equal-weighted average stock returns of these peers provide a measure of peer performance.

We begin by verifying whether TNIC-based peers indeed reflect product-market factors (e.g., demand or supply shocks) better than those based on pre-defined industry classifications. To do so, we estimate the correlation between firm sales and peer average sales using three alternative industry classifications – TNIC, SIC, and GICS. We also estimate the correlation between operating costs of the firm, i.e., the sum of the cost of goods sold and SG&A expenses, and peer firm average operating costs. Consistent with our expectation and prior evidence in Hoberg and Phillips (2016), we find that the correlation between firm sales and peer sales is the strongest based

[4] We refine the set of product market peers to those with comparable size and book-to-market values as the focal firm. We do so for two reasons – (i) recent studies note that matching the focal firm to SIC-based peers using size and book-to-market helps uncover evidence in favor of RPE (Albuquerque, 2009; 2014), and (ii) using closely matched size and book-to-market peers helps filter out managers' strategic use of self-selected larger firms to pay themselves more compensation (e.g., Bizjak, Lemmon, and Naveen, 2008; Faulkender and Yang, 2010). Table 5 presents results comparing RPE evidence with or without size and book-to-market matching.

on TNIC as compared to SIC or GICS.[5] The differences in correlations between TNIC and SIC/GICS are not only statistically significant but also economically meaningful. For example, the correlation between firm sales and TNIC-peer average sales is 0.785 while that for SIC (GICS) is 0.249 (0.169) after controlling for TNIC-peer average sales. The evidence for operating costs (i.e., the sum of cost of goods sold and SG&A expenses) is more pronounced – the correlation between operating costs of the firm and TNIC-based peer average operating costs is 0.933 while that based on SIC (GICS) is 0.164 (0.133) after controlling for TNIC-based peer average operating costs. Further, consistent with TNIC classifications better capturing the evolving nature of product markets, the above results are generally stronger in more recent time periods.[6] Overall, these tests indicate that TNIC-based classifications better capture product market factors than pre-defined classifications such as SIC or GICS.

Next, we turn to the RPE evidence. We use a firm-year panel of 26,187 observations spanning from 1996 to 2015 and investigate the presence of RPE in CEO compensation (controlling for the existing level of equity-based incentives). Specifically, we regress the natural logarithm of total CEO compensation on firm return, equal-weighted peer firm average return, control variables, Firm-CEO fixed effects, and calendar year fixed effects. First, we find a significantly positive coefficient on own firm stock return and a significantly negative coefficient on peer firm average return, consistent with CEOs being rewarded by positive own firm stock return, but filtering out the common performance shocks experienced by the same peer group.

[5] Hoberg and Phillips (2016) use a slightly different method by examining the extent to which alternative peer group classifications (TNIC/SIC/NAICS) generate higher levels of across-industry variation in profitability, sales group and stock market betas – where greater variation indicates a more informative industry classification.

[6] For example, the correlation between firm operating costs and TNIC-peer average operating costs is 0.834 in the 1996-2000 period, 0.965 in the 2000-2005 period, and 0.958 in the 2006-2010 period. In contrast, operating costs correlations based on SIC are decreasing over time – 0.250 in 1996-2000, 0.113 in 2001-2005, and 0.158 in 2006-2010. These effects become weaker in the most recent post-crisis period of 2011-2015 – in particular, the TNIC-based correlations drop relative to prior periods but are still larger than those based on SIC/GICS.

While the latter result (i.e., negative coefficient on peer returns) is consistent with RPE, the magnitude of the own firm effect (coefficient = 0.211) is significantly larger than the absolute magnitude of the peer firm effect (coefficient = -0.101) indicating only *partial* filtering (i.e., weak-form RPE, see Gibbons and Murphy, 1990). In other words, a 1% increase in the firm's own stock return is associated with a 21.1 basis points increase in total CEO compensation when the peer group experiences a 0% return. On the other hand, if the peer group also experiences a 1% return, CEO compensation reduces by 47.9% (-0.101/0.211), and she still enjoys a net 11 basis points (i.e., 0.211 - 0 .101) increase in total annual compensation.

Second, we find that the extent to which common performance is filtered out increases with the number of firms in the peer group (e.g., Holmstrom, 1982). The idea is that common shocks can be estimated with less noise when the peer group consists of many firms. This is because noise can be more effectively averaged out, especially when aggregating a number of peers' performance, leading to more a precise measure of common uncertainty. Consistent with theory, we find that the negative coefficient on peer stock returns becomes larger in magnitude as the number of firms in the same peer group increases. This coefficient takes a value of -0.040 when the firm has a few peers, -0.110 when the firm has a moderate number of peers and -0.203 when the firm has many peers.[7] Reassuringly, the coefficient on own firm performance does not differ significantly based on the number of firms in the peer group. This coefficient takes a value of 0.214 when the firm has a few peers, 0.224 when the firm has a moderate number of peers and 0.207 when the firm has many peers, and none of them is statistically different from each other. In terms of economic significance, when the focal firm and peer group both experience performance of a similar magnitude, the sensitivity of pay-for-own-firm-performance falls by 18.7% (-0.040/0.214)

[7] "Few", "moderate" and "many" are defined based on terciles of the number of firms in the TNIC peer group.

in the “few peers” group, by 49.11% (-0.110/0.224) in the “moderate peers” group, and by 98.07% (-0.203/0.207) in the “many peers” group.

The latter result represents our third finding, i.e., common performance is completely filtered out when the firm has many peers. In particular, net CEO compensation increases by a statistically insignificant 0.4 (i.e., 0.207 – 0.203) basis point when both the firm and the peer group experience stock returns of a similar magnitude – but only in firms with many peers. This evidence is referred to as “strong-form” RPE where common performance does not figure in CEO compensation, and she is compensated solely based on idiosyncratic performance.[8] This evidence speaks to Abowd and Kaplan’s (1999) opening quote about CEO rent-extraction and indicates that firms do not allow CEOs to walk away with millions during bull markets. However, their ability to rein in pay-for-luck hinges on the presence of many peers to better estimate common performance.

Our main results are robust to various sensitivity checks. First, using SIC or GICS provides evidence of weak-form RPE (see also Antle and Smith, 1986; Gibbons and Murphy, 1990; Aggarwal and Samwick, 1999a, 1999b; Garvey and Milbourn, 2006; Rajgopal et al., 2006; Alburquerque, 2009), but not strong-form RPE. We also do not observe that common performance is filtered out more when the number of firms in the same SIC/GICS industries increases. In addition, we find that SIC/GICS based industry classifications depict evidence of weak-form RPE only because they are correlated with TNIC-based classifications. In other words, including TNIC-based peer performance in the regression drives out the marginal explanatory power of SIC/GICS-based industry classifications. In addition, we exploit the time-varying nature of TNIC-classifications. We find that current stock performance of firms that have already exited the focal

[8] Holmstrom (1982) specifically states that “we would expect that with many agents we would be able to achieve approximately the same solution as if there were no common uncertainty at all.”

firm's product markets (i.e., past peers) as well as current stock performance of firms that will enter the focal firm's product markets in future periods (i.e., future peers) do not in general provide information about common performance while only current performance of current peers does. These results not only indicate that TNIC classifications incorporate the evolving nature of product strategies but also suggest that we are not capturing some mechanical aspect connecting the focal firm with these peers.

To strengthen our inferences, we perform two cross-sectional tests and examine situations where corporate boards might not want to filter out common performance. First, Aggarwal and Samwick (1999b) show that RPE is weaker when there is a need to soften product market competition. Consistent with this prediction, we find weaker evidence of RPE when rival firms' products are strategic complements because RPE, in this case, decreases shareholders' returns by encouraging more aggressive product market strategies. Second, Gopalan et al. (2010) theoretically show that common performance should not be filtered out when a firm's exposure to common external shocks is the CEO's choice (i.e., strategic flexibility). Consistent with their theory, we find that our RPE effect is weaker in situations where the CEO has greater strategic flexibility.

To further strengthen our argument, we investigate the evidence of RPE in forced CEO turnover decisions. Prior research finds mixed and limited evidence on RPE in the CEO turnover decisions. For instance, Jenter and Kanaan (2015) find that CEOs are fired after bad industry and market performance. Using hand-collected forced CEO turnover data following Parrino (1997) and Peters and Wagner (2014), we find evidence consistent with our pay results. Specifically, we find a significantly negative coefficient on own firm stock return and a significantly positive coefficient on peer firm average return, consistent with CEOs being fired for negative own firm performance while common performance shocks experienced by the same peer group are filtered out. Similar to the pay results, the magnitude of the own firm effect is significantly larger than the absolute magnitude of the peer firm effect, indicating only partial filtering, which is consistent with Jenter and Kanaan (2015). Next, we find that our primary results for the pay regressions translate to CEO replacement decisions: the filtering of common shocks increases with the number of peers, and complete filtering is observed when the number of peers is sufficiently large. This evidence corroborates the previous findings in CEO pay.

Our study contributes to the RPE literature in three ways. First, we hypothesize (and find confirmatory evidence) that a key identification strategy to testing RPE theory lies in accurately defining the peer group. Our study builds on the recent literature arguing that prior RPE research has failed to find the empirical evidence of RPE due to the incorrect identification of RPE peers (e.g., Albuquerque, 2009; Gong et al., 2011; Lewellen, 2013). In contrast to using static, predefined industry classifications, we employ new identification strategies by exploiting recent advances in textual analysis to identify product-market peers and find evidence consistent with RPE. In this sense, our study is also in line with Albuquerque, De Franco, and Verdi (2013) who suggest that an appropriate identification strategy helps researchers to draw clearer implications regarding efficient or opportunistic executive compensation practices.

Second, in both the CEO pay and the forced CEO turnover settings, we can confirm more stringent predictions of RPE theory, i.e., the extent of common performance filtering should increase with the number of peers, and that with a sufficiently large number of peers, the optimal contract should resemble one with no common uncertainty (Holmstrom, 1982). Prior attempts to find evidence supporting these predictions examine whether product market competition is positively associated with RPE, but ultimately only find mixed evidence (Aggarwal and Samwick, 1999b; DeFond and Park, 1999; Bushman and Smith, 2001; Ali et al., 2009).[9] We provide direct evidence consistent with RPE increasing with the number of peers operating in the firm's product markets and with the firm optimally using RPE in the presence of a sufficiently large number of peers.

Third, our results also speak to the long-standing debate about optimal contracting versus rent-extraction in explaining CEO compensation. While the presence of "pay-for-luck" is often cited as evidence in favor of CEO rent-extraction, our results suggest that this phenomenon is less prevalent in firms where market participants are privy to a relatively large number of reference points concerning CEO compensation. A fuller exploration of the role of corporate governance in the use of product-market peers based RPE is a fruitful area for further exploration.

This paper proceeds as follows. In the next section, we discuss the relevant literature and develop our hypotheses. Next, we discuss empirical specifications to test RPE theory in CEO compensation contracts, and then we present the estimation results, including robustness checks. Lastly, we conclude and summarize.

I. Literature Review and Hypothesis development

A. Relevant Literature

Holmstrom (1979) predicts that when the agent's efforts are unobservable and non-contractible, the second-best contracting mechanism is to provide an incentive contract where the

[9] Aggarwal and Samwick (1999b) find a negative association between product market competition and RPE in compensation contracts. DeFond and Park (1999) find a positive association between competition and RPE in CEO turnover decisions, while Ali et al, (2009) fail to replicate DeFond and Park (1999). Ali et al. (2009) point out that the competition measure used in DeFond and Park (i.e., Sales-based HHI) is based on sales of only publicly-traded firms, resulting in a biased measure of competition. Overall, it is a still open question whether greater product market competition is positively associated with RPE in both CEO compensation and turnover decisions. Bushman and Smith (2001), for example, call for research to resolve conflicting results in Aggarwal and Samwick (1999b) and DeFond and Park (1999).

agent's compensation is contingent on observable measures of firm performance. Consistent with this prediction, prior research shows that CEOs are rewarded by increases in the own firm stock returns (i.e., positive pay-for-performance sensitivity; Jensen and Murphy, 1990; Aggarwal and Samwick, 1999a). This incentive contract, however, imposes an unnecessary risk on the risk-averse agent to the extent that firm performance is influenced by external shocks that are not under the agent's control. These uncontrollable shocks potentially decrease the utility of the agent thereby reducing contracting efficiency. One solution proposed by Holmstrom (1982) is to filter out these external shocks from firm performance, thereby resulting in a greater "signal-to-noise" ratio, which in turn results in greater contracting efficiency. That is, the agent should not be rewarded solely for her own *total* performance but rather for performance relative to that of her peers. This is the idea behind the Relative Performance Evaluation (RPE) theory.

Prior studies have attempted to test this RPE theory in CEO compensation contracts (e.g., Antle and Smith, 1986; Gibbons and Murphy, 1990; Jensen and Murphy, 1990; Janakiraman, Lambert, and Larcker, 1992; Aggarwal and Samwick, 1999a, among others). However, this evidence is mixed at best (Prendergast, 1999; Lambert, 2001; Frydman and Jenter, 2013), which in turn has resulted in alternative theories that seek to explain this "RPE puzzle."

For example, Bertrand and Mullainathan (2001) argue the lack of RPE is attributed to the rent-seeking behavior of managers. They argue that firms with weak corporate governance are less likely to use RPE because CEOs in these firms can affect their pay-setting process and are paid for positive external shocks, but not similarly penalized for negative external shocks (i.e., pay-for-luck). Another stream of research seeks to find cross-sectional evidence of RPE by identifying factors that alter the costs and benefits of using RPE. For instance, Gopalan et al. (2010) show that if the exposure to common external shocks is a strategic choice of the CEO (i.e., strategic flexibility), then RPE is less likely to be used in compensation contracts.

The third stream of research seeks to test RPE by limiting the set of firms within the industry group that can be considered peers. For instance, Albuquerque (2009) argues that using the entire SIC group as peers is problematic because all firms in this group may not face common external shocks and firms' abilities to respond to common shocks is likely to vary substantially within the same industry. Albuquerque (2009) refines the set the peers within the focal firm's two-digit SIC industry group to those in the same size quartile portfolio and finds evidence consistent with RPE. In a similar vein, Dikolli, Hofmann, and Pfeiffer (2011) show analytically that aggregating heterogeneous firm performance within the same industry adds significant summarization bias in the measure of common shocks, leading to the failure in detecting RPE.

While the above studies use firms in the same industry as RPE peers, another stream uses peer firms that are self-disclosed by the firm (e.g., Murphy, 1999; Bannister and Newman, 2003; Carter et al., 2009; Gong et al., 2011; Lewellen, 2013).[10] Similar to the above, these studies also argue that using all firms in the same industry might lead to a noisy measure of common external shocks, failing to detect RPE in the data.

B. Product Market Peers

In this study, we hypothesize that the prior lack of RPE evidence is primarily on account of the use of pre-defined industry classifications such as SIC, GICS, and NAICS to define the peer

[10] For example, Gong et al. (2011) use compensation disclosures mandated by SEC after 2006 and examine the RPE theory. Interestingly, Gong et al. (2011) do not find evidence of RPE following the method used in Albuquerque (2009), but find evidence of weak-form evidence of RPE using self-selected RPE peers by the firm. Similar to Gong et al. (2011), Lewellen (2013) collects a firm's significant competitors disclosed in the firm's 10-K filings, and finds evidence consistent with RPE.

group. We reason that these pre-defined industry classifications group firms based on inputs rather than the similarity in products or outputs (e.g., Bhojraj et al., 2003, Guenther and Rosman, 1994).[11] RPE theory assumes homogenous agents in the same team that shares the same common uncertainty (Holmstrom, 1982). Arguably, the empirical counterpart of common uncertainty is common demand and supply shocks that affect all firms producing similar products, rather than, for example, having similar production functions. The distinction is important because having similar production functions does not necessarily imply those firms producing similar products (e.g., Bernard and Skinner, 1996; Brickley and Zimmerman, 2010).

In addition, pre-defined industry classifications rarely change over time and consequently do not capture the evolving nature of the firm's product markets as the firm's product offerings change (Hoberg and Phillips, 2016). A firm enters or exits its peers' product market space if the latter starts or stops producing similar products (not whether or not it uses similar production processes). Although the firm's product market peers also change accordingly, in this case, traditional industry classifications do not reflect this as these classifications do not evolve rapidly. Accordingly, pre-defined industry classifications fail to capture this dynamic nature of evolving product markets. If so, RPE tests relying on these pre-defined industry classifications might fail to detect consistent evidence. Gibbons and Murphy (1990, p. 49) allude to this possibility by stating that "…our inability to detect an industry effect after controlling for market movements may reflect the inappropriateness of industry definitions based on SIC codes for purposes of relative performance evaluation."

[11] http://www.naics.com/info.htm. The Census Department states "NAICS was developed to classify units according to their production function. NAICS results in industries that group units undertaking similar activities using similar resources but does not necessarily group all similar products or outputs." Prior studies also note how these standard classifications ignore within-industry heterogeneity in production processes, resulting in misclassification (e.g., Clarke, 1989; Kahle and Walkling, 1996; Bhojraj et al., 2003; Dopuch et al., 2008; Hoberg and Phillips, 2016).

Hence, we argue that the empirical analysis should identify peer firms producing similar products who face similar demand and supply shocks as RPE peers. To this end, we use Text-based Network Industry Classifications (TNIC) recently developed by Hoberg and Phillips (2016) to identify RPE peers. Hoberg and Phillips (2016) identify peer firms based on the pairwise product similarity scores among firms by parsing firms' product descriptions in annual 10-K filings (Item 1 or 1A). They argue that firms producing similar products are more likely to be peer firms competing in the same product markets. Hoberg and Phillip (2016) validate that TNIC better explains differences in industry characteristics such as profitability, sales growth, and market risk across the industry. They show that positive (negative) industry demand shocks lead to more (less) firms entering into those industries. They also show that these classifications better reflect competitors identified by managers. Several other studies using TNIC find that this classification scheme provides new insights regarding a firm's product market peers. For example, Hoberg and Phillip (2010) show that M&A transactions are more likely between firms having similar product descriptions and long-term outcome such as profitability is better when the target and the acquirer have similar product descriptions ex-ante, possibly due to product market synergies. Foucault and Fresard (2014) show that a firm's investment is sensitive to the stock returns of product market peers.

To compute the product similarity, Hoberg and Phillips (2016) specifically convert each firm's product description in 10-K filings into a word vector and calculate product cosine similarity scores for every pair of firms (i.e., the distance between two-word vectors for every pair of firms). For example, a firm *i*'s product similarity score with a firm *j* is calculated as the dot product of the word vector of the firm *i*, which consists of vocabularies describing the firm *i*'s products, and that of the firm *j*. This cosine product similarity score between firm *i* and firm *j* is bounded in [0,1] and increases with the number of same words that both firm *i* and firm *j* use, implying that firm pairs with high cosine similarity scores are likely to operate in the similar product markets. Firm *j* is classified as firm *i*'s product market peer if product similarity score between firm *i* and firm *j* is above a pre-specified minimum similarity threshold.[12] This classification yields a group of product market peers for every firm, which allows peer group composition to vary year-to-year and firm-by-firm. Hoberg and Phillips (2016) argue that this procedure can capture the notion that the most appropriate peer firms are firms producing similar products. In addition, Hoberg and Phillips (2016) also argue that TNIC captures the changing nature of product markets over time because all firms' update their product descriptions annually and the updates are required to be correct and timely by SEC. Hence, we test the RPE theory using TNIC-based peers. This leads to our first testable RPE prediction:

H1: Firms base CEO compensation not only on own firm performance but also filter out the performance of their product market peers.

C. Implication of the Number of Product Market Peers in RPE

While prior studies also find evidence in favor of the RPE hypothesis above, we go one step further and devise more stringent tests based on RPE theory. In particular, we hypothesize that the extent of RPE (i.e., filtering out of peer performance) increases with the number of firms in the peer group. In addition, we predict that common performance is completely filtered out in the presence of a large number of peers (i.e., strong-form RPE).

[12] Hoberg and Phillips (2016) state that "Although one can use any minimum similarity threshold to construct a classification, we focus on thresholds generating industries with the same fraction of membership pairs as SIC-3 industries, allowing us to compare our industries to SIC-3 in an unbiased fashion."

In Holmstrom (1982), each agent's performance (x_i) is determined by effort a_i, common uncertainty parameter η, which affects all agents in the same team, and idiosyncratic error term e_i, which is determined by the agent-specific efforts (i.e., $x_i = a_i + \eta + e_i$). Hence, each agent's uncertainty is determined by common uncertainty parameter η and idiosyncratic error term e_i. By aggregating performance of all agents in the same team, the idiosyncratic error terms are averaged out in the aggregate performance index, and thus the common uncertainty parameter η can be estimated. Holmstrom (1982) proceeds to predict that if the number of agents is large enough to infer the precise value of the common uncertainty parameter, the principal can completely filter out common uncertainty in evaluating the agent's performance. On the other hand, if the number of agents in a team is small, then the idiosyncratic performance of agents is not sufficiently eliminated in the aggregation process, resulting in the principal only partially filtering out common uncertainty in evaluating the agent's performance (Gibbons and Murphy, 1990). This prediction forms our second and third testable hypotheses that are stated as follows:

H2: ***The extent of filtering of common performance in CEO compensation increases with the number of product market peers.***

H3: ***Firms completely filter out common performance in CEO compensation in the presence of a large number of product market peers.***

II. Research Design

A. Empirical Specification

We use the empirical specification proposed by Holmstrom and Milgrom (1987) and widely used in prior RPE studies (e.g., Gibbons and Murphy, 1990; Albuquerque, 2014).

$$ln(Total\ Comp_t) = \beta_1\ Firm\ Ret_t + \beta_2\ Peer\ Ret_t + \beta_3\ Size_{t-1} + \beta_4\ BM_{t-1} + \beta_5\ Vol_{t-1} + \beta_6\ Tenure_t + \beta_7\ Age_t + \beta_8\ Ownership_{t-1} + \beta_9\ lnDelta_{t-1} + \varepsilon_{i,t} \quad (1)$$

The dependent variable is the natural logarithm of one plus annual total CEO compensation, measured as the sum of salary, bonus, the grant-date fair value of stock and option grants, long-term incentive payouts, other annual compensation, and all other annual compensation (i.e., variable TDC1 in ExecuComp). *Firm* Ret_t captures firm *i*'s own stock price performance and is defined as the annual buy-and-hold stock return including dividends. *Peer* Ret_t captures the average stock performance of firm *i*'s product market peers and is measured as the equal-weighted average of annual stock returns of product market peers excluding firm *i*. To define product market peers, we choose one-quarter of TNIC peers with the smallest the Mahalanobis distance using the market value of equity and the book-to-market ratio within each focal firm's TNIC group in each fiscal period (e.g., Albuquerque, 2009; Lys and Sabino, 1992).[13] Firms in the same quartile portfolio as firm *i* (excluding firm *i*) are defined as firm *i*'s product market peers in period *t*.

Following prior studies, we include several control variables in the model. We include $Size_{t-1}$, which is measured as the natural log of total revenue for firm *i* at the beginning of period *t* (Smith and Watts, 1992). BM_{t-1} proxies for growth options and is measured as the book-to-market ratio for firm *i* at the beginning of period *t* (Smith and Watts, 1992). Following Aggarwal and Samwick (1999a), we also include idiosyncratic volatility (Vol_{t-1}), which is measured as the standard deviation of the residuals obtained from a regression of monthly firm return on the

---

[13] Specifically, we take the following steps to choose the closest RPE peers in terms of size and book-to-market. First, we merge the latest market value of equity and book-to-market of TNIC peers as of the beginning of the focal firm *i*'s fiscal period. We drop peer firm observations with missing values of market value of equity and book-to-market. We truncate stock returns at the .5st and 99.5th percentiles before computing averages to mitigate the influence of extreme observations. We compute the Mahalanobis distance between the focal firm and each peer firm using the market value of equity and book-to-market in each year. Finally, we choose one quarter of TNIC peers that are closest to firm *i* in terms of the distance. Lys and Sabino (1992) show that researchers can maximize the power of their tests by placing 27% of the sample on each of the extreme portfolios. We require a minimum of two peers for each focal firm in each year.

monthly equal-weighted average peer return using the preceding 12 months of observations for firm *i* in period *t* (a minimum of 6 observations is required). We include (the natural logarithm of) CEO tenure ($Tenure_t$), CEO age (Age_t), and CEO stock ownership (Own_{t-1}) to control for the effects of CEO characteristics on firms' compensation policies. Portfolio delta is included in the regression model to control for any potential influences of existing incentives on corporate boards' decisions to filter out common shocks. Portfolio delta measures the dollar change in wealth experienced by the CEO for a 1% change in the firm's stock price (Core and Guay, 2002; Coles, Daniel, and Naveen, 2006). $lnDelta_{t-1}$ is defined as the natural logarithm of one plus portfolio delta for the CEO at the beginning of period *t*. We also include Firm-CEO and calendar year fixed effects to control for unobservable time-invariant individual Firm-CEO characteristics and time-varying common factors. Also, including Firm-CEO fixed effects allows us to identify whether changes in total CEO compensation are associated with firm return and peer return (Gormley and Matsa, 2014). We winsorize all continuous variables at the 0.5% and 99.5% levels to mitigate the influence of extreme observations. We cluster standard errors by firm (Petersen, 2009).

The coefficient on $Firm\ Ret_t$ in Equation 1 is expected to be positive (i.e., pay-performance sensitivity), while that on $Peer\ Ret_t$ captures the RPE effect and is expected to be negative (i.e., lower compensation for greater common performance). In addition, optimal contracting theory predicts that the sum of the coefficients on $Firm\ Ret_t$ and $Peer\ Ret_t$ is statistically zero if common performance is completely filtered out while compensating the CEO, and thus she is evaluated solely based on idiosyncratic performance (Holmstrom, 1982; Gibbons and Murphy 1990).

B. Data and Descriptive Statistics

We retrieve market values of equity and stock returns data from CRSP, financial statement data from Compustat, and CEO compensation data from ExecuComp. We adjust delisting returns following Beaver, McNichols, and Price (2007). Following Garvey and Milbourn (2006), we use a sample of ExecuComp firms with non-negative CEO tenure. We also delete observations with missing financial and compensation data. The above data requirements yield a sample of 26,187 firm-year observations. The sample period ranges from 1996 to 2015 because TNIC are only available for this sample period.[14]

Panel A of Table 1 presents descriptive statistics for our main dependent and independent variables. The mean (median) of total compensation is $5.45 million ($3.18 million), which shows significant right skewness as in prior compensation studies (e.g., Albuquerque, 2009). Hence, we take the natural logarithm of total compensation to reduce skewness. Panel B of Table 1 presents Pearson correlations among main variables. We note that total compensation is significantly positively correlated with firm return (Pearson correlation of 0.08) and peer return (Pearson correlation of 0.04). We find that total CEO compensation is significantly positively correlated with firm size (Pearson correlation of 0.61) and negatively correlated with the book-to-market ratio (-0.15), suggesting that both large firms and growth firms incur greater compensation costs to hire talented managers (Smith and Watts, 1992).

III. Empirical Results

A. Validity Check

We begin by validating our main assumption that TNIC better captures a firm's demand and supply shocks in product markets relative to SIC and GICS. Using average sales as a measure

[14] TNIC data is obtained from http://hobergphillips.usc.edu/.

of demand shocks, we run a regression of firm *i*'s sales on peer firm average sales based on TNIC, SIC, and GICS (excluding firm *i*). This test allows us to examine the correlation between firm *i*'s sales and peer sales holding the effect of other industry classifications fixed. With regard to supply shocks, we estimate correlations between firm *i's* operating costs (i.e., the sum of cost of goods sold and SG&A expenses) and average operating costs based on TNIC, SIC, and GICS.

Table 3 tabulates these results. In panel A, we examine sales correlations. In column (1) of Panel A, we find that the coefficient on *Average Sales* (*TNIC*) is 0.785 while the coefficients on *Average Sales* (*SIC*) and *Average Sales* (*GICS*) are 0.249 and 0.169, respectively. This result is consistent with our expectation and Hoberg and Phillip (2016) and suggests that TNIC better captures firms' demand shocks than SIC or GICS. Correlations based on operating cost are even stronger. In column (1) of Panel B, the coefficient on *Average Costs* (*TNIC*) is 0.933 while the coefficient on *Average Costs* (*SIC*) and *Average Costs* (*GICS*) is 0.164 and 0.133, respectively. This evidence suggests that TNIC better captures firms' supply shocks than SIC or GICS.

We further examine whether the above results are stronger in more recent periods. If TNIC better captures the evolving nature of product markets and is updated annually, we would expect the TNIC-based correlations to be stronger in more recent periods. To this end, we partition our full sample into four subsamples based on 5-year periods, and estimate sales and operating costs correlations for each period (columns (2) through (5) in each panel). In general, we find that sales and operating costs correlations based on TNIC are increasing in more recent periods. In particular, in Panel B, we find that the correlation between firm *i*'s operating costs and TNIC-peer average operating costs is 0.834 in the 1996-2000 period, 0.965 in the 2001-2005 period and 0.958 in the 2006-2010 period. In contrast, COGS correlations based on SIC are decreasing over time: – 0.250 in 1996-2000; 0.113 in 2001-2005 and 0.158 in 2006-2010. These effects become weaker in the most recent post-crisis period of 2011-2015 – in particular, the TNIC-based correlations drop relative to prior periods, but remain larger than those based on SIC/GICS.

Overall, the findings in Table 3 suggest that TNIC better captures a firm's supply and demand shocks as evidenced by stronger correlations between firms' sales and TNIC-peer average sales and between firms' operating costs and TNIC-peer average operating costs. In the next section, we present results of our main RPE hypotheses.

B. Main Results

Table 3 presents the estimation results using Equation 1. We start with the full-sample results in column 1. Consistent with positive pay-for-performance sensitivity, the coefficient on *Firm Ret*$_t$ is positive (coefficient of 0.211) and significant at the 1% level. Furthermore, consistent with our first RPE prediction, the coefficient on *Peer Ret*$_t$ is negative (coefficient of -0.101) and also significant at the 1% level. The evidence supports the weak-form version of RPE (i.e., partial filtering of common performance, Gibbons and Murphy 1990), given that the absolute value of the coefficient on *Peer Ret*$_t$ is significantly less than that on *Firm Ret*$_t$ (F-stat = 30.99). Stated in economic terms, CEO compensation increases by 21.1 basis points when the firm experiences a 1% increase in its own stock price and its peers experience a 0% stock return during the fiscal year. However, if peer stock returns also increase by 1%, the CEO only experiences an increase in annual total compensation of 11 basis points (0.211-0.101).

To test our second prediction that the extent of common performance filtering increases with the number of suitable peers, we divide the sample into three subsamples based on the tercile of the number of TNIC peers in period *t* and estimate Equation 1 within each subsample. Firm-year observations that belong to the first, second, and third tercile of the number of TNIC peers are classified as the Few, Moderate, and Many groups, respectively. We expect the coefficient on *Peer Ret*$_t$ to become increasingly more negative as we move from the Few group to the Moderate group to the Many groups, and that is exactly what we find. The coefficient on *Peer Ret*$_t$ is -0.040 in the Few peer group and statistically insignificant, -0.110 in the Moderate peer group and significant at the 1% level, and -0.203 in the Many peer group and significant at the 1% level. The coefficient on *Peer Ret*$_t$ is insignificantly different between the Few and the Moderate groups (difference of -0.069), and significantly different between the Moderate and the Many groups at the 10% level (difference of -0.094), and also between the Few and the Many groups at the 1% level (difference of -0.163). Importantly, the coefficient on *Firm Ret*$_t$ does not vary significantly across these subsamples (0.214, 0.224 and 0.207). This is comforting because there is no a priori reason for the extent of pay-for-performance sensitivity to differ based on the size of the peer group – only the extent of RPE is predicted to be affected by the size of the peer group. In terms of economic significance, when both the firm and its peers experience the same magnitude of common stock return, the positive pay-for-performance sensitivity decreases by 18.69% (-0.040/0.214) in the Few peer group, by 49.11% (-0.110/0.224) in the Moderate peer group, and by 98.07% (-0.203/0.207) in the Many peer group.

The latter result is consistent with our third prediction of complete filtering (i.e., strong-form evidence on RPE, Holmstrom, 1982). In particular, the sum of the coefficients on *Firm Ret*$_t$ and *Peer Ret*$_t$ is indistinguishable from zero (F-stat = 0.010 p-value = 0.925), suggesting that common performance is completely filtered out while evaluating the CEO. Overall, these results suggest that firms use RPE in rewarding their CEOs, and the extent of RPE usage depends on the presence of a large enough number of true product market peers.

IV. Additional Tests

A. Alternative Industry Classifications

In this section, we replicate our results using pre-defined industry classifications. In Panel A of Table 4, we use three-digit SIC codes and calculate *SIC Peer Ret*$_t$ based on the same method used in the construction of our main peer return variable using TNIC, *Peer Ret*$_t$.[15] In column 1, we find the same result documented in column 1 of Table 3, i.e., weak-form evidence of RPE. This result is consistent with prior RPE research (e.g., Gibbons and Murphy, 1990; Albuquerque, 2009). Next, we partition the sample into three subsamples based on the tercile of the number of firms in the same SIC industry and estimate Equation 1 within each subsample. In this analysis, we do not find evidence consistent with our two predictions relating the efficacy of RPE to the number of peers. In particular, the coefficient on *SIC Peer Ret*$_t$ does not show monotonicity as we move from the Few group to the Moderate group to the Many groups based on the number of SIC peers – rather the coefficient on *SIC Peer Ret*$_t$ is most negative and significant in the moderate peers subsample. Furthermore, we are also unable to find evidence consistent with our third prediction of complete filtering in the Many peers subsample. In particular, the coefficient on *Firm Ret*$_t$ is 0.134 while that on *SIC Peer Ret*$_t$ is -0.058 and marginally insignificant, indicating that the CEO continues to enjoy a 7.6 basis points increase in annual compensation even when both the firm and the peer group experience a 1% stock return during the year (F-statistic = 4.230 p-value 0.040).

In Panel B of Table 4, we use GICS industry codes to define peers. GICS industry codes are the most recent and improved industry classification method developed by MSCI Inc. and S&P (e.g., Bhojraj et al., 2003). Here again, while we find evidence consistent with weak-form RPE in

[15] TNIC is comparable with three-digit SIC because the pre-specified minimum product similarity threshold use in constructing TNIC is set to generate industries with the same fraction of industry pairs as three-digit SIC industries (Hoberg and Phillips, 2016). Results using NAICS are similar and are not tabulated.

the full-sample, we are unable to find evidence consistent with our other two predictions. The coefficient on *GICS Peer* Ret_t shows patterns similar to those in Panel A. They do not show monotonicity as we move from few peers to moderate peers to many peers, and the most negative and significant one is found in the moderate peers subsample. Once again, there is no evidence of complete filtering of common performance in the many peers subsample – CEOs continue to enjoy 9.0 basis points increase in annual compensation when the firm and the peer group both enjoy a 1% annual stock return.

If SIC and GICS industries are poor proxies for the firm's peer group, why do we observe evidence consistent with partial filtering (i.e., weak-form evidence of RPE) using these classifications? And also some weak evidence of monotonicity between the Few and the Moderate peer subsamples? We conjecture that this is due to these proxies being correlated with TNIC-based classifications. To examine this possibility, we include both *Peer* Ret_t, which is based on TNIC, and *SIC Peer* Ret_t or *GICS Peer* Ret_t that is based on these alternative pre-defined industry classifications simultaneously in the same regression. If our conjecture is true, we would observe the statistically negative coefficient only on *Peer* Ret_t but not on *SIC Peer* Ret_t or *GICS Peer* Ret_t.

Table 5 presents the estimation results comparing peer return variables. In column 1, we include *Peer* Ret_t and *SIC Peer* Ret_t simultaneously in the same regression and find that only the coefficient on *Peer* Ret_t is significantly negative, while the coefficient on *SIC Peer* Ret_t is statistically insignificant. In column 2, we find a similar result when we replace *SIC Peer* Ret_t with *GICS Peer* Ret_t. In column 3, we include all three peer return variables and find that only *Peer* Ret_t is significantly negatively correlated with the dependent variable at 1% level while coefficients on both *SIC Peer* Ret_t and *GICS Peer* Ret_t are statistically insignificant.

In our main analysis, we use the characteristics-matched peer firms based on size and book-to-market to construct our measure of peer returns. One potential concern is that the matching would be the primary factor resulting in better identification of peer group rather than TNIC groupings (i.e., Albuquerque, 2009). To rule out this possibility, in columns 4, 5, and 6, we do not use the matching, but rather use all firms in each TNIC, SIC, and GICS group to calculate *All Peer* Ret_t, *All SIC Peer* Ret_t, and *All GICS Peer* Ret_t variables. In columns 4 through 6, we confirm our conjecture and find that only the coefficient on *All Peer* Ret_t remains negative and statistically significant, while those on *All SIC Peer* Ret_t and *All GICS Peer* Ret_t become insignificant. These results corroborate our argument that TNIC provides a better identification of peers relative to pre-defined industry classifications.

Finally, in column 7, we include the peer return variable based on the characteristics-matched TNIC peers (*Peer* Ret_t) and the peer return variable based on all TNIC peers (*All Peer* Ret_t) simultaneously in the same regression. We find that the coefficient on *All Peer* Ret_t becomes insignificant, while the coefficient on *Peer* Ret_t remains statistically significant at 1% level. This result is consistent with the prior literature and suggests that matching based on firm characteristics in the same peer group also improves the identification of peers that are subject to common external shocks (Albuquerque 2009).

Overall, the results in Table 5 suggest that TNIC-based peers provide a better proxy for common peer group performance than those based on pre-defined industry classifications in the CEO pay setting. Also, these findings suggest that proxies based on pre-defined industry classifications appear to provide some evidence consistent with weak-form RPE because they are correlated with product-market based TNIC industry classifications.

B. Dynamic Peer Groupings and RPE

As noted earlier, one of the key advantages of using TNIC to identify RPE peers is that TNIC captures the evolving nature of product markets. Therefore, we can examine whether current stock returns of past, current, and future product market peers contain information about common performance. For example, consider *past* peer firm *j* that was the product market peer of firm *i* in period *t-1*, but not in period *t* (i.e., firm *j* exited firm *i*'s product space in period *t-1*). In this case, firm *j*'s current stock return in period *t* is less likely to contain information regarding common demand and supply shocks that firm *i* faces in period *t*. Similarly, if firm *k* is not a product market peer of firm *i* in period *t* but only becomes a peer in period *t+1* (i.e., future peer), then the stock returns of firm *k* in period *t* are also less likely to contain relevant information about common shocks that firm *i* is experiencing in period *t*. In reality, entering new product markets takes time, and hence firm *k* is most likely taking some activities to enter the new product market in the current period *t* (e.g., investments), resulting in firm *k*'s stock returns in period *t* presumably containing information regarding common external shocks. Foucault and Fresard (2014) adopt this approach and show that past (future) peers' stock price is not (weakly) associated with the focal firm's investment, while present peers' stock price is informative to the focal firm's investment.

Similar to Foucault and Fresard (2014), we classify peer firm observations that are used to construct the $Peer\ Ret_t$ variable into four sets of peer firms: (1) past peers, (2) new peers, (3) current peers, and (4) future peers. We define *Past Peers* as firms that were firm *i*'s product market peers in period *t-1* but are not in the same TNIC group in period *t*. *New Peers* are firm *i*'s product market peers in period *t* but were not in the same TNIC group in period *t-1*. *Current Peers* are firm *i*'s product market peers in period *t-1* as well as in period *t.* Lastly, we define *Future Peers* as firms that will be firm *i*'s product market peers in period *t+1* but are not in the same TNIC group in period t. We then calculate equal-weighted stock returns of each set of peers using stock returns in period t and replace *Peer Ret*$_t$ in Equation 1 with each of these stock returns.

Table 6 reports the results. Consistent with our expectations, in column 1 of Panel A, the current period stock returns of past peers (*Past Peer Ret*$_t$) is not statistically significant, suggesting that current stock returns of past peers do not contain information regarding common shocks. In columns 2 and 3, we find that the coefficients on *New Peer Ret*$_t$ and *Current Peer Ret*$_t$ are significantly negative at 1% level, respectively. In column 4, we include the two variables simultaneously in the same regression and find that the coefficient on *New Peer Ret*$_t$ becomes statistically insignificant, while the coefficient on *Current Peer Ret*$_t$ remains statistically significant at 5% level. In column 5, we find that *Future Peer Ret*$_t$ is not statistically significant, suggesting that current stock returns of firms that are expected to enter firm *i*'s product markets in the next period also do not contain information concerning common shocks in period t. Overall, this time-series evidence not only corroborates our RPE hypothesis, but also mitigates concerns that we are capturing some mechanical feature linking our focal firm to these product market peers.

C. Cross-sectional Tests

To further strengthen our inferences, we perform several cross-sectional tests. First, Aggarwal and Samwick (1999b) show that the nature of product market competition affects the extent to which corporate boards use RPE. Specifically, they show that if a firm's product market outputs are strategic complements with rivals, then corporate boards are less likely to use RPE since RPE, in this case, incentivizes managers to take aggressive price strategies which in turn lower shareholders' returns. In contrast, if the firm's output is a strategic substitute with that of its competitors, then managers have weaker incentives to maximize own firm value but stronger incentives to increase all other firms' value (i.e., collusion). In this case, corporate boards are more likely to use RPE.

To examine this prediction, we use the Competitive Strategic Measure (CSM) to identify whether the product market competition is described as one of strategic complements or substitutes (Sundaram et al. 1996; Chod and Lyandres 2011). CSM is defined as the coefficient of correlation between the ratio of the change of a firm's profits to the change of its sales, and the change in the combined sales of its rivals. Intuitively, CSM captures the cross-partial derivative of firm value with regards to industry peers' strategic actions (as measured by changes in sales). If the CSM has a positive value, it indicates that the competition is one of strategic complements; otherwise, the competition is one of strategic substitutes.

Specifically, in columns 1 and 2 of Table 7, we divide the full sample into two subsamples based on the sign of CSM: column 1 uses a subsample of firms with positive CSM (strategic complements), while column 2 uses a subsample of firms with negative CSM (strategic substitutes). Consistent with our expectations, we find that the coefficient on *Peer Ret*$_t$ in column 2 is -0.144, and the absolute value of this coefficient is statistically greater at the 5% level than the absolute value of the coefficient on *Peer Ret*$_t$ in column 1 (-0.053). These findings are consistent with Aggarwal and Samwick (1999b) and suggest that firms use less RPE when the product market competition is characterized as one of strategic complements.

To provide further support, we use an alternative competition measure to test this prediction. Specifically, we use sales-based Herfindahl-Hirschman Index (HHI) of TNIC to divide the full sample into two subsample assuming that if a firm's product market is more concentrated, then the competition is more likely to be one of strategic complements, while if a firm's product market is less concentrated, then the competition is more likely to be one of strategic substitutes.

In columns 3 and 4, again, we divide the full sample into two subsamples based on the sample median of sales-based HHI. Again, consistent with our expectations, we find that the coefficient on *Peer* Ret_t in column 4 is -0.165, and the absolute value of this coefficient is statistically greater at the 5% level than that column 3 (-0.072). This result corroborates our earlier finding and suggests the nature of product market competition affects the use of RPE in compensation contracts.

Second, we examine the theoretical predictions of RPE in Gopalan et al. (2010). Gopalan et al. propose a model showing that the use of RPE decreases if firms want to provide strategic flexibility to their CEOs. They argue that "the board of directors is not primarily concerned with how hard the CEO is actually working, but whether she has the vision to choose the right strategy for deploying the firm's assets. In doing so, the CEO's concern is with the firm's strategic direction in lieu of its surrounding market environment." Put differently, if the effect of common external shocks on firm performance is not random, but due to specific actions undertaken by the CEO, then the effect of common external shocks should not be excluded in evaluating the CEO's efforts.

To test this prediction, we follow Gopalan et al. and use the following three measures to identify firms that offer greater strategic flexibility to the CEO: peer-adjusted market-to-book ratio, peer-adjusted stock returns during the previous period, and the asset growth in the next period. First, firms with high market-to-book ratios are more likely to have greater growth options and thus are more likely to provide their CEOs with the greater strategic flexibility to allow for more discretion in exercising those options. Therefore, we classify firm-years with peer-adjusted market-to-book ratios above the median as offering greater strategic flexibility to the CEO. Second, RPE is reduced for more talented CEOs due to the decreasing disutility of effort for more talented CEOs. Thus, we classify firm-years with positive peer-adjusted stock returns during period *t-1* as having more talented CEOs because firms managed by more talented CEOs are more likely to exhibit better peer-adjusted stock performance. Third, if less RPE allows CEOs to have greater strategic flexibility, we expect to observe some evidence that CEOs with less RPE exploit the strategic flexibility to a greater extent at the firm level such as greater asset growth. Hence, we classify firm-years with positive asset growth in period *t+1* as exploiting their strategic flexibility to a greater extent and examine whether firms with positive (negative) asset growth in period *t+1* are less (more) likely to use RPE in period *t*. [16]

Table 8 presents the results. In columns 1 and 2, we divide the full sample into two subsamples based on median peer-adjusted market-to-book. Consistent with our expectations, we find evidence of less filtering of common shocks for high market-to-book firms in column 2 and the difference between columns 1 and 2 is statistically significant at the 1% level. This evidence suggests that firms providing their CEOs with greater strategic flexibility use less RPE. In columns 3 and 4, we use the peer-adjusted stock returns as a conditioning variable and find similar results. The absolute magnitude of the coefficient on $Peer\ Ret_t$ in column 4 (positive peer-adjusted returns) is statistically lower at the 5% level relative to that in column 3 (negative peer-adjusted returns). Lastly, in columns 5 and 6, we use the asset growth rate in period *t+1* to investigate whether CEOs with less RPE in period *t* exploit their strategic flexibility in the subsequent period to a greater extent at the firm level. Consistent with our expectations, we find that firms with greater asset growth in period *t+1* filter out common shocks to a lesser extent in period *t* as evidenced by the

[16] Gopalan et al. also examine whether multi-segment firms (based on the SIC industry) are less likely to use RPE. We do not examine this variable because segment information in Compustat is only based on SIC industries, which we do not rely on in our study. Gopalan et al. also use R&D expenditures to test the theory. In untabulated tests, we find that the extent of RPE in firms with high R&D expenditures is not significantly different from RPE in firms with low R&D expenditures. This result could be attributed to R&D expenditures being a noisier measure of the firm's growth options because a significant portion of firms in Compustat universe does not report R&D expenditures separately. Kob and Reeb (2015) find that approximately 10.5% of firms with missing R&D from Compustat actually have active patenting activities, suggesting that R&D expenditures from Compustat are a noisy measure of growth options.

significantly lower absolute magnitude of the coefficient on *Peer Ret*$_t$ at 10% level in column 6 (positive asset growth rate) relative to that in column 7 (negative asset growth rate). In sum, the results in Table 8 are consistent with predictions and findings in Gopalan et al. and suggest that the use of RPE is attenuated by the board's desire to promote strategic flexibility on the part of the CEO.

D. RPE Tests for Forced CEO Turnover Decisions

Lastly, we investigate the evidence of RPE in CEO turnover decisions. Prior research finds mixed evidence of RPE in CEO turnover decisions. For example, Jenter and Kanaan (2015) find that CEOs are fired after bad industry and market performance using Fama-French 48 industry classifications. Also, while DeFond and Park (1999) find a positive association between product market competition (based on sales-based HHI using 2-digit SIC) and RPE in CEO turnover decisions, Ali et al. (2009) use an alternative measure of product market completion and fail to replicate these results in DeFond and Park (1999).

We examine the RPE hypothesis in forced CEO turnover decisions by identifying peer groups based on TNIC classifications. Following prior studies, we examine RPE in forced CEO turnover decisions using a linear probability model with firm and year fixed effects (Cornelli, Kominek, and Ljungqvist 2012; Guo and Masulis 2015):

$$\begin{aligned} Forced_t \quad = \quad & \beta_1\, Firm\ Ret_{t-1} + \beta_2\, Peer\ Ret_{t-1} + \beta_3\, Size_{t-1} + \beta_4\, BM_{t-1} + \beta_5\, Vol_{t-1} \\ & + \beta_6\, Tenure_t + \beta_7\, Age_t + \beta_8\, Ownership_{t-1} + \varepsilon_{i,t}. \end{aligned} \tag{2}$$

The dependent variable, *Forced*$_t$, is an indicator variable equal to one if a forced CEO turnover occurs in period *t*, and zero otherwise. We use hand-collected forced CEO turnover data for all ExecuComp CEOs for the period 1996 to 2015 (see Peters and Wagner, 2014).[17] *Firm Ret*$_{t-1}$ is measured as the natural logarithm of one plus annual buy-and-hold stock return in period *t-1*. If a CEO turnover occurs in period *t*, annual returns are measured over a period that covers the 12 months before the CEO departure date. *Peer Ret*$_{t-1}$ is measured as the natural logarithm of one plus equal-weighted annual average returns based on the characteristics-matched TNIC peers over the same period that *Firm Ret*$_{t-1}$ is measured.

Panel A of Table 9 provides descriptive statistics of CEO turnover. We first note that the mean value of forced CEO turnover in our sample is approximately 2%, which is similar to that found in prior studies (Guo and Masulis 2015; Jenter and Kanaan 2015; Peters and Wagner 2014). Panel B presents Pearson correlations among main variables. We note that the *Forced*$_t$ variable is negatively correlated with *Firm Ret*$_{t-1}$, suggesting that poor stock performance is associated with forced CEO turnover decisions.

Table 10 demonstrates the estimation results using Equation 2. Similar to the analyses in Table 3, we start with the full-sample results in column 1. Consistent with CEOs fired for the poor performance, the coefficient on *Firm Ret*$_{t-1}$ is negative (coefficient of -0.038) and significant at the 1% level. The coefficient estimate suggests that a one standard deviation decrease in *Firm Ret*$_{t-1}$ is associated with an increase in the forced CEO turnover likelihood of 1.7%, which represents an 85.5% increase relative to the mean. We find that the coefficient on *Peer Ret*$_{t-1}$ is positive (coefficient of 0.022) and also significant at the 1% level, consistent with our first RPE hypothesis. The evidence supports the weak-form version of RPE in the forced CEO turnover decisions, which is consistent with findings in prior research (Jenter and Kanaan 2015).

[17] The procedure to classify turnovers as forced follows Parrino (1997) and uses press reports along with an age criterion and further refinements. See Peters and Wagner (2014) for details.

We next test our second prediction that the extent of common performance filtering increases with the number of peers by dividing the sample into three subsamples based on the tercile of the number of TNIC peers in period *t-1* and estimate Equation 2 within each subsample. Here again, we find the coefficient on *Peer Ret*$_{t-1}$ to become increasingly more positive as we move from the Few group to the Moderate group to the Many group. The coefficient on *Peer Ret*$_{t-1}$ is 0.007 in the Few peer group and statistically insignificant, 0.028 in the Moderate peer group significant at the 1% level, and 0.043 in the Many peer group significant at the 1% level. The coefficient on *Peer Ret*$_{t-1}$ is significantly different between the Few and the Moderate groups at the10% level (difference of 0.021) and also between the Few and the Many groups at the 1% level (difference of 0.036). Similar to the pay results in Table 3, the coefficient on *Firm Ret*$_{t-1}$ does not vary significantly across these subsamples (-0.038, -0.047 and -0.038).

In column 4, we find evidence consistent with our third prediction of complete filtering in the forced CEO turnover decisions. The sum of the coefficients on *Firm Ret*$_{t-1}$ and *Peer Ret*$_{t-1}$ is indistinguishable from zero (*F*-statistic = 0.170, *p*-value = 0.681), suggesting that boards completely filter out common performance when deciding on CEO replacement in an environment with many peer firms. Overall, these results support our RPE hypotheses and corroborate the CEO pay results documented in Table 3.

Similar to Table 4, we replicate our results using alternative industry classifications. Table 11 reports the replication results. In Panel A, we use three-digit SIC codes and calculate *SIC Peer Ret*$_{t-1}$ based on the same method used in the construction of our main peer return variable using TNIC, *Peer Ret*$_{t-1}$. In column 1, we find the same result documented in column 1 of Table 10, i.e., weak-form evidence of RPE in the forced CEO turnover decisions. This result is consistent with Jenter and Kanaan (2015). Next, we partition the sample into terciles based on of the number of firms in the same SIC industry, and estimate Equation 2 within each subsample. We find evidence consistent with our two predictions. In particular, the coefficient on *SIC Peer Ret*$_{t-1}$ monotonically decreases as we move from the Moderate group to the Many group (coefficient difference of 0.022; *p*-value 0.025). We also find evidence consistent with complete filtering in the Many peers subsample. In particular, the coefficient on *Firm Ret*$_{t-1}$ is -0.038 and that on *SIC Peer Ret*$_{t-1}$ is -0.032, and the sum of the two coefficients is statistically indistinguishable from zero (*F*-statistic 0.310; *p*-value 0.579).

In Panel B, we use GICS codes to define peers and compute *GICS Peer Ret*$_{t-1}$. Here again, we find evidence consistent with weak-form RPE in the full sample. In the subsample analysis, the most positive and significant one is found in the moderate peers subsample, which is consistent with the results in Panel B of Table 4. In the moderate peer subsample, we find evidence of complete filtering of common performance in the forced CEO turnover decisions.

Findings in Table 11 suggest that evidence of strong-form RPE in forced CEO turnover decisions can be found using alternative industry classifications. As we mentioned before, we conjecture that this might be due to these alternative proxies being correlated with product market peers. Similar to Table 5, we examine our conjecture by including both *Peer Ret*$_{t-1}$, which is based on TNIC, and *SIC Peer Ret*$_{t-1}$ or *GICS Peer Ret*$_{t-1}$ that is based on these alternative pre-defined industry classifications simultaneously in the same regression.

Table 12 presents the estimation results comparing peer return variables in the forced CEO turnover regressions. In column 1, we include *Peer Ret*$_{t-1}$ and *SIC Peer Ret*$_{t-1}$ simultaneously in the same regression and find that the coefficient on *Peer Ret*$_{t-1}$ is significantly positive at the 1% level, and the coefficient on *SIC Peer Ret*$_{t-1}$ is statistically significant at the 10% level. In column 2, we replace *SIC Peer Ret*$_{t}$ with *GICS Peer Ret*$_{t}$, and we find that the coefficient on *Peer Ret*$_{t-1}$ is significantly positive at the 5% level and the coefficient on *GICS Peer* Ret_{t-1} is significantly positive at the 1% level. In column 3, we include all three peer return variables and find that *Peer* Ret_{t-1} and *GICS Peer* Ret_{t-1} are significantly positively correlated with the dependent variable at the 10% level and the 1% level, respectively. These results suggest that the peer return variables using the characteristics-matched SIC and GICS peers provide information regarding common shocks additional to that provided by the characteristics-matched TNIC peers.

Next, we use all firms in the same industry group to construct the peer return variable in columns 4-7. In contrast to the findings in columns 1-3, we find that only the coefficient on *All Peer* Ret_{t-1} remains positive and statistically significant at the 1% level, while those on *All SIC Peer* Ret_{t-1} and *All GICS Peer* Ret_{t-1} become statistically insignificant in columns 4-6. In column 7, we include the peer return variables based on the characteristics-matched TNIC peers (*Peer* Ret_{t-1}) and the peer return variable based on all TNIC peers (*All Peer* Ret_{t-1}) simultaneously in the same regression. We find that the coefficient on *Peer* Ret_{t-1} becomes insignificant while the coefficient on *All Peer* Ret_{t-1} remains statistically significant at the 1% level.[18]

Collectively, the findings in Table 12 confirm our hypothesis that TNIC provides a better identification of peers in the forced CEO turnover setting. The evidence suggests that corporate boards seem to use the entire set of product market peers in evaluating firm performance, especially when they have to make a CEO retention decision rather than to use a smaller subset of peer firms with similar firm-specific characteristics.

[18] We re-examine results in Table 10 and Table 11 using peer return variables based on all peer firms in each industry classification and find similar results.

V. Conclusion

This study re-examines the RPE hypothesis using product-market peers identified by a textual analysis of firms' product descriptions in 10-K filings (Hoberg and Phillip, 2016). In contrast to the mixed evidence of RPE documented in prior studies, we find three pieces of evidence consistent with RPE in both CEO pay and forced CEO turnover decisions – (i) firms on average filter out common shocks to performance measures, (ii) the extent of filtering increases with the number of peers, and (iii) firms completely filter out common shocks in the presence of a large number of peers. We can replicate the first finding but not the other two using the pre-defined industry classifications such as SIC and GICS especially in the CEO pay regressions. We find evidence that product market peers, in general, provide a better identification than pre-defined industry classifications. Overall, our results suggest that a key identification strategy to testing RPE theory lies in accurately defining the peer group.

APPENDIX: Variable Definitions

Variable	Definition
Age_t	*Age* is defined as CEO age variable in ExecuComp in period *t*.
BM_{t-1}	BM_{t-1} is measured as firm *i*'s Book-to-Market ratio as of the beginning of period *t*. Book-to-Market is measured as the book value of equity divided by market value of equity. Book value of equity is measured by shareholders' equity plus deferred tax and investment credit minus preferred stock. The market value of equity is obtained from CRSP and is calculated by the number of common shares outstanding multiplied by share price.
$Firm\ Ret_t$	$Firm\ Ret_t$ is measured as the natural logarithm of one plus firm *i*'s annual buy-and-hold stock return in period *t*.
$Forced_t$	$Forced_t$ is an indicator equal to one if a forced CEO turnover occurs in period *t*, zero otherwise. Forced CEO turnover is identified following Parrino (1997) and Peters and Wagner (2014).
$lnDelta_{t-1}$	$lnDelta_{t-1}$ is the natural logarithm of one plus portfolio delta for the CEO at the beginning of the period *t*. Portfolio delta measures the dollar change in wealth experienced by the CEO for a 1% change in the firm's stock price (Core and Guay, 2002; Coles et al., 2006).
$\#\ of\ Peers_t$	*# of Peers* is measured as the number of firms in the same industry group for firm *i* in period *t*.
$Ownership_{t-1}$	$Ownership_{t-1}$ is calculated as the number of shares owned by CEO excluding option divided by the number of shares outstanding for firm *i* as of the beginning of period *t*.
$Peer\ Ret_t$	*Peer Ret* is measured as the natural logarithm of one plus equal-weighted annual returns of firm *i*'s characteristics-matched TNIC peers in period *t*. To define characteristics-matched TNIC peers, we choose one-quarter of TNIC peers based on the closeness of the Mahalanobis distance using the market value of equity (i.e., Size) and book-to-market as of the beginning of the fiscal period. We require firm *i* to have a minimum of two peer firms in each period.
$SIC\ Peer\ Ret_t$	*SIC Peer Ret* is measured as the natural logarithm of one plus equal-weighted annual returns of firm *i*'s characteristics-matched SIC peers in period *t*. To define characteristics-matched SIC peers, we choose one-quarter of firms in the same three-digit SIC industry based on the closeness of the Mahalanobis distance using the market value of equity (i.e., Size) and book-to-market as of the beginning of the fiscal period. We require firm *i* to have a minimum of two peer firms in each period.

GICS Peer Ret$_t$	*GICS Peer Ret* is measured as the natural logarithm of one plus equal-weighted annual returns of firm *i*'s characteristics-matched GICS peers in period *t*. To define characteristics-matched GICS peers, we choose one-quarter of firms in the same eight-digit GICS industry based on the closeness of the Mahalanobis distance using market value of equity (i.e., Size) and book-to-market as of the beginning of the fiscal period. We require firm *i* to have a minimum of two peer firms in each period.
Size$_{t-1}$	*Size*$_{t-1}$ is measured as the natural logarithm of one plus firm *i*'s total revenue in period *t-1*.
Tenure$_t$	*Tenure*$_t$ is defined as the natural logarithm of one plus the difference between the *BECAMECEO* variable in ExecuComp and the date of fiscal year-end for firm *i* as of the beginning of period *t* divided by 365.
Total Comp$_t$	*Total Comp* is *TDC1* in ExecuComp, which is measured by the sum of salary, bonus, long-term incentive payouts, the fair value of stock and option grants, and all other compensation for firm *i* in period *t*.
ln(*Total Comp*$_t$)	*ln*(*Total Comp*) is measured as the natural logarithm of one plus total CEO compensation for firm *i* in period *t*.
Vol$_{t-1}$	*Vol*$_{t-1}$ measures idiosyncratic return volatility and defined as the standard deviations of residuals from the regression of firm *i*'s monthly returns on monthly equal-weighted peer firm average returns using preceding past 12 months (a minimum of 6 observations is required).

REFERENCES

Abowd, J. M., and Kaplan, D. S., 1999. Executive Compensation: Six Questions That Need Answering. *Journal of Economic Perspectives*, *13*(4), 145-168.

Aggarwal, R., and Samwick, A., 1999a. The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation. *Journal of Political Economy* 107, no. 1: 65–105.

Aggarwal, R., and Samwick, A., 1999b. Executive Compensation, Strategic Competition, and Relative Performance Evaluation: Theory and Evidence. *The Journal of Finance* 54, no. 6.

Albuquerque, A., 2009. Peer Firms in Relative Performance Evaluation. *Journal of Accounting and Economics* 48, no. 1: 69–89.

Albuquerque, A., 2014. Do growth-option firms use less relative performance evaluation? *The Accounting Review*, 89(1), 27-60.

Albuquerque, A. M., De Franco, G., and Verdi, R. S. 2013. Peer choice in CEO compensation. *Journal of Financial Economics*, 108(1), 160-181.

Ali, A., Klasa, S., and Yeung, E., 2009. The limitations of industry concentration measures constructed with Compustat data: Implications for finance research. *Review of Financial Studies*, *22*(10), 3839-3871.

Antle, R., and Smith, A., 1986. An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives. *Journal of Accounting Research* 24, no. 1: 1–39.

Bannister, J., and Newman, H., 2003. Analysis of Corporate Disclosures on Relative Performance Evaluation. *Accounting Horizons* 17, no. 3: 235–246.

Barro, J. R., and Barro, R. J., 1990. Pay, Performance, and Turnover of Bank CEOs. *Journal of Labor Economics*, 448-481.

Beaver, W., McNichols, M., and Price, R., 2007. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics*, *43*(2), 341-368.

Bernard, V., and Skinner, D., 1996. What Motivates Managers’ Choice of Discretionary Accruals? *Journal of Accounting and Economics* 22, no. 1: 313–325.

Bertrand, M., and Mullainathan, S., 2001. Are CEOs Rewarded for Luck? The Ones Without Principals Are. *The Quarterly Journal of Economics* 116, no. 3: 901–932.

Bhojraj, S., Lee, C., and Oler, D., 2003. What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research* 41, no. 5: 745–774.

Bizjak, J. M., 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.

Brickley, J., and Zimmerman, J., 2010. Corporate Governance Myths: Comments on Armstrong, Guay, and Weber. *Journal of Accounting and Economics* 50, no. 2: 235–245.

Bushman, R., and Smith, A., 2001. Financial Accounting Information and Corporate Governance. *Journal of Accounting and Economics* 32, no. 1: 237–333.

Carter, M., Ittner, D., and Zechman, S., 2009. Explicit Relative Performance Evaluation in Performance-vested Equity Grants. *Review of Accounting Studies* 14, no. 2–3: 269–306.

Clarke, R., 1989. SICs as Delineators of Economic Markets. *Journal of Business*: 17–31.

Coles, J.L., Daniel, N.D., and Naveen, L., 2006. Managerial Incentives and Risk-Taking. *Journal of Financial Economics* 79, 431-468

Core, J., Guay, W., 2002. Estimating the value of employee stock option portfolios and their sensitivities to price and volatility. *Journal of Accounting Research* 40, 613-630.

Cornelli, F., Kominek, Z., and Ljungqvist, A. 2013. Monitoring managers: Does it matter?. *The Journal of Finance*, 68(2), 431-481.

DeFond, M., and Park, C., 1999. The Effect of Competition on CEO Turnover. *Journal of Accounting and Economics* 27, no. 1: 35–56.

Dikolli, S., Hofmann, C., and Pfeiffer, T., 2011. Relative Performance Evaluation and Peer-performance Summarization Errors. *Review of Accounting Studies*: 1–32.

Dopuch, N., Mashruwala, R., Seethamraju, C., and Zach, T., 2012. The Impact of a Heterogeneous Accrual-generating Process on Empirical Accrual Models. *Journal of Accounting, Auditing & Finance* 27, no. 3: 386–411.

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.

Foucault, T., and Fresard, L., 2014. Learning from peers' stock prices and corporate investment. *Journal of Financial Economics*, 111(3), 554-577.

Frydman, C., and Jenter, D., 2010. CEO Compensation. *Annual Review of Financial Economics*, *2*(1), 75-102.

Garvey, G., and Milbourn, T., 2003. Incentive Compensation When Executives Can Hedge the Market: Evidence of Relative Performance Evaluation in the Cross Section. *The Journal of Finance* 58, no. 4: 1557–1582.

Garvey, G., and Milbourn, T., 2006, Asymmetric Benchmarking in Compensation: Executives Are Rewarded for Good Luck but Not Penalized for Bad. *Journal of Financial Economics* 82, no. 1: 197–225.

Gibbons, R., and Murphy, K., 1990. Relative Performance Evaluation for Chief Executive Officers. *Industrial and Labor Relations Review*: 30S–51S.

Gormley, T. A., and Matsa, D. A., 2014. Common errors: How to (and not to) control for unobserved heterogeneity. *Review of Financial Studies*, 27(2), 617-661.

Gong, G., Li, L. Y., and Shin, J. Y., 2011. Relative Performance Evaluation and Related Peer Groups in Executive Compensation Contracts. *The Accounting Review* 86, no. 3 (2011): 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, no. 5: 2060–2098.

Guenther, D., and Rosman, A., 1994. Differences Between COMPUSTAT and CRSP SIC Codes and Related Effects on Research. *Journal of Accounting and Economics* 18, no. 1: 115–128.

Guo, L., and Masulis, R. W. 2015. Board structure and monitoring: New evidence from CEO turnovers. *The Review of Financial Studies*, 28(10), 2770-2811.

Hoberg, G., and Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, *23*(10), 3773-3811.

Hoberg, G., and Phillips, G., 2016. Text-Based Network Industries and Endogenous Product Differentiation, *Journal of Political Economy*, 124 (5), 1423-1465.

Holmstrom, B., 1979. Moral Hazard and Observability. *The Bell Journal of Economics*: 74–91.

Holmstrom, B., 1982. Moral Hazard in Teams. *The Bell Journal of Economics*: 324–340.

Holmstrom, B., and Milgrom, P., 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica: Journal of the Econometric Society*: 303–328.

Janakiraman, S., Lambert, R., and Larcker, D., 1992. An Empirical Investigation of the Relative Performance Evaluation Hypothesis. *Journal of Accounting Research* 30, no. 1: 53–69.

Jensen, M., and Murphy, K., 1990. Performance Pay and Top-Management Incentives. *The Journal of Political Economy* 98, no. 2: 225–264.

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

Kahle, K. and Walkling, R., 1996. The Impact of Industry Classifications on Financial Research. *Journal of Financial and Quantitative Analysis* 31, no. 3.

Koh, P. S., and Reeb, D. M. 2015. Missing R&D. *Journal of Accounting and Economics*, 60(1), 73-94.

Lambert, R., 2001. Contracting Theory and Accounting. *Journal of Accounting and Economics* 32, no. 1: 3–87.

Lewellen, S., 2013, Executive compensation and peer effects, *working paper*, Yale University

Lys, T., and Sabino, J. S., 1992. Research design issues in grouping-based tests. *Journal of Financial Economics*, *32*(3), 355-387.

Milbourn, T., 2003. CEO Reputation and Stock-based Compensation. *Journal of Financial Economics* 68, no. 2: 233–262.

Murphy, K., 1999. Executive Compensation. *Handbook of Labor Economics* 3: 2485–2563.

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

Peters, F. S., and Wagner, A. F. 2014. The executive turnover risk premium. *The Journal of Finance*, 69(4), 1529-1563.

Petersen, M. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22, 435-480.

Prendergast, C., 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37, no. 1: 7–63.

Rajgopal, S., Shevlin, T., and Zamora, V., 2006. CEOs' Outside Employment Opportunities and the Lack of Relative Performance Evaluation in Compensation Contracts. *The Journal of Finance* 61, no. 4: 1813–1844.

Smith, C. W., and Watts, R. L., 1992. The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics*, *32*(3), 263-292.

Table 1
Descriptive Statistics

Panel A presents descriptive statistics for main variables. *Total Comp*$_t$ is total CEO compensation in period *t* (*TDC1* in ExecuComp). *ln*(*Total Comp*$_t$) is measured as the natural logarithm of one plus total CEO compensation in period *t*. *Firm Ret*$_t$ is measured as the natural logarithm of one plus annual buy-and-hold stock return in period *t*. *Peer Ret*$_t$ is measured as the natural logarithm of one plus equal-weighted annual average returns based on characteristics-matched TNIC peers in period *t*. *SIC Peer Ret*$_t$ and *GICS Peer Ret*$_t$ are measured using the same method as *Peer Ret*$_t$ except for the use of SIC and GICS industries, respectively. *# of Peers*$_t$ is measured as the total number of TNIC peers in period *t*. *# of SIC Peers*$_t$ and *# of GICS Peers*$_t$ are the total number of SIC and GICS peers in period *t*, respectively. *Size*$_{t-1}$ is measured as the natural logarithm of one plus total revenues in period *t-1*. *BM*$_{t-1}$ is measured as the Book-to-Market ratio as of the beginning of period *t*. *Tenure*$_{t-1}$ is defined as the natural logarithm of one plus the difference between the *BECAMECEO* variable in ExecuComp and the date of fiscal year-end in period *t* divided by 365. *Age*$_t$ is defined as the age of CEO in period *t*. *Ownership*$_{-1}$ is calculated as the number of shares owned by CEO excluding option divided by the number of shares outstanding as of the beginning of period *t*. *lnDelta*$_{t-1}$ is the natural logarithm of one plus portfolio delta for the CEO at the beginning of the period t. Portfolio delta measures the dollar change in wealth experienced by the CEO for a 1% change in the firm's stock price. Panel B presents Pearson correlations Correlations that are significant at 5% level are bolded. The sample period is between 1996 and 2015. All variables are defined in the Appendix A.

Panel A Descriptive Statistics

	N	Mean	Std	Q1	Median	Q3
Total Comp$_t$	26,187	5,453.41	9,884.44	1,495.08	3,188.28	6,458.25
ln(*Total Comp*$_t$)	26,187	8.04	1.06	7.31	8.07	8.77
Firm Ret$_t$	26,187	0.05	0.45	-0.15	0.09	0.30
Peer Ret$_t$	26,187	0.06	0.32	-0.09	0.09	0.25
SIC Peer Ret$_t$	25,524	0.05	0.33	-0.10	0.09	0.25
GICS Peer Ret$_t$	25,846	0.06	0.32	-0.09	0.10	0.25
# of Peers$_t$	26,187	82.18	113.27	16.00	41.00	94.00
# of SIC Peers$_t$	25,524	88.01	119.87	12.00	28.00	121.00
# of GICS Peers$_t$	25,846	69.01	93.70	20.00	42.00	79.00
Size$_{t-1}$	26,187	7.24	1.62	6.14	7.15	8.31
BM $_{t-1}$	26,187	0.57	0.45	0.28	0.48	0.75
Vol$_{t-1}$	26,187	0.09	0.05	0.05	0.08	0.11
Tenure$_t$	26,187	1.91	0.77	1.34	1.91	2.47
Age$_t$	26,187	55.73	7.29	51.00	56.00	60.00
Ownership$_{t-1}$	26,187	0.02	0.06	0.00	0.00	0.01
lnDelta$_{t-1}$	26,187	4.99	1.97	4.07	5.19	6.23

Panel B Pearson Correlations

	Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	*ln*(*Total Comp*$_t$)	-										
(2)	*Firm Ret*$_t$	**0.08**	-									
(3)	*Peer Ret*$_t$	**0.04**	**0.58**	-								
(4)	*SIC Peer Ret*$_t$	**0.05**	**0.55**	**0.74**	-							
(5)	*GICS Peer Ret*$_t$	**0.04**	**0.58**	**0.77**	**0.78**	-						
(6)	*Size*$_{t-1}$	**0.61**	**0.01**	0.04	0.03	0.03	-					
(7)	*BM* $_{t-1}$	**-0.15**	**0.08**	**0.12**	**0.10**	**0.11**	**0.02**	-				
(8)	*Vol* $_{t-1}$	**-0.24**	**-0.03**	**-0.02**	-0.01	0.00	**-0.38**	**0.06**	-			
(9)	*Tenure*$_t$	**-0.05**	**0.03**	**0.02**	**0.02**	**0.02**	**-0.09**	**-0.04**	**-0.03**	-		
(10)	*Age*$_t$	**0.03**	**0.03**	**0.04**	**0.04**	**0.03**	**0.12**	**0.05**	**-0.13**	**0.39**		-
(11)	*Ownership*$_{t-1}$	**-0.23**	0.01	-0.01	**-0.02**	0.00	**-0.14**	**-0.03**	**0.10**	**0.34**	**1.00**	
(12)	*lnDelta*$_{t-1}$	**0.37**	**-0.03**	**-0.04**	**-0.04**	**-0.04**	**0.33**	**-0.26**	**-0.16**	**0.36**	**0.13**	**0.23**

Table 2
Sales / Costs Correlations

Panel A presents estimation results from the regression of firm *i*'s sales ($Sales_t$) in period *t* on average sales using all peer firms in TNIC ($Avg\ Sales\ TNIC_t$), SIC ($Avg\ Sales\ SIC_t$), or GICS ($Avg\ Sales\ GICS_t$) in period *t* excluding firm *i*. Panel B presents estimation results from the regression of firm *i*'s operating costs in period t ($Costs_t$), which is the sum of cost of goods sold and SG&A expenses on average operating costs using all peer firms in TNIC ($Avg\ Costs\ TNIC_t$), SIC ($Avg\ Costs\ SIC_t$), or GICS ($Avg\ Costs\ GICS_t$) in period *t* excluding firm *i*. In each panel, column (1) reports estimation results using the full sample, and column (2) through (5) present estimation results conditional on time periods denoted in each column. The sample period is between 1996 and 2015. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

Panel A Sales Correlations

	Dependent Variable: $Sales_t$				
Independent Variables	(1)	(2)	(3)	(4)	(5)
	Full Sample	1996 - 2000	2001 - 2005	2006 - 2010	2011 - 2015
Avg Sales TNIC$_t$	**0.785*** **	**0.816*** **	**0.842*** **	**0.792*** **	**0.634*** **
	(35.982)	**(30.641)**	**(29.983)**	**(22.538)**	**(16.251)**
Avg Sales SIC$_t$	0.249***	0.231***	0.208***	0.235***	0.366***
	(10.813)	(8.267)	(7.191)	(6.759)	(9.543)
Avg Sales GICS$_t$	0.169***	0.152***	0.149***	0.186***	0.218***
	(7.484)	(5.208)	(5.218)	(4.996)	(5.156)
# of observations	74,004	23,428	19,973	16,379	14,224
Adjusted R-squared	0.540	0.518	0.547	0.552	0.550

Panel B Cost Correlations

	Dependent Variable: $Costs_t$				
Independent Variables	(1)	(2)	(3)	(4)	(5)
	Full Sample	1996 - 2000	2001 - 2005	2006 - 2010	2011 - 2015
Avg Costs TNIC$_t$	**0.933*** **	**0.834*** **	**0.965*** **	**0.958*** **	**0.944*** **
	(50.656)	**(27.850)**	**(38.258)**	**(32.824)**	**(31.083)**
Avg Costs SIC$_t$	0.164***	0.250***	0.113***	0.158***	0.162***
	(9.959)	(8.452)	(5.108)	(6.005)	(7.259)
Avg Costs GICS$_t$	0.133***	0.149***	0.134***	0.133***	0.139***
	(7.866)	(4.685)	(6.385)	(4.779)	(5.440)
# of observations	74,004	18,336	19,973	16,379	14,224
Adjusted R-squared	0.433	0.424	0.443	0.432	0.428

Table 3
Tests of Relative Performance Evaluation Hypothesis

This table presents the estimation results from the following regression model.

$$ln(Total\ Comp_t) = \beta_1\ Firm\ Ret_t + \beta_2\ Peer\ Ret_t + \beta_3\ Size_{t-1} + \beta_4\ BM_{t-1} + \beta_5\ Vol_{t-1} + \beta_6\ Tenure_t + \beta_7 Age + \beta_8\ Ownership_{t-1} + \beta_9\ lnDelta_{t-1} + \varepsilon_t$$

In column 1, the estimation uses the full sample. In column 2, 3, and 4, the full sample is divided into three subsamples based on the tercile of the number of TNIC peers. Results testing strong-form evidence of RPE and the coefficient differences are summarized toward the bottom of the table. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent Variable: *ln*(*Total Comp*$_t$)			
Independent Variables	(1)	(2)	(3)	(4)
		# of TNIC Peers		
	Full Sample	Few	Moderate	Many
Firm Ret$_t$	0.211***	0.214***	0.224***	0.207***
	(13.684)	(9.256)	(6.759)	(7.051)
***Peer Ret*$_t$**	**-0.101*** **	**-0.040**	**-0.110*** **	**-0.203*** **
	(-5.036)	**(-1.468)**	**(-2.629)**	**(-4.113)**
Size$_{t-1}$	0.243***	0.278***	0.198***	0.217***
	(9.344)	(7.124)	(3.988)	(5.331)
BM$_{t-1}$	-0.309***	-0.270***	-0.275***	-0.342***
	(-13.386)	(-9.349)	(-5.512)	(-6.877)
Vol$_{t-1}$	-0.220	-0.233	-0.437	-0.020
	(-1.453)	(-0.858)	(-1.536)	(-0.073)
Tenure$_t$	0.009	0.062	-0.018	-0.033
	(0.283)	(1.273)	(-0.291)	(-0.489)
Age$_t$	-0.016	0.006	-0.024	-0.036
	(-1.030)	(0.459)	(-0.649)	(-1.168)
Ownership$_{t-1}$	-0.305	-0.787*	-0.154	0.060
	(-1.187)	(-1.734)	(-0.289)	(0.129)
lnDelta$_{t-1}$	0.007	-0.003	0.012	0.010
	(0.980)	(-0.300)	(0.752)	(0.798)
Strong RPE F-Stat	**30.990**	**40.550**	**8.020**	**0.010**
p-value ($\beta_1 + \beta_2 = 0$)	**0.000**	**0.000**	**0.005**	**0.925**
Firm-CEO FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of observations	26,187	8,889	8,623	8,675
Adjusted R-squared	0.761	0.788	0.775	0.738
Coefficient Difference	$\Delta\beta_1$	p-value	$\Delta\beta_2$	p-value
Few versus Moderate	0.009	(0.782)	-0.069	(0.102)
Moderate versus Many	-0.016	(0.660)	-0.094*	(0.090)
Few versus Many	-0.007	(0.827)	-0.163***	(0.001)

Table 4
RPE tests using alternative industry classifications

This table presents the estimation results from the following regression model.

$$ln(Total\ Comp_t) = \beta_1\, Firm\ Ret_t + \beta_2\, Alternative\ Peer\ Ret_t + \beta_3\, Size_{t-1} + \beta_4\, BM_{t-1} + \beta_5\, Vol_{t-1} + \beta_6\, Tenure_t + \beta_7\, Age + \beta_8\, Ownership_{t-1} + \beta_9\, lnDelta_{t-1} + \varepsilon_t$$

In Panel A, the *Alternative Peer Ret*$_t$ is *SIC Peer Ret*$_t$ that uses three-digit SIC industries to define peers. In Panel B, the *Alternative Peer Ret*$_t$ is *GICS Peer Ret*$_t$ that uses six-digit GICS industries to define peers. In each panel, column 1 presents estimation resulting using the full sample, and in columns 2-4, the full sample is divided into three subsamples based on the tercile of the number of industry peers that is used in each panel. Results testing strong-form evidence of RPE and the coefficient differences are summarized toward the bottom of the table. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

Panel A RPE tests using SIC industries

	Dependent Variable: $ln(Total\ Comp_t)$			
Independent Variables	(1)	(2)	(3)	(4)
		# of SIC Peers		
	Full Sample	Few	Moderate	Many
Firm Ret$_t$	0.202***	0.241***	0.249***	0.134***
	(12.861)	(10.299)	(9.821)	(4.479)
***SIC Peer Ret*$_t$**	**-0.073*** **	**-0.042**	**-0.116*** **	**-0.058**
	(-4.007)	**(-1.576)**	**(-3.123)**	**(-1.517)**
Size$_{t-1}$	0.245***	0.219***	0.244***	0.247***
	(9.231)	(6.197)	(6.425)	(5.498)
BM$_{t-1}$	-0.310***	-0.284***	-0.314***	-0.298***
	(-13.164)	(-9.599)	(-8.788)	(-5.207)
Vol$_{t-1}$	-0.179	-0.492**	-0.379	-0.069
	(-1.155)	(-1.982)	(-1.285)	(-0.250)
Tenure$_t$	0.010	0.064	0.103**	-0.108
	(0.298)	(1.436)	(2.101)	(-1.480)
Age$_t$	-0.015	-0.029	-0.008	-0.031
	(-0.930)	(-0.846)	(-0.418)	(-0.882)
Ownership$_{t-1}$	-0.310	-0.313	-0.589	-0.182
	(-1.177)	(-0.896)	(-1.059)	(-0.394)
lnDelta$_{t-1}$	0.006	-0.000	-0.014	0.016
	(0.856)	(-0.038)	(-1.336)	(1.250)
Strong RPE F-Stat	**47.620**	**55.360**	**14.290**	**4.230**
p-value ($\beta_1 + \beta_2 = 0$)	**0.000**	**0.000**	**0.000**	**0.040**
Firm-CEO FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of observations	25,524	8,661	8,424	8,439
Adjusted R-squared	0.761	0.793	0.777	0.736
Coefficient Difference	$\Delta\beta_1$	p-value	$\Delta\beta_2$	p-value
Few versus Moderate	0.008	(0.799)	-0.074	(0.061)
Moderate versus Many	-0.114	(0.001)	0.058	(0.218)
Few versus Many	-0.107	(0.001)	-0.016	(0.690)

Panel B RPE tests using GICS industries

Independent Variables	(1)	(2)	(3)	(4)
	Dependent Variable: $ln(Total\ Comp_t)$			
		# of GICS Peers		
	Full Sample	Few	Moderate	Many
$Firm\ Ret_t$	0.206***	0.225***	0.227***	0.160***
	(13.345)	(8.959)	(8.309)	(6.246)
GICS Peer Ret $_t$	**-0.084*** **	**-0.051***	**-0.124*** **	**-0.070***
	(-4.197)	**(-1.659)**	**(-3.302)**	**(-1.687)**
$Size_{t-1}$	0.245***	0.218***	0.191***	0.270***
	(9.387)	(3.756)	(4.443)	(6.207)
BM_{t-1}	-0.312***	-0.240***	-0.384***	-0.339***
	(-13.461)	(-6.065)	(-9.777)	(-7.519)
Vol_{t-1}	-0.214	-0.565**	-0.797***	0.378
	(-1.412)	(-2.250)	(-2.956)	(1.483)
$Tenure_t$	0.012	0.023	0.026	-0.002
	(0.372)	(0.386)	(0.493)	(-0.032)
Age_t	-0.014	-0.019	-0.007	-0.000
	(-0.821)	(-0.537)	(-0.140)	(-0.007)
$Ownership_{t-1}$	-0.287	-0.519	-0.475	-0.073
	(-1.120)	(-1.098)	(-0.985)	(-0.162)
$lnDelta_{t-1}$	0.007	0.004	-0.017	0.006
	(0.888)	(0.257)	(-1.587)	(0.473)
Strong RPE F-Stat	**38.870**	**25.760**	**8.770**	**6.140**
p-value ($\beta_1 + \beta_2 = 0$)	**0.000**	**0.000**	**0.003**	**0.013**
Firm-CEO FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of observations	25,846	8,745	8,521	8,580
Adjusted R-squared	0.761	0.791	0.770	0.747
Coefficient Difference	$\Delta\beta_1$	p-value	$\Delta\beta_2$	p-value
Few versus Moderate	0.002	(0.960)	-0.072	(0.088)
Moderate versus Many	-0.067	(0.039)	0.054	(0.266)
Few versus Many	-0.065	(0.038)	-0.019	(0.678)

Table 5
Comparison with alternative industry classifications

This table presents the estimation results from the following regression model.

$$ln(Total\ Comp_t) = \beta_1\ Firm\ Ret_t + \beta_2\ Peer\ Ret_t + \beta_3\ Alternative\ Peer\ Ret_t + \beta_4\ Size_{t-1} + \beta_5\ BM_{t-1} + \beta_6\ Vol_{t-1} + \beta_7\ Tenure_t + \beta_8\ Age + \beta_9\ Ownership_{t-1} + \beta_9\ lnDelta_{t-1} + \varepsilon_t$$

In column 1, the *Alternative Peer Ret*$_t$ is *SIC Peer Ret*$_t$ that uses SIC industries to define peers. In column 2, the *Alternative Peer Ret*$_t$ is *GICS Peer Ret*$_t$ that uses GICS industries to define peers. In column 3, *SIC Peer Ret*$_t$ and *GICS Peer Ret*$_t$ are included in the same regression simultaneously. In columns 4-6, we use all firms in the same industry group to construct peer return variables rather than characteristics-matched peers. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent variable: *ln*(*Total Comp*$_t$)						
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm Ret$_t$	0.206***	0.206***	0.207***	0.215***	0.217***	0.219***	0.212***
	(12.957)	(12.983)	(12.854)	(13.481)	(13.750)	(13.610)	(13.571)
***Peer Ret*$_t$**	**-0.114*** **	**-0.106*** **	**-0.116*** **	-	-	-	**-0.011**
	(-3.585)	**(-3.322)**	**(-3.342)**	-	-	-	**(-0.340)**
SIC Peer Ret$_t$	0.017	-	0.017	-	-	-	-
	(0.628)	-	(0.536)	-	-	-	-
GICS Peer Ret$_t$	-	0.014	0.002	-	-	-	-
	-	(0.467)	(0.059)	-	-	-	-
***All Peer Ret*$_t$**	-	-	-	**-0.089*** **	**-0.084*** **	**-0.082*** **	**-0.093*** **
	-	-	-	**(-3.648)**	**(-3.498)**	**(-3.212)**	**(-3.163)**
All SIC Peer Ret$_t$	-	-	-	-0.027	-	-0.015	-
	-	-	-	(-1.214)	-	(-0.597)	-
All GICS Peer Ret$_t$	-	-	-	-	-0.033	-0.026	-
	-	-	-	-	(-1.381)	(-0.953)	-
Controls	Included	Included	Included	Included	Included	Included	Included
Firm-CEO FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	25,524	25,846	25,224	25,524	25,846	25,224	26,187
Adjusted R-squared	0.761	0.761	0.761	0.761	0.762	0.761	0.761

Table 6
Dynamic peer groups

This table presents the estimation results from the following regression model.

$$ln(Total\ Comp_t) = \beta_1\, Firm\ Ret_t + \beta_2\, Dynamic\ Peer\ Ret_t + \beta_3\, Size_{t-1} + \beta_4\, BM_{t-1} + \beta_5\, Vol_{t-1} + \beta_6\, Tenure_t + \beta_7\, Age + \beta_8\, Ownership_{t-1} + \beta_9\, lnDelta_{t-1} + \varepsilon_t$$

Dynamic Peer Ret$_t$ is defined as equal-weighted average stock returns of past peers (*Past Peer Ret$_t$*), new peers (*New Peer Ret$_t$*), current peers (*Current Peer Ret$_t$*), or future peers (*Future Peer Ret$_t$*) for firm *i* as of period *t*. Past peers are firms that were used to construct the *Peer Ret$_t$* variable in the past period *t-1* but are not in the same product market in the current period *t*. New peers are firms that are used to construct the *Peer Ret$_t$* variable in the current period *t* but were not in the same product market in the past period *t-1*. Current peers are firms that are used in both the past period *t-1* and the current period *t* to construct the *Peer Ret$_t$* variable. Future peers are firms that will be used to construct the *Peer Ret$_t$* variable in the future period *t+1* but are not in the same product market in the current period *t*. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent Variable: *ln*(*Total Comp*)				
Independent Variables	(1)	(2)	(3)	(4)	(5)
Firm Ret$_t$	0.146***	0.193***	0.214***	0.224***	0.175***
	(4.554)	(5.751)	(8.585)	(5.740)	(4.367)
Past Peer Ret$_t$	**-0.046**	-	-	-	-
	(-1.114)	-	-	-	-
New Peer Ret$_t$	-	**-0.112****	-	**-0.025**	-
	-	**(-2.505)**	-	**(-0.394)**	-
Current Peer Ret$_t$	-	-	**-0.156*****	**-0.156****	-
	-	-	**(-4.594)**	**(-2.072)**	-
Future Peer Ret$_t$	-	-	-	-	**-0.070**
	-	-	-	-	**(-1.417)**
Controls	Included	Included	Included	Included	Included
Firm-CEO FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
# of observations	8,462	9,049	14,349	7,351	7,581
Adjusted R-squared	0.711	0.725	0.759	0.725	0.716

Table 7
Cross-sectional variation: Strategic interactions and RPE

This table presents the estimation results from the following regression model.

$$ln(Total\ Comp_t) = \beta_1\, Firm\ Ret_t + \beta_2\, Peer\ Ret_t + \beta_3\, Size_{t-1} + \beta_4\, BM_{t-1} + \beta_5\, Vol_{t-1} + \beta_6\, Tenure_t + \beta_7\, Age + \beta_8\, Ownership_{t-1} + \beta_9\, lnDelta_{t-1} + \varepsilon_t$$

In column 1 and 2, the full sample is divided into two subsamples based on the Competitive Strategic Measure (CSM). CSM is defined as the coefficient of correlation between the ratio of the change of a firm's profits to the change of its sales, and the change in the combined sales of its rivals. CSM captures the cross-partial derivative of firm value with regards to industry peers' strategic actions as measured by changes in sales. If the CSM has the positive value, it indicates that the competition is strategic complements; otherwise, the competition is strategic substitutes. In column 3 and 4, the full sample is divided into two subsamples based on the median value of the revenue-based Herfindahl-Hirschman Index (HHI) using TNIC peers. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent Variable: *ln*(*Total Comp*$_t$)			
Independent Variables	(1)	(2)	(3)	(4)
	Competitive Strategic Measure (*CSM*)		*HHI*	
	Complements	Substitute	Concentrated	Competitive
Firm Ret$_t$	0.199***	0.194***	0.203***	0.211***
	(8.526)	(7.966)	(8.822)	(8.760)
Peer Ret$_t$	-0.053	-0.144***	-0.072***	-0.165***
	(-1.636)	(-4.652)	(-2.606)	(-4.395)
Size$_{t-1}$	0.252***	0.231***	0.215***	0.254***
	(7.166)	(6.206)	(4.998)	(8.119)
BM$_{t-1}$	-0.317***	-0.298***	-0.301***	-0.286***
	(-9.012)	(-9.156)	(-8.420)	(-9.074)
Vol$_{t-1}$	-0.112	-0.098	-0.353	-0.121
	(-0.512)	(-0.428)	(-1.547)	(-0.549)
Tenure$_t$	-0.000	0.012	-0.012	0.065
	(-0.008)	(0.269)	(-0.230)	(1.446)
Age$_t$	-0.017	0.001	-0.024	-0.030
	(-0.675)	(0.031)	(-0.742)	(-1.398)
Ownership$_{t-1}$	-0.083	-0.476	-0.575	0.021
	(-0.243)	(-1.255)	(-1.520)	(0.066)
lnDelta$_{t-1}$	-0.000	0.008	-0.000	0.008
	(-0.014)	(0.955)	(-0.014)	(0.955)
Firm-CEO FE	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
# of observations	13,886	12,301	13,093	13,094
Adjusted R-squared	0.776	0.770	0.769	0.766
Coefficient difference	ΔCoeff.	p-value	ΔCoeff.	p-value
Firm Ret$_t$	-0.006	(0.832)	0.008	(0.769)
Peer Ret$_t$	-0.090	(0.016)	-0.093	(0.020)

Table 8
Cross-sectional variation: Strategic flexibility and RPE

This table presents the estimation results from the following regression model.

$$ln(Total\ Comp_t) = \beta_1 Firm\ Ret_t + \beta_2 Peer\ Ret_t + \beta_3 Size_{t-1} + \beta_4 BM_{t-1} + \beta_5 Vol_{t-1} + \beta_6 Tenure_t + \beta_7 Age + \beta_8 Ownership_{t-1} + \beta_9 lnDelta_{t-1} + \varepsilon_t$$

In column 1 and 2, the full sample is divided into two groups based on the median value of peer-adjusted market-to-book ratio in period *t*. In column 3 and 4, the full sample is divided into two groups based peer-adjusted annual stock return in period *t-1*. In column 5 and 6, the full sample is divided into two groups based on the median value of the firm's future asset growth rate in period *t+1*. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent Variable: $ln(Total\ Comp_t)$					
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Peer-Adjusted MTB		*Peer-Adjusted Return*		*Asset Growth*	
	Low	High	Negative	Positive	Low	High
$Firm\ Ret_t$	0.211***	0.195***	0.207***	0.270***	0.227***	0.217***
	(8.659)	(6.352)	(7.566)	(6.648)	(8.888)	(7.533)
$Peer\ Ret_t$	**-0.157*** **	**-0.037**	**-0.115*** **	**-0.023**	**-0.139*** **	**-0.065* **
	(-4.542)	**(-1.121)**	**(-3.292)**	**(-0.613)**	**(-3.880)**	**(-1.870)**
$Size_{t-1}$	0.222***	0.305***	0.272***	0.228***	0.231***	0.260***
	(6.150)	(8.520)	(8.186)	(5.318)	(5.323)	(6.144)
BM_{t-1}	-0.289***	-0.346***	-0.364***	-0.291***	-0.285***	-0.367***
	(-10.401)	(-8.481)	(-9.632)	(-7.501)	(-9.369)	(-7.310)
Vol_{t-1}	-0.205	-0.193	-0.102	-0.519**	-0.422*	-0.204
	(-0.907)	(-0.762)	(-0.401)	(-2.067)	(-1.729)	(-0.744)
$Tenure_t$	0.028	0.013	-0.015	0.040	0.000	0.067
	(0.664)	(0.267)	(-0.340)	(0.864)	(0.007)	(1.407)
Age_t	0.004	-0.050**	-0.017	-0.006	-0.048	-0.020
	(0.215)	(-1.963)	(-0.721)	(-0.166)	(-1.303)	(-0.751)
$Ownership_{t-1}$	-0.197	-0.361	-0.412	-0.172	-0.253	-0.313
	(-0.455)	(-1.073)	(-1.057)	(-0.383)	(-0.452)	(-0.982)
$lnDelta_{t-1}$	-0.000	0.008	0.011	0.017	-0.007	0.015
	(-0.014)	(0.955)	(1.226)	(1.238)	(-0.683)	(1.216)
Firm-CEO FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	13,092	13,092	14,328	11,859	12,073	12,073
Adjusted R-squared	0.767	0.777	0.757	0.768	0.763	0.773
Coefficient difference	ΔCoeff.	p-value	ΔCoeff.	p-value	ΔCoeff.	p-value
$Firm\ Ret_t$	-0.016	(0.630)	0.063	(0.104)	-0.010	(0.743)
$Peer\ Ret_t$	0.119	(0.003)	0.093	(0.027)	0.074	(0.067)

Table 9
Descriptive statistics: Forced CEO turnovers

This table reports descriptive statistics for all sample firms with available information for forced CEO turnover tests. Panel A presents descriptive statistics for main variables. $Forced_t$ is an indicator equal to one if the forced CEO turnover occurs in period *t*, zero otherwise. $Firm\ Ret_{t-1}$ is measured as the natural logarithm of one plus annual buy-and-hold stock return in period *t-1*. If a CEO turnover occurs in period *t*, annual returns are measured over a period that covers the 12 months prior to the CEO departure date. $Peer\ Ret_t$ is measured as the natural logarithm of one plus equal-weighted annual average returns based on the characteristics-matched TNIC peers over the same period that $Firm\ Ret_{t-1}$ is measured. $SIC\ Peer\ Ret_{t-1}$ and $GICS\ Peer\ Ret_{t-1}$ are measured using the same method as $Peer\ Ret_{t-1}$ except for the use of SIC and GICS industries, respectively. Panel B presents Pearson correlations among main variables. Correlations that are significant at 5% level are bolded. The sample period is between 1996 to 2015. All variables are defined in the Appendix A.

Panel A Descriptive Statistics

	N	Mean	Std	Q1	Median	Q3
$Forced_t$	26,187	0.02	0.14	0.00	0.00	0.00
$Firm\ Ret_{t-1}$	25,993	0.06	0.45	-0.14	0.10	0.30
$Peer\ Ret_{t-1}$	26,187	0.07	0.32	-0.08	0.11	0.26
$SIC\ Peer\ Ret_{t-1}$	25,565	0.06	0.33	-0.09	0.10	0.25
$GICS\ Peer\ Ret_{t-1}$	25,857	0.07	0.32	-0.08	0.11	0.26

Panel B Pearson Correlations

	Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	$Forced_t$	-									
(2)	$Firm\ Ret_{t-1}$	**-0.10**	-								
(3)	$Peer\ Ret_{t-1}$	**-0.02**	**0.57**	-							
(4)	$SIC\ Peer\ Ret_{t-1}$	**-0.02**	**0.54**	**0.74**	-						
(5)	$GICS\ Peer\ Ret_{t-1}$	**-0.02**	**0.57**	**0.77**	**0.78**	-					
(6)	$Size_{t-1}$	0.00	**0.02**	**0.04**	**0.02**	**0.03**	-				
(7)	BM_{t-1}	**0.03**	**-0.32**	**-0.15**	**-0.15**	**-0.15**	**0.02**	-			
(8)	Vol_{t-1}	**0.03**	**-0.11**	**-0.09**	**-0.08**	**-0.08**	**-0.38**	**0.06**	-		
(9)	$Tenure_t$	**-0.02**	**0.05**	0.01	0.01	0.01	**-0.09**	**-0.04**	**-0.03**	-	
(10)	Age_t	**-0.03**	**0.02**	**0.03**	**0.03**	**0.03**	**0.12**	**0.05**	**-0.13**	**0.39**	-
(11)	$Ownership_{t-1}$	**-0.03**	**0.02**	0.00	-0.01	0.01	**-0.14**	**-0.03**	**0.10**	**0.34**	**0.15**

Table 10
Tests of Relative Performance Evaluation in forced CEO turnover decisions

This table presents the estimation results from the following regression model.

$$Forced_t = \beta_1 \, Firm\ Ret_{t-1} + \beta_2 \, Peer\ Ret_{t-1} + \beta_3 \, Size_{t-1} + \beta_4 \, BM_{t-1} + \beta_5 \, Vol_{t-1} + \beta_6 \, Tenure_t + \beta_7 \, Age + \beta_8 \, Ownership_{t-1} + \varepsilon_t$$

In column 1, the estimation uses the full sample. In column 2, 3, and 4, the full sample is divided into three subsamples based on the tercile of the number of TNIC peers. Results testing strong-form evidence of RPE and the coefficient differences are summarized toward the bottom of the table. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent Variable: *Forced*$_t$			
Independent Variables	(1)	(2)	(3)	(4)
		# of TNIC Peers		
	Full Sample	Few	Moderate	Many
Firm Ret$_{t-1}$	-0.038***	-0.038***	-0.047***	-0.038***
	(-8.264)	(-5.215)	(-5.191)	(-4.301)
***Peer Ret*$_{t-1}$**	**0.022*** **	**0.007**	**0.028*** **	**0.043*** **
	(4.352)	**(0.894)**	**(2.809)**	**(3.951)**
Size$_{t-1}$	0.002	-0.004	0.003	0.004
	(0.517)	(-0.786)	(0.400)	(0.658)
BM$_{t-1}$	-0.006	-0.011	-0.022**	0.002
	(-1.088)	(-1.328)	(-2.468)	(0.215)
Vol$_{t-1}$	0.047	0.069	0.074	0.036
	(1.425)	(1.069)	(1.065)	(0.637)
Tenure$_t$	0.023***	0.020***	0.027***	0.032***
	(8.814)	(3.961)	(4.641)	(6.506)
Age$_t$	-0.002***	-0.001*	-0.002***	-0.002***
	(-5.531)	(-1.942)	(-2.704)	(-3.770)
Ownership$_{t-1}$	-0.036	-0.064	-0.114	-0.072
	(-1.349)	(-1.230)	(-1.268)	(-1.496)
Strong RPE F-Stat	**8.320**	**10.850**	**4.140**	**0.170**
p-value ($\beta_1 + \beta_2 = 0$)	**0.004**	**0.001**	**0.042**	**0.681**
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of observations	25,993	8,728	8,620	8,645
Adjusted R-squared	0.044	0.070	0.086	0.046
Coefficient Difference	$\Delta\beta_1$	p-value	$\Delta\beta_2$	p-value
Few versus Moderate	-0.009	(0.414)	0.021	(0.064)
Moderate versus Many	0.008	(0.473)	0.014	(0.290)
Few versus Many	-0.000	(0.978)	0.036	(0.004)

Table 11
RPE in forced CEO turnover decisions using alternative industry classifications

This table presents the estimation results from the following regression model.

$$Forced_t \;=\; \beta_1\,Firm\ Ret_{t-1} + \beta_2\,Alternative\ Peer\ Ret_{t-1} + \beta_3\,Size_{t-1} + \beta_4\,BM_{t-1} + \beta_5\,Vol_{t-1} + \beta_6\,Tenure_t + \beta_7\,Age + \beta_8\,Ownership_{t-1} + \varepsilon_t$$

In Panel A, the *Alternative Peer Ret*$_t$ is *SIC Peer Ret*$_{t-1}$ that uses SIC industries to define peers. In Panel B, the *Alternative Peer Ret*$_{t-1}$ is *GICS Peer Ret*$_{t-1}$ that uses GICS industries to define peers. In each panel, column 1 presents estimation resulting using the full sample, and in columns 2-4, the full sample is divided into three subsamples based on the tercile of the number of industry peers that is used in each panel. Results testing strong-form evidence of RPE and the coefficient differences are summarized toward the bottom of the table. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

Panel A RPE tests in forced CEO turnover decisions using SIC industries

	Dependent Variable: *Forced*$_t$			
Independent Variables	(1)	(2)	(3)	(4)
		# of SIC Peers		
	Full Sample	Few	Moderate	Many
Firm Ret$_{t-1}$	-0.038***	-0.045***	-0.032***	-0.038***
	(-8.449)	(-5.390)	(-4.027)	(-4.850)
***SIC Peer Ret*$_{t-1}$**	**0.018*** **	**0.010**	**0.014**	**0.032*** **
	(4.053)	**(1.427)**	**(1.365)**	**(3.789)**
Size$_{t-1}$	0.002	-0.003	-0.002	0.005
	(0.544)	(-0.412)	(-0.449)	(0.944)
BM$_{t-1}$	-0.008	-0.023***	-0.004	-0.000
	(-1.548)	(-2.644)	(-0.523)	(-0.023)
Vol$_{t-1}$	0.051	0.088	0.075	0.018
	(1.479)	(1.216)	(1.162)	(0.373)
Tenure$_t$	0.024***	0.028***	0.023***	0.025***
	(8.776)	(5.118)	(4.934)	(5.056)
Age$_t$	-0.002***	-0.003***	-0.001*	-0.002**
	(-5.731)	(-4.417)	(-1.656)	(-2.560)
Ownership$_{t-1}$	-0.034	-0.036	-0.052	-0.025
	(-1.220)	(-0.726)	(-1.355)	(-0.511)
Strong RPE F-Stat	**13.870**	**14.740**	**2.880**	**0.310**
p-value ($\beta_1 + \beta_2 = 0$)	**0.000**	**0.000**	**0.090**	**0.579**
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of observations	25,375	8,903	8,066	8,398
Adjusted R-squared	0.043	0.049	0.050	0.039
Coefficient Difference	$\Delta\beta_1$	p-value	$\Delta\beta_2$	p-value
Few versus Moderate	0.013	(0.224)	0.004	(0.710)
Moderate versus Many	-0.005	(0.593)	0.018	(0.131)
Few versus Many	0.008	(0.476)	0.023	(0.025)

Panel B RPE tests in forced CEO turnover decisions using SIC industries

Independent Variables	Dependent Variable: $Forced_t$ (1)	(2)	(3)	(4)
		# of GICS Peers		
	Full Sample	Few	Moderate	Many
$Firm\ Ret_{t-1}$	-0.038***	-0.033***	-0.052***	-0.031***
	(-8.360)	(-3.810)	(-5.726)	(-3.993)
$GICS\ Peer\ Ret_{t-1}$	**0.024*** **	**0.013***	**0.040*** **	**0.026*** **
	(4.890)	**(1.712)**	**(3.919)**	**(2.965)**
$Size_{t-1}$	0.001	-0.002	-0.001	0.005
	(0.495)	(-0.463)	(-0.086)	(1.074)
BM_{t-1}	-0.006	-0.003	-0.019**	0.003
	(-1.124)	(-0.295)	(-2.019)	(0.244)
Vol_{t-1}	0.046	0.040	0.075	0.036
	(1.387)	(0.616)	(0.984)	(0.775)
$Tenure_t$	0.023***	0.023***	0.030***	0.030***
	(8.801)	(4.551)	(5.117)	(5.880)
Age_t	-0.002***	-0.002***	-0.002***	-0.002***
	(-5.529)	(-2.707)	(-3.115)	(-3.964)
$Ownership_{t-1}$	-0.035	-0.065*	-0.003	-0.039
	(-1.328)	(-1.883)	(-0.048)	(-0.724)
Strong RPE F-Stat	**7.130**	**4.000**	**1.300**	**0.360**
p-value ($\beta_1 + \beta_2 = 0$)	**0.008**	**0.046**	**0.254**	**0.551**
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
# of observations	25,663	8,777	8,355	8,530
Adjusted R-squared	0.050	0.085	0.054	0.040
Coefficient Difference	$\Delta\beta_1$	p-value	$\Delta\beta_2$	p-value
Few versus Moderate	-0.020	(0.089)	0.027	(0.022)
Moderate versus Many	0.021	(0.058)	-0.014	(0.260)
Few versus Many	0.002	(0.888)	0.013	(0.230)

Table 12
Comparison with alternative industry classifications in forced CEO turnover decisions

This table presents the estimation results from the following regression model.

$$Forced_t = \beta_1 \, Firm\ Ret_{t-1} + \beta_2 \, Peer\ Ret_{t-1} + \beta_3 \, Alternative\ Peer\ Ret_{t-1} + \beta_4 \, Size_{t-1} + \beta_5 \, BM_{t-1} + \beta_6 \, Vol_{t-1} + \beta_7 \, Tenure_t + \beta_8 \, Age + \beta_9 \, Ownership_{t-1} + \varepsilon_t$$

In column 1, the *Alternative Peer Ret*$_{t-1}$ is *SIC Peer Ret*$_{t-1}$ that uses SIC industries to define peers. In column 2, the *Alternative Peer Ret*$_{t-1}$ is *GICS Peer Ret*$_{t-1}$ that uses GICS industries to define peers. In column 3, *SIC Peer Ret*$_{t-1}$ and *GICS Peer Ret*$_{t-1}$ are included in the same regression simultaneously. In columns 4-6, we use all firms in the same industry group to construct peer return variables rather than characteristics-matched peers. The sample period between 1996 and 2015. All variables are defined in the Appendix. Standard errors are clustered by firm. ***, **, and * represent significance level at the 1%, 5%, and 10% level, respectively. Robust t-statistics are in parentheses.

	Dependent variable: *Forced*$_t$						
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm Ret$_{t-1}$	-0.040***	-0.040***	-0.041***	-0.040***	-0.039***	-0.040***	-0.039***
	(-8.723)	(-8.523)	(-8.833)	(-8.762)	(-8.660)	(-8.758)	(-8.607)
***Peer Ret*$_{t-1}$**	**0.017*** **	**0.012** **	**0.011* **	**-**	**-**	**-**	**0.005**
	(2.842)	**(2.201)**	**(1.829)**	**-**	**-**	**-**	**(0.658)**
SIC Peer Ret$_{t-1}$	0.009*	-	0.004	-	-	-	-
	(1.871)	-	(0.676)	-	-	-	-
GICS Peer Ret$_{t-1}$	-	0.016***	0.015***	-	-	-	-
	-	(3.110)	(2.578)	-	-	-	-
***All Peer Ret*$_{t-1}$**	**-**	**-**	**-**	**0.029*** **	**0.024*** **	**0.026*** **	**0.024*** **
	-	**-**	**-**	**(4.132)**	**(3.658)**	**(3.734)**	**(2.780)**
All SIC Peer Ret$_{t-1}$	-	-	-	0.002	-	-0.001	-
	-	-	-	(0.291)	-	(-0.111)	-
All GICS Peer Ret$_{t-1}$	-	-	-	-	0.006	0.006	-
	-	-	-	-	(0.980)	(0.733)	-
Controls	Included	Included	Included	Included	Included	Included	Included
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	25,375	25,663	25,084	25,402	25,672	25,118	25,993
Adjusted R-squared	0.044	0.050	0.049	0.045	0.050	0.050	0.044