What Makes a Winner?
Toward Resolving the Role of Luck and Skill in Sustained CEO Performance

Seoyoung Kim†
Assistant Professor of Finance
Santa Clara University
Leavey School of Business

Robert Eberhart‡
Assistant Professor of Organization Analysis
Santa Clara University
Leavey School of Business

Daniel Armanios§
Assistant Professor
Department of Engineering and Public Policy
Carnegie Mellon University

May 2017

Keywords: luck; CEO evaluation; upper-echelon theory; sustained performance; performance evaluation.

* We thank Sanjiv Das, David Denis, Diane Denis, Mara Faccio, Byoung-Hyoun Hwang, Di Li, John McConnell, Max Moroz, Renee Rottner, Bill Starbuck, Jin Xu, Deniz Yavuz, Xiaoyan Zhang, and seminar participants at the Sustainability, Ethics, & Entrepreneurship (SEE) 2017 Conference for helpful comments.
† Leavey School of Business, Santa Clara University; 500 El Camino Real; Santa Clara, CA 95053. Phone: (408) 554-5171. Email: srkim@scu.edu.
‡ Leavey School of Business, Santa Clara University; 500 El Camino Real; Santa Clara, CA 95053. Phone: (408) 554-4574. Email: reberhart@scu.edu
§ Department of Engineering and Public Policy, Carnegie Mellon University; 5000 Forbes Avenue; Pittsburgh, PA 15213. Phone: (412) 268-2670. Email: darmanios@cmu.edu
What Makes a Winner?
Toward Resolving the Role of Luck and Skill in Sustained CEO Performance

Abstract

This study is an empirical examination of the extent to which luck, controlling for skill, explains sustained performance. Using a unique empirical strategy of bootstrap simulations that we compare to actual performances of public companies, we observe that over 95% of the differences in performance outcomes between “top” versus “average” performers can be attributed to luck, even if all CEOs are equally skilled. Through this novel empirical approach, we can better incorporate the role of luck into studies of sustained performance, and our findings suggest that more attention should be placed on the role of unanticipated and even uncontrollable changes on performance.

Keywords: luck; CEO evaluation; upper-echelon theory; sustained performance; performance evaluation.
INTRODUCTION

The role of luck has come to the forefront of scholarly discussions aimed at explaining the role of a CEO in organizational performance (Liu et al., 2016). Academic and policy discussions – that often justify CEO compensation - rely on the idea that CEO skills explain observed sustained exceptional performance differences between firms (Denrell, 2004a; Pfeffer et al., 2003). Challenging this idea, empirical work is increasingly indicating that, instead of CEO skill, simple luck can produce observed superior performance with sufficient frequency to convince observers that CEO skill is involved (Denrell, 2004b; Denrell et al., 2014; Henderson et al., 2012). What this work seems to indicate is that unanticipated conditions beyond a CEO’s control may actually allow certain CEOs to continuously outperform their peers, irrespective of their skill. Not surprisingly, this stream supports a provocative discussion within the literature surrounding executive performance and pay (Fitza, 2017; Quigley et al., 2017).

Disentangling “luck” from “skill” is a notoriously challenging empirical exercise, and ours is not the first study to embark on this path. Recent literature in strategic management highlights this difficulty. One common approach to separate CEO effects from luck is the use of variance partitioning methods. This method employs archival data to infer the proportion of firm performance that is attributable to the CEO after controlling for other contextual factors and levels of analysis, such as the type of industry (Hambrick et al., 2014). Variance partitioning methods have tended to find that CEO skill is a predominant explanatory factor. More recently, researchers have employed “bootstrapping” methods to compare actual CEO-firm performance to simulated or randomly generated baselines of performance (Denrell, 2004b; Fitza, 2014;
Henderson et al., 2012). These studies explore how random simulations can explain sustained exceptional performance, suggesting a more predominant role for luck over skill. In sum, separating luck from skill while evaluating CEOs remains a concern, while sustained exceptional performance is often taken as an axiomatic indication of CEO skill.

Given the difficulty of disentangling luck from skill, and the importance of these studies in policy discussions, the validity of estimates of the role of luck versus skill in CEO performance has come into question. Nonetheless, these studies continue to face the challenge that luck and skills are difficult to distinguish and are difficult to equitably incorporate into these techniques. Moreover, these studies still consign the role of luck to explaining performance variance. What has yet to be considered is that if luck can explain performance variance, then it may also help to explain differences in sustained exceptional performance. Thus, an important question to ask is: controlling for CEO skill, to what degree can luck explain differences in sustained firm performance across CEOs?

The purpose of this study is to contribute to this discussion by offering a novel empirical perspective. That is, we examine how much of a sustained exceptional performance run can be attributed to differences in luck as opposed to differences in skill. We do this via bootstrap simulations based on empirical data for each of the CEOs and their respective firms. This method provides us the baseline performance variation that we can expect when all CEOs are equally skilled. We compare this baseline against the actual performance variation we observe to draw inferences about the proportion of luck that is responsible for sustained performance. In other words, because we control for skill with this unique baseline grounded in observed performance, our method uncovers the true underlying differences in ability across CEOs, which allows us to

---

1 See Schwab and Starbuck (2012) and Schwab and Starbuck (2013) for more in depth discussion regarding baseline modeling.
more precisely infer the portion of sustained performance variation that is attributable to differences in luck rather than to differences in skill.

To show this, we form simulated cross-sections of turnaround performances under the assumption that the incoming and departing CEOs are equally skilled. Our results indicate that luck is often indistinguishable from true skill even in sustained long-run performance across all except the most successful CEOs. Our results also suggest that the differences in performances surrounding CEO turnovers are also largely attributable to differences in luck rather than to differences in skill between the incoming and outgoing CEOs.

Thus, while prior literature predominantly views differences in skill as explaining differences in long-run sustained performance (while consigning luck to noise that, on average, is zero), our study indicates this prediction is far from being settled. In sum we adopt the bootstrap simulation methods of recent literature, but offer methodological improvements to increase the validity of these comparisons and to more convincingly estimate the potential role of luck in sustained performance above and beyond that of skill.

Given that we are trying to better understand a core phenomenon in strategy research, i.e., that of sustained exceptional performance, and trying to adjudicate among multiple perspectives on this phenomena (mainly skill versus luck), our paper follows a “red” state approach (Mitchell et al., 2012). As such, this paper begins with a theoretical discussion of the skill and luck perspectives as a means to distinguish between these theoretical underpinnings as to what drives exceptional performance and to the relative contribution of each perspective. We then proceed to outline our methodology and results, and we conclude by discussing our contribution to the literature.
LITERATURE REVIEW

The role of luck in sustained performance variation

A key concern among researchers as well as in policy discussions of top management teams surrounds whether the variation in observed performance is largely attributable to luck rather than to a CEO’s skill. However, a key difficulty in this debate lies in how to adequately account for differences in skill across CEOs, and also, in how to account for the extent to which luck can explain sustained long-run performance (Denrell, 2004b; Henderson et al., 2012). This is consequential because CEO compensation and retention are heavily influenced by firm performance in general, and stock-return performance, in particular (e.g., Coughlan et al., 1985; Fee et al., 2004; Warner et al., 1988; Weisbach, 1988).

Scholars who emphasize internal firm capabilities rely on unique capabilities or skills as the key mechanism leading to sustained performance. For example, sustained superior performance is seen as evidence of the CEO’s capabilities in recognizing what markets to enter and when (Gompers et al., 2010). Similarly, acquisition returns are seen as evidence of an experienced CEOs ability to “capture a larger fraction of the merger surplus” by negotiating a lower premium for the target firm (Custódio et al., 2013) Evidence also suggests that a well-connected CEO can improve firm performance by exploiting his personal connections to improve financing costs and investment policy for the firm (Engelberg et al., 2012). Conversely, poor performance may be indicative of shirking CEOs who do not place sufficient efforts in their work for the firm, as operationalized by the personal use of corporate jets (Yermack, 2006). Evidence also suggests that replacing a CEO is costly, adding incentives for boards to view success as favoring a CEO’s skills (Taylor, 2010)
Even scholars who consider more random forces external to the firm also generally rely on the interaction of these effects with more internal skill-based processes to explain sustained performance. For example, in a turbulent business environment, dominant CEOs seem to perform poorly compared to diverse top management teams (Haleblian et al., 1993). Here again, internal capabilities are presumed as crucial to explain differences in performance outcomes. The implication is that if a firm or management team sustains superior performance, differences in skill-based internal processes must explain these sustained exceptional successes. Overall, the reliance on internal skills of the CEO and the top management team to explain performance variance can obscure the role of simple luck through attribution error.

*Disentangling luck from skill*

In response to concerns about the attribution of skill to luck, researchers are turning to luck-based simulations to reproduce performance variation and sustained performance in their efforts to explain the turnover among top performers (Liu et al., 2016). This discussion is of interest to academics and practitioners alike, because until only recently, the convention has been to ascribe internal idiosyncratic capability to exceptional performance, while ascribing these same performance levels to luck has been the exception. Accordingly, researchers have embarked on studies to quantify the relative weight of luck’s potential effect on firm performance.

One common method is to use archival data to assess firm performance over time. In this method, external effects on firm performance are controlled for and the remaining variance is analyzed via variance partitioning methodologies (Crossland et al., 2007; McGahan et al., 1997). These studies typically report the allocation of variance amongst annual-, industry-, firm-, and CEO-specific effects, with the unexplained remainder being attributed to estimation error and
luck. However, questions have arisen as to whether the performance variation attributed to CEOs is overstated because CEO effects are difficult to distinguish from random effects when each CEO is associated with a relatively small number of observations (Fitza, 2014; Quigley et al., 2017).

Responding to the limitations of variance partitioning methods while exploring the portion attributable to luck, researchers have turned to methods that compare the estimated CEO effects on the actual dispersion of firm performances with the estimated CEO effects on the dispersion generated by simulated performances. For instance, comparing a panel of randomly generated annual ROA (i.e., return on assets) to actual performance data, Fitza (2014) provides evidence suggesting that CEO effects may have been overstated by more than 13% in prior variance-decomposition studies.

Subsequent researchers have called this result into question (Quigley et al., 2017). One concern is that the simulated ROA data in Fitza (2014) is randomly generated without regard to the underlying economic structure and process from which this metric is born. That is, the data is generated in a random fashion without regard to the time-varying industry/market effects, the firm-specific persistence factor, and the accrual-accounting principles that drive this performance metric in reality. Second, the annual granularity of the approach employed by Fitza (2014) limits the interpretation of his results, since there are far more CEO dummies for the panel dataset than there are industry or firm dummies. In response, Quigley and Graffin (2017) “also removed CEOs who served just one year as their effects are perfectly predicted by their CEO dummy, and as a result, artificially inflate the CEO effect”. Quigley and Graffin also employed a multilevel

---

2 We note that even this improvement may artificially inflate the variation attributed to the CEO dummies, since a tenure of just two years still only spans two observations when working with annual performance data.
modeling (MLM) approach to account for the inherently nested structure of firm-performance data; namely, that the CEO is part of a firm, which in turn, is part of an industry.

While insightful and adding important nuances, Quigley and Graffin (2017) continue to generate annual ROA in a completely random fashion, thereby calling into question whether CEO effects are understated when applying the MLM approach to simulated data that is not predicated on the economic forces that drive performance, particularly for metrics such as ROA which tend to be highly persistent. Consistent with this notion, Fitza (2017) finds again that CEO effects may be greatly overstated when he expands upon his earlier 2014 study, this time applying the MLM approach and also allowing for industry-level persistence in the simulated annual ROA.

Overall, a natural question remains as to whether randomly generated performance variation would continue to be (falsely) attributed to CEO effects if performance data were observed at greater frequencies or if simulated data were predicated on finer firm-specific parameters. Furthermore, in evaluating the CEO-specific variation in a panel of annual performance metrics, the studies to date do not address the question as to whether differences in sustained performance are predominantly attributable to differences in CEO skill or simply to sampling variation (i.e., luck). We now proceed to outline the intuition behind our approach to disentangle the components of this important but hairy question.

**Simulations Predicated on Actual Data**

Our proposed approach is to compare the differences in long-run performance outcomes (which we adjust for industry and market benchmarks) between the best and worst performers that we observe in the actual data, with the differences expected in a simulated economy (via a unique
bootstrapping simulation approach) where all CEOs are equally skilled. In this way, we explicitly control for CEO skill with a novel mixed approach that uses observed empirical data to parameterize a simulation analysis to evaluate long-run performance outcomes that can repeat over time.

Our approach offers four distinct advantages. First, we complement and expand on earlier studies by using a time-series of monthly stock returns for publicly traded U.S. firms, allowing a higher frequency of observations that are not plagued by issues inherent in accounting metrics such as ROA (as we discuss later when we detail our empirical strategy). Second, by imposing the null hypothesis that all CEOs are equally capable we can attribute any performance differences observed in a given simulation to differences in luck rather than to differences in CEO skill. Third, we ground our simulated performance data based on actual stock-return data informing the industry and market parameters specific to each CEO/firm. Fourth, we measure the long-run benchmark-adjusted performance of each CEO based on his entire tenure at the firm in question, thereby allowing us to assess the extent to which variation in sustained superior performance may be attributed to differences in luck as opposed to differences in CEO skill.

Overall, prior research has closely examined external and internal forces that influence firm outcomes. However, questions still loom as to whether stable CEO-specific effects prevail, rather than the random external forces that, at least, have some degree of “luck”. This distinction is important because, if luck plays a conceivably large role, it can confound the earlier decomposition attempts of external versus internal effects on CEO performance, which, in turn, has an impact on the compensation that CEOs demand and the resources to which a firm might be able to gain access. This paper aims to empirically assess the role attributable to luck by better controlling for skill and time.
EMPIRICAL STRATEGY

To properly evaluate and interpret differences in firm performance outcomes, we must account for the fact that any large cross-section of firms can guarantee that some CEOs/firms will be repeatedly (and significantly) lucky or unlucky, thereby giving rise to substantial differences in sustained, long-run performance outcomes, even if all CEOs are equally skilled. To accomplish this objective, we design tests that control for the inevitable luck-based variation that arises in large samples. This methodological approach, also referred to as “baseline modeling”, has recently begun to gain traction across an array of disciplines (Schwab et al., 2012, 2013)

In our case, using parameters from the actual historical performance data, we repeatedly simulate an economy in which we impose the null hypothesis that all CEOs/firms are equally skilled. As such, any variations in performance that we observe in a given simulation are entirely due to differences in luck rather than to differences in skill or effort. Taking the average luck-based variation across many simulated economies thus provides a baseline against which we can compare the actual variation in performance outcomes that we observe in reality. This allows us to determine the extent to which historical performance data may be indicative of luck as opposed to skill. The formal model and methodology for doing this are described in detail further below in the subsections titled “Performance-evaluation methodology using actual historical performance” and “Bootstrap simulations imposing the null of equal skill across CEO-firm pairs”. We now proceed to outline our data sources, to define our key performance metrics and benchmarks, and to describe our bootstrap procedure in greater detail.
Data

We obtain our data from the Standard & Poor’s Executive Compensation database (hereafter, Execucomp), which consists of the S&P 1500 CEO-firm pairs and spans the period encompassing 1992 through 2015. For instance, Clifton A. Pemble / Garmin Ltd. constitutes a CEO-firm pair for the period spanning January 2013 to present. Stephen Riggio / Barnes & Noble Inc. constitutes another CEO-firm pair for the period spanning February 2002 to March 2010. This classification allows us not only to evaluate performances across firms, but also to evaluate changes in performances across CEOs within the same firm.

Although the Execucomp database begins in 1992, we consider a CEO’s entire tenure when evaluating his performance, even if his term begins before 1992. For example, if a CEO of firm X began his term in March of 1985, we use the returns from April of 1985 to the penultimate month of the CEO’s term to calculate his alpha. We obtain monthly returns data from the Center for Research in Security Prices (hereafter, CRSP) database, and we require that each CEO-firm pair have at least 24 months of returns data. As recommended by other studies, we exclude those CEO-firm pairs whose stock prices dip below $5.00 during that CEO’s term to avoid microstructure issues and estimation problems that accompany small, illiquid stocks (Chan et al., 2001; Jegadeesh et al., 2001). Our final sample consists of 22,617 firm-years, with 1,918 distinct firms and 4,080 distinct CEO-firm pairs. Of these, we have 2,717 turnover events for which we have at least 24 months of returns data preceding and following the new CEO’s arrival.

---

3 The Execucomp database does not provide data for the entire S&P 1500 prior to 1994.
4 To demonstrate and to ensure that our findings are not specific to a particular sample period, we also provide supplementary sub-period analyses in a separate online appendix to complement our main analyses.
Variables

Our key dependent variable is *Excess Returns* (denoted $r_{i,t}$), which we measure as the monthly stock return in excess of the risk-free rate of return for each CEO-firm pair $i$ at each point in time $t$. We use stock returns rather than accounting profitability metrics, because stock returns provide more frequent data points, which is necessary for reliable bootstrapping. That is, the use of even quarterly accounting figures would require a minimum CEO tenure of six years for sample selection, thereby restricting our analysis to only the most successful or entrenched CEOs.

Moreover, insofar as the market can only estimate the value added or destroyed by an incoming CEO/management team, stock returns provide timely, ongoing assessments of the team’s accomplishments or failures. Stock return performance is a crucial factor in how CEOs are evaluated, and strong returns are associated with higher compensation and subsequent CEO appointments at other firms as well as with retaining board appointments after retirement (Brickley *et al*., 1999; Fee *et al*., 2004; Murphy, 1999). Likewise, poor stock-return performance is associated with a greater likelihood of CEO turnover (Jenter *et al*., 2006; Kaplan *et al*., 1994; Warner *et al*., 1988; Weisbach, 1988). Furthermore, because we expect luck to have less of a role in large, public firms, ours is a more conservative test. In addition, stock returns do not require adjustments for seasonality inherent in accounting profitability as exemplified by the retail industry (i.e., consider the spike in earnings and sales for a retailer during the reporting period encompassing Black Friday and Christmas).

---

5 We also note that to properly account for the seasonality in quarterly accounting metrics would inevitably result in fewer degrees of freedom, and would require an even greater minimum tenure of at least 32 quarters (i.e., at least eight years) for sample CEOs. In addition, we would need to account for the natural differences in discretionary accruals and accounting choices, which are based on a multitude of characteristics beyond simply considering industry standards (e.g., firm size, volatility, leverage, profitability, capital expenditures, research-and-development expenditures, and governance structure, to name a few). These controls would result in yet another increase in the minimum tenure required for a proper bootstrap analysis (i.e., in the order of a 10- to 11-year minimum required for inclusion in our analysis).
We have two key independent variables, which we obtain from the CRSP database. First, to account for the market-wide economic conditions in which a firm is embedded, we control for \( MKTRF_t \), which represents the monthly excess market-portfolio return (i.e., the average excess stock return across all publicly traded firms) at time \( t \). Second, because prior research has found that the economic cycles and conditions across industries account for a substantial portion of observed performance variation (McGahan, 1999), we also control for \( INDRF_{i,t} \), which represents the monthly excess industry-portfolio return for firm \( i \)'s industry (i.e., the average excess stock return across all publicly traded firms in the same industry as firm \( i \)) at time \( t \).

**Performance-evaluation methodology using actual historical performance**

We begin by estimating the actual benchmark-adjusted performance (i.e., alpha) of each CEO-firm pair with an OLS model that allows us to account for general market- and industry-wide price movements. That is, we estimate the following time-series regression equation for each CEO-firm pair \( i \),

\[
    r_{i,t} = \alpha_i + \beta_{i,MKT} \cdot MKTRF_t + \beta_{i,IND} \cdot INDRF_{i,t} + \varepsilon_{i,t},
\]

where \( r_{i,t} \) is the monthly excess stock return for CEO-firm pair \( i \) at time \( t \), \( MKTRF_t \) is the monthly excess market return at time \( t \), and \( INDRF_{i,t} \) is the monthly excess industry portfolio return for firm \( i \)'s industry at time \( t \) based on the Fama-French (1997) 30-industry classification.\(^6\) In using the benchmark-adjusted performance, i.e., alpha, across the CEO’s entire tenure, we note that we are in fact comparing differences in **sustained** performance across CEO-firm pairs. That is, to achieve a high alpha, a CEO must persistently outperform benchmarks, as a few ad-hoc months of stellar performance will be washed out if most months during the CEO’s tenure are lackluster.

**Bootstrap simulations imposing the null of equal skill across CEO-firm pairs**

\(^6\) Obtained from Ken French’s website: [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We observe very similar results using blunter or finer classifications (e.g., such as the 49-industry classification).
Next, we implement our non-parametric bootstrap procedure as follows. This non-parametric bootstrap procedure that we employ is also commonly referred to as a non-parametric Monte-Carlo simulation, semi-parametric bootstrapping, or simply as a “re-sampling procedure”, and is a widely accepted and commonly used approach in obtaining numerical approximations via simulation (e.g., see Davidson and MacKinnon, 2003). 7

From regression equation (1), we obtain OLS parameter estimates and a time series of residuals, \( \hat{\epsilon}_{i,t}, \ldots, \hat{\epsilon}_{i,T_i} \). We then sample with replacement from this residual vector, assigning equal probability to each \( \hat{\epsilon}_{i,t} \), and we construct a simulated time-series of residuals, \( \varepsilon_{i,t}^*, \ldots, \varepsilon_{i,T_i}^* \). Although the sample mean of the residuals from equation (1) is zero by construction, the sample mean of the bootstrap error terms need not be, since some of the \( \hat{\epsilon}_{i,t} \)'s may be selected multiple times and others may not be selected at all. 8 Our bootstrap data-generating process is then:

\[
r_{i,t}^* = \alpha_{\text{NULL}} + \hat{\beta}_{i,\text{MKT}} \cdot \text{MKTR}_i + \hat{\beta}_{i,\text{IND}} \cdot \text{INDR}_i + \varepsilon_{i,t}^*, \quad \varepsilon_{i,t}^* \sim \text{EDF}(\hat{\epsilon}_i)
\]  

(2)

in which \( \hat{\beta}_{i,\text{MKT}} \) and \( \hat{\beta}_{i,\text{IND}} \) are the OLS parameter estimates obtained from regression equation (1), and \( \text{EDF}(\hat{\epsilon}_i) \) is the empirical distribution function that assigns equal probability, \( T_i^{-1} \), to each \( \hat{\epsilon}_{i,t} \) in the residual vector \( \hat{\epsilon}_i \). 9 Using this nonparametric bootstrap procedure, we simulate a time-series of returns for CEO \( i \) on whom we impose the constraint that her true skill is equal to some pre-specified magnitude (i.e., true \( \alpha_i = \alpha_{\text{NULL}} \)). Notably, in our method a bootstrap sample

---

7 Other studies using this bootstrap procedure include Kosowski, Timmerman, Werners, and White (2006) and Kosowski, Naik, and Teo (2007), who employ this approach to analyze mutual-fund and hedge-fund performances, respectively.

8 For example, if the residual vector from equation (1) is \([-1, 0, 1]\), then different iterations could yield bootstrap error-term vectors of \([-1, 0, 1],[-1, -1, -1],[1, 0, 0]\), and so on.

9 In our implementation, we scale the residual vector by a factor of \( [T_i/(T_i-K)]^{1/2} \), because the empirical distribution of the residuals from equation (1) has variance \( T_i^{1/2} \Sigma \hat{\epsilon}_{i,t}^2 = T_i^{1/2} (T_i-K) \hat{s}^2 \), in which \( K \) equals the number of regressors and \( \hat{s}^2 \) is the unbiased estimator of \( \sigma_e^2 \).
may yield an estimated alpha that is substantially different from $\alpha_{null}$ since, by chance, it may have drawn more of the positive (or more of the negative) $\hat{\epsilon}_{i,t}$'s.

Finally, to evaluate CEO performances accounting for the large cross-section to which they belong, we employ a cross-sectional bootstrap procedure. Specifically, we follow the above process for all CEO-firm pairs, $i = 1, ..., N$, and we repeat this process $B=999$ times to form $B$ cross-sections of $N$ bootstrapped alphas and $t$-statistics.\(^{10}\) Because the bootstrap data-generating process in equation (2) imposes the constraint that all CEO-firm pairs are equally skilled (i.e., true $\alpha_i = \alpha_{null}$ for all $i$), any cross-sectional variation in simulated sample alphas is entirely due to differences in luck as opposed to differences in skill, providing a benchmark of the extent to which CEOs are expected to achieve highly positive or negative benchmark-adjusted performances purely by chance. Thus, the nonparametric, cross-sectional bootstrap procedure that we employ allows us to evaluate CEO-firm performances within context of the entire cross-section to which they belong.

**RESULTS**

*Basic summary statistics*

In Table 1, we present summary statistics on basic CEO and firm characteristics. The average CEO in our sample is 55.8 years of age. 29% of CEOs are 60 years of age or older, with an average tenure of 7.1 years. The average firm in our sample has $17.1$ billion in total assets, with an average market capitalization of $15.3$ billion. Overall, the sample we study consists of large,

\(^{10}\) $B$ is chosen such that $\lambda \cdot (B+1)$, in which $\lambda$ is the size of the test, is an integer. Otherwise, the probability of Type I error cannot be exactly $\lambda$. Much like scaling by $N$ versus $(N-1)$ for a sample-variance calculation, this distinction becomes increasingly irrelevant for large $B$. 
publicly traded firms comprising the S&P 1500 index over the period spanning 1992 through 2015, a setting in which we anticipate that noise and luck should play a less prominent role.

\[\text{--- Insert Table 1 here ---}\]

Differences in CEO performances

We begin by examining the extent to which the actual best and worst benchmark-adjusted, long-run CEO performance outcomes differ from the average, and we compare this difference to the performance differential that we can expect in a simulated economy where all CEOs are equally skilled. That is, in each of the \(B=999\) simulated cross-sections of individual CEO performances, we see the differences in performance outcomes between the top and average performers, the average of which provides the expected performance differential due solely to differences in luck as opposed to differences in skill. We then calculate bootstrapped \(p\)-values as the proportion of bootstrap iterations yielding a performance differential that is at least as extreme as the actual difference observed in the actual data.

The results, presented in Table 2 (and depicted graphically in Appendix A, of our Online Appendix), show that, empirically, the difference in benchmark-adjusted performances between the 90\(^{th}\) percentile and median CEOs was 1.39\% per month, or 16.68\% annually. In comparison, results from bootstrap simulations indicate that, purely by sampling variation, we would expect the 90\(^{th}\) percentile CEO to outperform the median by 1.34\% per month. This suggests that an annualized performance differential of 16.08\% is easily attributable to greater luck, leaving a more modest 0.60\% that is not attributable to luck. Based on the average sample market capitalization of $15 billion, this 0.60\% skill-related difference translates to an average dollar
amount of $90 million that is due to greater skill. While $90 million is not trivial, it pales in comparison to the potential economic gain of $2.4 billion that is attributable to greater luck as opposed to greater skill.

Moreover, the bootstrapped p-value of 0.09 indicates that the simulated performance differential between the 90th percentile and median CEOs exceeds the actual observed difference of 1.39% in 9% of all bootstrapped cross-sections. Thus, while the performance differential observed in actuality often exceeds the simulated differences in performance (under the assumption that all CEOs are equally skilled), our estimate of the effect of luck provides a more conservative and holistic decomposition of sustained, superior performance. Overall, our results suggest that the performance differential due to differences in skill is not as large as implied by the raw difference in outperformance between the top and median outcomes.

The difference between the top 10% of performers and the mean cross-sectional performance tells a similar story: in actuality, the top 10% of performance outcomes exceeded the mean by 2.49% per month; by sampling variation alone, this difference in benchmark-adjusted performances is expected to be 2.36% per month, suggesting an annualized performance differential of 1.56% (or, 0.13% per month) that is attributable to greater skill (bootstrapped p-value = 0.01). Similarly, the 99th percentile performers exhibit not only greater luck but also greater skill relative to the 90th percentile performers, with an annualized performance

--- Insert Table 2 here ---

11 In contrast, we find that in simply evaluating the variation in performance across firms, the simulated performance differential between the firm at the 90th percentile and the median firm exceeded the actual observed difference in 50% of all bootstrapped cross-sections (results available in separate online appendix). Thus, although a large portion of the differences in performances across CEOs is still attributable to luck, this more nuanced examination of performance variation is less driven by differences in luck than the performance variation across firms in general.
differential of 4.68% (or, 0.39% per month) that is attributable to greater skill (bootstrapped \(p\)-value = 0.02).

On the other hand, although the actual best performer outperformed the 99\(^{th}\) percentile CEO by 6.60% per month, the bootstrap simulations indicate that even if all CEOs are equally skilled, the best performer is expected to outperform the 99\(^{th}\) percentile CEO by 6.90% per month (and in 46% of simulated cross-sections, the performance differential was at least as large as the 6.60% observed in actuality).

Similar observations apply when we extend our analysis to poor performance outcomes (Panel B). We observe in our sample companies that the benchmark-adjusted performance differential between the median and 10\(^{th}\) percentile CEOs was 1.16% per month. In comparison, our simulation results indicate that, by poor luck alone, we would expect the 10\(^{th}\) percentile CEO to underperform the median by 1.33% per month, and in each of the 999 bootstrapped cross-sections, the performance differential between the 10\(^{th}\) percentile and median CEOs was at least as extreme as the observed difference empirically (bootstrapped \(p\)-value = 1.00).

The results suggest that the best CEOs in actuality perform too well (relative to the average performer) to be completely explained by greater luck. However, a substantial portion of the observed differences in performance outcomes are attributable to differences in luck, indicating that the true underlying differences in skill are substantially smaller than suggested by simply looking at the raw difference in benchmark-adjusted performance outcomes.\(^{12}\) Overall, the results highlight the importance of gauging the extent to which CEOs are expected to outperform benchmarks and peers by sampling variation alone; this particularly applies to large

\(^{12}\) We draw similar conclusions in supplementary sub-period analyses, which we provide separately in an online appendix.
samples, which guarantee the occurrence of seemingly exceptional performance outcomes even if all CEOs are truly, equally skilled.

*Best / worst CEO performances, and the frequency of extreme outcomes*

To further illustrate the importance of accounting for large cross-sections when evaluating the role of luck in CEO performances, we consider the highest performing CEO in our sample of S&P 1500 firms, who enjoyed a market- and industry-adjusted performance of 11.19% per month, with a standard error of 4.16%, over his five-year tenure prior to the recent sub-prime financial crisis. Evaluated in isolation, the probability of a particular CEO performing so well by luck alone is very low, since a $t$-statistic of 2.69 indicates a 0.36% chance that a true zero-alpha performer would generate the observed outcome (or better). However, in a large cross-section of equally (un)skilled CEOs, it would be very surprising if *no one* produced such an extreme outcome. In fact, across 999 bootstrapped cross-sections of CEO performances, whereby we impose the assumption that all CEOs have zero skill, 34% produced a maximum performer whose monthly benchmark-adjusted performance was at least as high. That is, in 34% of simulations, the best performer, whose ‘achievement’ is solely an artifact of sampling variation, performed at least as well as the best performer observed empirically (un-tabulated).

------------- Insert Table 3 here -------------

To expound upon this example, we examine the actual frequency of CEOs who achieve extreme performance outcomes, and we compare this to the number of simulated CEOs who are expected to experience such extreme outcomes purely by luck. The results, presented in Table 3, show that, in actuality, 126 CEO-firm pairs (out of our sample of 4,080 CEO-firm pairs) had
benchmark-adjusted performances of at least 3% per month (Panel A). In comparison, bootstrap simulations indicate that under the assumption that all CEOs have zero skill (i.e., true $\alpha=0$), 84 CEOs are still expected to enjoy such positive outcomes by sampling variation alone. Thus, although a large portion of differences in performance are attributable to differences in luck, too many CEOs perform too well to be entirely attributed to luck alone.

With regard to poor performance outcomes (Panel B), we observe in our sample of firms, 38 CEO-firm pairs had benchmark-adjusted performances of less than -3% per month. Simulations indicate that, by poor luck alone, 69 zero-skill CEOs are expected to experience such negative performance outcomes. Similar observations apply when we extend our analysis to the frequency of extreme $t$-statistics.

Overall, the results show that in a large sample, many unskilled performers are sure to be sufficiently and repeatedly lucky (or unlucky) so as to produce outcomes that are statistically significant at conventional cutoffs.\textsuperscript{13} That is, a substantial number of CEOs are expected to achieve extreme performance outcomes based on extreme luck (as opposed to extreme skill), pointing to the importance of evaluating individual CEO performances within context of the large cross-section to which they belong. Consistent with the previous analyses, the simulations suggest that luck generates a substantial portion of benchmark-adjusted performance outcomes, and further highlight the importance of gauging not only the extent to which CEOs outperform industry benchmarks and peers, but also the extent to which they are expected to do so by sampling variation alone.

\textsuperscript{13} We draw similar conclusions in supplementary sub-period analyses, which we provide in separately in an online appendix.
Changes in firm performance surrounding CEO turnovers

In our last set of analyses, we assess changes in performance surrounding CEO turnovers. Changes in performance are often used to gauge the success of managerial replacement decisions, and provide an interesting additional setting in which to assess the roles of luck versus skill in observed performance outcomes. Specifically, we begin by examining benchmark-adjusted performances over a three-year window for the departing and incoming CEOs. The results, presented in Table 4, indicate that negative-alpha departing CEOs tend to be followed by better performing replacements, with an average difference in monthly alpha of 1.43% (Panel A). On the other hand, positive-alpha departing CEOs tend to be followed by worse performers, with an average difference in monthly alpha of -0.94% (Panel B).

However, when evaluating CEOs by the extent to which they improve or impair performance, the question remains as to the performance changes we can expect to observe even if the departing CEO is no better or worse than his replacement. That is, suppose a poorly performing CEO is replaced with a CEO who performs substantially better. How much of this improvement in performance can be attributed to differences in skill as opposed to differences in luck between the departing and incoming CEOs?

To explore this question, we use a similar procedure as before to form simulated cross-sections of turnaround performances under the assumption that the incoming and departing CEOs are equally skilled. Specifically, for each CEO-turnover event, we use the monthly time-series of (excess) stock returns from the 36 months prior to through the 36 months following the turnov
new CEO’s arrival (we skip the month of the new CEO’s arrival / prior CEO’s departure), and we estimate the following regression equation:

\[
\begin{bmatrix}
    r_{\text{prior,1}} \\
    \vdots \\
    r_{\text{prior,T} \text{old}} \\
    r_{\text{new,1}} \\
    \vdots \\
    r_{\text{new,T}_{\text{new}}}
\end{bmatrix}
= \begin{bmatrix}
    0 \\
    \vdots \\
    0
\end{bmatrix}
+ \beta_{\text{MKTRF}} \cdot MKTRF + \beta_{\text{IND}} \cdot INDRF + \epsilon \quad \text{for } i = 1, 2, \ldots, T_{\text{new}} \tag{3}
\]

such that \( i \) is a vector of ones. Using the OLS parameter estimates, \( \{\hat{\beta}_{\text{MKTRF}}, \hat{\beta}_{\text{IND}}\} \), and residual vector, \( \hat{\epsilon} \), from equation (3), our new bootstrap DGP is then:

\[
\begin{bmatrix}
    r_{*_{\text{prior,1}}} \\
    \vdots \\
    r_{*_{\text{prior,T} \text{old}}} \\
    r_{*_{\text{new,1}}} \\
    \vdots \\
    r_{*_{\text{new,T}_{\text{new}}}}
\end{bmatrix}
= \beta_{\text{MKTRF}} \cdot MKTRF + \hat{\beta}_{\text{IND}} \cdot INDRF + \epsilon^* , \quad \epsilon^* \sim \text{EDF}(\hat{\epsilon}) \quad \text{for } i = 1, 2, \ldots, T_{\text{new}} \tag{4}
\]

That is, we simulate a time-series of returns on which we impose the constraint that there is no difference in true skill between the departing and incoming CEOs (i.e., true \( \Delta \alpha = \alpha_{\text{new}} - \alpha_{\text{prior}} = 0 \)). Nonetheless, a bootstrap sample may yield an estimated change that is substantially different from zero when regressed on the augmented regression equation (3) since, by chance, it may have drawn more of the positive (or more of the negative) \( \hat{\epsilon}_{\ell,t} \)'s for the new CEO’s term than for the previous CEO’s term.

Because the bootstrap data-generating process in equation (4) imposes the assumption that there is no difference in true skill between a predecessor and his replacement, any perceived improvements or declines in performance come solely from differences in luck as opposed to

---

14 Our results are robust to allowing for differences in market and industry loadings between the new and old CEOs’ terms.

15 The bootstrapped changes in excess performance (\( \hat{\Delta}^*_\alpha \)) remain unaltered for any imposed baseline performance \( \alpha_i \in \mathcal{R} \).
differences in skill. In contrast to the previous simulations, this analysis allows for differing levels of skill across CEO-firm pairs; the only assumption is that a departing CEO and his direct replacement are equally skilled.

The results, presented in Table 5, show that in actuality, 194 CEOs outperformed their negative-alpha predecessors by at least 3% per month (Panel A), and 175 underperformed their positive-alpha predecessors by at least 3% per month (Panel B). In comparison, results from bootstrap simulations indicate that, by luck alone, 289 incoming CEOs are expected to enjoy such positive improvements in monthly performance, and 303 are expected to suffer such declines, even though all incoming CEOs have the same ability as their predecessors (i.e., true Δα=0).16 Similar observations apply when we extend our analysis to the frequency of extreme t-statistics (of the difference in alphas).

Overall, the simulations demonstrate that a substantial number of CEO turnover events are expected to result in significant benchmark-adjusted performance changes, even in an economy where departing CEOs are always followed by equally capable successors, and further highlight the importance of adequately accounting for luck when interpreting performance outcomes. To be clear, this is not to say that the CEO replacement decision is an inconsequential one; rather, these results highlight the fact that we can expect to see significant performance changes.

--- Insert Table 5 here ---

16 Appendix B of our Online Appendix provides a visual representation of these results, and Appendix C provides a graphical representation of the overall distribution of actual versus bootstrapped changes in performance surrounding CEO turnovers.
improvements or impairments even in cases where an incoming CEO has been carefully selected to be of equal quality as his predecessor.

**Boundary conditions**

For the avoidance of doubt, skill is not a uniquely defined concept, and in removing the market and industry component of returns, we may remove other aspects of CEO skill. For instance, part of a CEO’s job is to determine the firm’s optimal exposure to external factors (Gopalan *et al.*, 2010). In addition, CEOs can collectively influence the aggregate performance measures used as benchmarks (Aggarwal *et al.*, 1998). Nonetheless, our focus in this paper is to examine how much of the observed differences in actual performance outcomes can be attributed to differences in luck, while remaining cognizant of the fact that being attributable to luck does not necessarily preclude the role of skill.

Furthermore, we clarify that our analysis tracks only CEO-firm pairs, and does not track other changes within the “C”-suite or top management team, wherein turnover may occur at greater frequency. We also clarify that evidence of luck does not negate evidence of skill. That is, our results demonstrate that a wide dispersion in long-run performance outcomes is actually consistent with a world where all CEOs are equally skilled, which is not necessarily equivalent to suggesting that there are no cross-sectional differences in skill. Overall, our results simply suggest that the true underlying differences in skill are substantially smaller than suggested by looking at the raw performance differential.

**Robustness Analyses**

To explore the distinction between firm-level versus CEO/firm-level analyses, we replicate our simulation analyses from Tables 2 and 3 at the more granular firm level (as opposed
to the CEO/firm-level). The results, which we present in Appendices D, E, and F of our Online Appendix, show demonstrably different results than when we focus on CEO-firm pairs. Specifically, we find that although a large portion of differences in performances across firms is also attributable to luck, the more nuanced examination of performances across CEO/firm pairs is less driven by differences in luck than the performance variation across firms in general.

In addition, to demonstrate and ensure that our findings are not specific to a particular sample period, we also provide supplementary analyses to complement our main analyses in Tables 2 and 3, whereby we bifurcate our sample along the pre- versus post-2008 (i.e., pre-crisis versus post-crisis) sub-periods. The results can be found in Appendix G and Appendix H of our Online Appendix.

**DISCUSSION**

The objective of this paper is to better understand the contribution of luck, controlling for skill, in explaining sustained CEO/firm performance. Both skill-based (often attributed to aspects internal to the firm) and luck-based (often attributed to aspects external to the firm) explanations are used in interpreting differences across performance outcomes. However, prior literature has made little headway with respect to isolating the contribution of luck from skill, particularly in evaluating sustained long-run performance in a large cross-sections of CEO-firm pairs. We see this as particularly important given firm performance is critical to top management evaluation and compensation, as well as to the assessment of firm value. Such assessments perceive performance outcomes as predominantly a consequence of skill and that luck is a minor factor in such outcomes.
Our results employ a method that uniquely controls for skill. We show that there are still substantial differences in realized performance outcomes even if all CEOs in a given sample are equally skilled. This finding suggests that the true underlying differences in CEO skill are substantially smaller and attributable to luck than implied by simply looking at the raw difference in benchmark-adjusted performance outcomes. That is, empirically (i.e., based on the actual observed data), the difference in benchmark-adjusted performances between the 90th percentile and median CEOs was 1.39% per month, or 16.68% annually. In comparison, results from bootstrap simulations indicate that by sampling variation alone, we would, on average, expect to observe a performance differential of 1.34% per month, suggesting an annualized performance differential of 16.08% that is entirely attributable to luck; i.e., luck alone is able to explain over 95% of the observed differences in actual sustained performances. In other words, although the best-performing CEOs perform too well relative to the median to be completely explained by luck alone, the simulations operating under the (extreme) assumption that all CEOs are equally skilled approximate the observed actual data quite well.

We confirm the robustness of this finding when we extend our analysis to poor performance outcomes. We observe that the benchmark-adjusted performance differential between the median and 10th percentile CEOs was 1.16% per month. Simulation results indicate that, by luck alone, we would expect the 10th percentile CEO to underperform the median by 1.33% per month; in fact, in all 999 simulated cross-sections, the difference between the 10th percentile and median performers was at least as extreme as the actual, observed difference.

Furthermore, by examining changes in performance surrounding CEO turnovers, we observe that, 194 incoming CEOs outperformed their negative-alpha predecessors by at least 3%

---

17 Specifically, these benchmark-adjusted performances are the alphas obtained from regressing excess monthly returns on excess market and excess industry returns.
per month. In comparison, simulation results indicate that, by luck alone, we would expect 289 incoming CEOs to enjoy such positive improvements in monthly performance. In a similar vein, we observe that 175 incoming CEOs underperformed their positive-alpha predecessors by at least 3% per month, yet simulation results indicate that 303 are expected to suffer such declines by poor luck alone.

Together, these findings demonstrate that the wide dispersion in observed poor performance outcomes is not statistically distinguishable from a world where all CEOs are equally skilled. Moreover, these results highlight the importance of accounting for the role of luck borne by large cross-sections when interpreting performance outcomes, particularly in the context of CEO replacement decisions, which is arguably one of the most important (and potentially costliest) decisions made by firms. That said, although luck can explain a significant part of the performance differential, it does not completely explain sustained exceptional performance. That is, the number of sustained exceptional performances is too large to be explained or replicated by luck alone. As such the prior literature on CEO skill becomes even more germane as new research begins to better unpack performance heterogeneity. We now turn to specific implications and contributions.

*Implications for research on the drivers of firm performance*

First, we add to prior research on sustained performance differences across firms. Our core contribution is that we develop a method to isolate and quantify the influence of luck from skill as a source of sustained performance differences. This is an important endeavor, given that earlier theoretical work suggests that random influences can produce sustained patterns of performance (Denrell, 2004b; Henderson et al., 2012; Levinthal, 1991). There is also a related stream of work suggesting that skilled managers can take advantage of luck through anticipation.
in order to sustain performance (Denrell et al., 2012; Mauboussin, 2012; Powell et al., 2011). Yet, to date, solid empirical grounding to quantitatively evaluate the role of luck while controlling for skill has eluded these contributions (Liu et al., 2016). We approach this problem through a novel empirical method of comparing simulated outcomes to actual performance data. We thus contribute to studies of luck and performance with the view that luck plays an identifiable role in performance differences that must be considered by researchers and managers.

Second, we help to better bound skill-based explanations of sustained performance that usually characterize luck as an inconsequential, unexplained variance (i.e. “noise”) in sustained performance. As we reviewed above, because luck is treated as stochastic noise, prior work attributes sustained performance differences as evidence of skill. These earlier works are, thus, used to attribute sustained performance variance to CEOs and to explain and justify elevated compensation. Our findings, in contrast, demonstrate that sustained differences in luck comprise a measureable part of sustained differences in performance outcomes. As we described in our literature review above, related studies also offer important and valid industry- and firm-specific mechanisms to explain sustained performance differences (Barney et al., 2001; Eisenhardt et al., 2000; Gompers et al., 2010; Van de Ven et al., 1984). We add the idea of a quantifiable influence of luck that contributes to these discussions by providing researchers a more complete view of performance heterogeneity. We thus contribute to the growing discussion on the role of luck versus CEO skill with respect to firm performance.

Furthermore, our study highlights and speaks to the many different dimensions and forms of luck that must be considered. Specifically, prior literature accounts for luck via industry-specific and market-wide influences (e.g., Bertrand and Mullanathan, 2001; Garvey et al., 2006);
however, as our results demonstrate, the remaining benchmark-adjusted performance attributed to the CEO-specific “skill” portion may still reflect a substantial amount of luck. Consequently, the measure of luck we introduce is distinct from and complementary to that considered by studies exploring whether executives are rewarded for system-wide economic events (Garvey et al., 2006).

Implications for studies of CEO compensation and performance evaluation

We also contribute to studies of CEO attribution and compensation. In particular, we highlight the importance of considering not only the extent to which CEOs outperform their benchmarks and peers but also the extent to which they are expected to do so by sampling variation alone. As such, our study adds to the vast literature on relative performance evaluation (Gibbons et al., 1990; Holmstrom, 1982; Jenter et al., 2006; Murphy, 1999), particularly given our findings that differences in luck (as distinct from differences in skill) can cause substantial differences in relative performances across CEOs/firms. That is, luck may be a large contributing factor to what is being deemed as managerial skill. Thus, our study also adds color to the findings that performance itself is not the predominant driver of variation in executive compensation (Tosi et al., 2000).

Evidence suggests that greater consideration is given to the outcome than to the process, with extreme performances attracting the most attention (Barney, 1997; Denrell et al., 2011). News articles praise or denounce CEOs for their firms’ stock-return performances, and Business Week annually publishes the best and worst chief executives based on profitability and

---

18 “Peter Cartwright of Calpine, a firm that develops and runs gas-fired power plants. This is not a profitable venture at the moment, and the average annual return to shareholders over the past six years has been -7%. For this unelectrifying performance Cartwright has pocketed an average annual $13 million” (Forbes, 2005).
changes in shareholder value. Similarly, Forbes ranks the best and worst CEO performances, declaring, “some [executives] are so bad they should be paying the shareholders” (Forbes, 2005). In sum, if luck is an important influence in performance outcomes, should superior performance, per se, be so highly lauded or rewarded?

Perhaps a final note on CEO compensation is that, as Barney (1997; p. 17) writes: “What prescriptive advice can we give to managers given that the role of luck is important, ‘that they should ‘be lucky’?” Our prescriptive advice is take the countervailing view: that management systems should not reward or punish on metrics that seem greatly influenced by luck. Given that we show how luck can drive a significant portion of performance compensation, perhaps there should be better limits on CEO compensation relative to the lowest employee salary as done in other national contexts, where collective effort is recognized. For instance, in Japan, CEOs at the largest public companies earned, on average, approximately $1.48 million in 2010. In the United States, similar CEOs’ total compensation was approximately ten times greater, at about $16.7 million (Milhaupt et al., 2014). In sum, since we show that luck can account for a material part of performance, such differentials in pay differentials might be best be attenuated both internally as luck affects the whole firm, and in comparison to other national contexts.

CONCLUSION

In this paper, we provide evidence that the true underlying differences in CEO skill in terms of sustained exceptional performance may be substantially smaller than suggested by the dispersion in actual performance outcomes. Firm performance is a crucial factor in how CEOs are evaluated, with extreme performances attracting the most attention. However, as our simulation results demonstrate, we find that a substantial number CEOs from a large cross-section are
certain to be repeatedly lucky or unlucky, guaranteeing extreme differences in long-term performance outcomes even if everyone were equally skilled and put forth the same amount of effort. This paper also speaks to how we interpret extreme changes in performance surrounding executive turnovers, since a substantial number of CEO-turnover events are expected to result in significant performance changes (even in a simulated economy where the departing CEOs are no better or worse than their replacements).

Evaluating how much variation in performances we can expect due solely to differences in luck, thus, has important implications for organizational strategy, for replacement decisions, for designing incentive contracts, and more broadly, for how firm performance outcomes are interpreted in attempting to measure managerial effort or ability.
References


Denrell J, Liu C. 2011. Are the highest performers the most impressive?


Mitchell W, Tsui AS. 2012. Research in Emerging Economy Contexts: Selected Articles from MOR and three SMS journals (GSJ, SEJ, and SMJ).


Table 1. CEO and firm characteristics

This table presents summary statistics on CEO and firm characteristics for our sample of S&P 1500 firms during the period of 1992 to 2015. Panel A presents CEO characteristics, where: CEO Age is the CEO’s age in years; CEO Age (≥60) Indicator is an indicator variable that equals one for CEOs who are at least 60 years of age, and zero otherwise; and CEO Tenure is the incumbent CEO’s tenure (as CEO of that firm) in years. Panel B presents firm characteristics, where: Total Assets is the book value of total assets in millions; MV(Equity) is the market value of equity in millions; Earnings Per Share (EPS) is net income per share outstanding, including extraordinary items; and Return on Assets (ROA) is net income scaled by total assets.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>(Std. Dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. CEO characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEO Age</td>
<td>55.81</td>
<td>(7.02)</td>
</tr>
<tr>
<td>CEO Age (≥60) Indicator</td>
<td>0.29</td>
<td>(0.45)</td>
</tr>
<tr>
<td>CEO Tenure</td>
<td>7.11</td>
<td>(7.10)</td>
</tr>
<tr>
<td><strong>Panel B. Firm characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Assets ($million)</td>
<td>17,132</td>
<td>(96,377)</td>
</tr>
<tr>
<td>MV(Equity) ($million)</td>
<td>15,287</td>
<td>(48,361)</td>
</tr>
<tr>
<td>Earnings Per Share (EPS)</td>
<td>1.806</td>
<td>(3.26)</td>
</tr>
<tr>
<td>Return on Assets (ROA)</td>
<td>0.058</td>
<td>(0.08)</td>
</tr>
<tr>
<td>No. of firm-years</td>
<td>22,617</td>
<td>---</td>
</tr>
</tbody>
</table>
Table 2. Differences in CEO performances

This table presents the empirical versus bootstrapped differences in performance outcomes across CEOs. For all CEO-firm pairs having at least 24 months of returns data (N=4,080), we estimate monthly industry/market-model alphas by regressing monthly excess returns on the excess market return and the relevant excess industry return (based on the Fama-French 30-industry classification). The ‘empirical difference’ is the actual observed difference. The ‘bootstrapped expected difference’ is the expected performance differential based on simulations under the assumption that all CEO-firm pairs are equally skilled (i.e., true alphas are equal). The bootstrapped p-value, presented below in brackets, reports the proportion of bootstrap iterations yielding a performance differential even more extreme than the actual observed difference. A visual representation of the key information contained in this table is provided in Figure 2.

| Panel A. Actual versus bootstrapped differences among best performing CEOs |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| | Top 10% (avg) minus mean | 90th pctl minus median | 99th pctl minus 90th pctl | Maximum minus 99th pctl |
| Empirical difference (%) | 2.49 | 1.39 | 2.82 | 6.60 |
| Bootstrapped expected difference (assuming no differences in skill) | 2.36 | 1.34 | 2.43 | 6.90 |
| Bootstrapped p-value | [0.01] | [0.09] | [0.02] | [0.46] |

| Panel B. Actual versus bootstrapped differences among worst performing CEOs |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| | Bottom 10% (avg) minus mean | 10th pctl minus median | 1st pctl minus 10th pctl | Minimum minus 1st pctl |
| Empirical difference (%) | -2.07 | -1.16 | -2.12 | -2.11 |
| Bootstrapped expected difference (assuming no differences in skill) | -2.30 | -1.33 | -2.30 | -5.96 |
| Bootstrapped p-value | [1.00] | [1.00] | [0.88] | [1.00] |
Table 3. Frequency of extreme empirical versus bootstrapped CEO performances

This table presents the cumulative number of CEO-firm performances above or below a certain threshold. For each CEO-firm pair having at least 24 months of returns data (N=4,080 pairs), we estimate the monthly alpha and corresponding t-statistic by regressing monthly excess returns on the excess market return and the relevant excess industry return (based on the Fama-French 30-industry classification). The row labeled ‘Empirical outcome’ reports the number of actual estimated alphas or t-statistics that exceed the reported threshold. The rows under ‘Bootstrapped outcome’ report the number of simulated alphas or t-statistics that are expected to exceed the reported threshold based on a bootstrapped distribution under the assumption that all performers have equal true alphas of zero. The corresponding 5th and 95th percentiles of each of these simulated distributions are reported below in brackets.

<table>
<thead>
<tr>
<th></th>
<th>Frequency of alphas ≥ ...</th>
<th>Frequency of t-statistics ≥ ...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Empirical outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>991</td>
<td>538</td>
<td>313</td>
</tr>
<tr>
<td>Bootstrapped outcome:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>338</td>
<td>201</td>
</tr>
<tr>
<td>[521, 589]</td>
<td>[312, 366]</td>
<td>[180, 222]</td>
</tr>
</tbody>
</table>

Panel B. Frequency of poor CEO performances

<table>
<thead>
<tr>
<th></th>
<th>Frequency of alphas ≤ ...</th>
<th>Frequency of t-statistics ≤ ...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.0%</td>
<td>-2.5%</td>
</tr>
<tr>
<td>Empirical outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>64</td>
<td>104</td>
</tr>
<tr>
<td>Bootstrapped outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>116</td>
<td>201</td>
</tr>
<tr>
<td>[56, 83]</td>
<td>[101, 134]</td>
<td>[179, 222]</td>
</tr>
</tbody>
</table>
Table 4. Performances surrounding CEO turnover

This table presents the average monthly alphas over a three-year period of the departing and incoming CEOs. For each turnover event having at least 24 months of returns data on each side (i.e., both pre- and post-arrival), we estimate the monthly alpha for the departing and incoming CEOs by regressing monthly excess returns on the excess market return, the relevant excess industry return (based on the Fama-French 30-industry classification). Panel A presents the average performances surrounding the departure of negative-alpha performers, and Panel B presents the average performances surrounding the departure of positive-alpha performers.

| Panel A. Departing CEO has negative-alpha performance |
|-----------------------------------------------|-------------------|-------------------|
| Departing CEO monthly alpha (%)               | -0.99             | (0.85)            |
| Incoming CEO monthly alpha (%)                | 0.44              | (1.55)            |
| Difference in performance                     | 1.43              | (1.91)            |
| No. of observations                           | 1,212             | ---               |

| Panel B. Departing CEO has positive-alpha performance |
|-----------------------------------------------|-------------------|-------------------|
| Departing CEO monthly alpha (%)               | 1.22              | (1.23)            |
| Incoming CEO monthly alpha (%)                | 0.27              | (1.40)            |
| Difference in performance                     | -0.94             | (1.85)            |
| No. of observations                           | 1,505             | ---               |
Table 5. Frequency of extreme empirical versus bootstrapped changes in performance surrounding CEO turnovers

This table presents the cumulative number of performance improvements (above a certain threshold) following the departure of a negative-alpha CEO (Panel A), and the cumulative number of performance declines (below a certain threshold) following the departure of a positive-alpha CEO (Panel B). For each turnover event having at least 24 months of returns data on each side, we estimate the change in monthly alpha (based on the 36 months prior to through 36 months following the arrival of the new CEO, skipping the turnover-event month) by regressing monthly excess returns on the excess market return, the relevant excess industry return (based on the Fama-French 30-industry classification), and an indicator variable that equals one in the months following the new CEO’s arrival, and zero otherwise. The row labeled ‘Empirical outcome’ reports the number of actual changes in estimated monthly alphas (or t-statistics thereof) that exceed the reported threshold, and the row labeled ‘Bootstrapped outcome’ reports the number of simulated changes (or t-statistics thereof) that are expected to exceed the reported threshold based on a bootstrapped distribution under the assumption that there is no difference in true skill between the incoming and outgoing CEOs (i.e., true $\Delta \alpha = 0$). The corresponding 5th and 95th percentiles of each of these simulated distributions are reported below in brackets.

<table>
<thead>
<tr>
<th>Panel A. Performance improvements following the departure of a negative alpha CEO (N = 1,212)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of turnaround performances $\geq$ ...</td>
</tr>
<tr>
<td>$1.0%$</td>
</tr>
<tr>
<td>Empirical outcome</td>
</tr>
<tr>
<td>Bootstrapped outcome:</td>
</tr>
<tr>
<td>[606, 654]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Performance declines following the departure of a positive alpha CEO (N = 1,505)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of turnaround performances $\leq$ ...</td>
</tr>
<tr>
<td>$-6.0%$</td>
</tr>
<tr>
<td>Empirical outcome</td>
</tr>
<tr>
<td>Bootstrapped outcome:</td>
</tr>
<tr>
<td>[59, 82]</td>
</tr>
</tbody>
</table>