

Measuring Luck in CEO Outperformance^{* †}

Emma Dong
Vice President
Analysis Group

Seoyoung Kim
Associate Professor of Finance
Santa Clara University

October 2026

Firm performance is a crucial factor in how CEOs are evaluated. However, a CEO can be repeatedly lucky or unlucky, adding noise even to long-run performance outcomes as a measure of managerial ability. In this study, we examine how much of the observed cross-sectional dispersion in long-run performances can be attributed to differences in luck rather than differences in skill. Using bootstrap simulations across 4,080 CEO-firm pairs, we find that the best performing CEOs perform too well relative to the median to be explained by luck alone. However, the true underlying differences in skill are substantially smaller than suggested by simply looking at the raw performance differential. Interestingly, repeatedly unlucky CEOs from the simulations vastly underperform the worst performances observed empirically, suggesting that unlucky performers are terminated in reality whereas they are allowed to continue to be significantly and persistently unlucky in a simulated economy.

JEL Classification: G3, G30, M40, M49

Keywords: luck, randomness, sustained performance, performance evaluation

^{*} We thank M. Apu II, Olly D., Sanjiv Das, David Denis, Diane Denis, Mara Faccio, Byoung-Hyoun Hwang, Di Li, John McConnell, Max Moroz, Jin Xu, Deniz Yavuz, and Xiaoyan Zhang for helpful comments.

[†] The authors can be reached at: emma.dong@analysisgroup.com, srkim@scu.edu.

1. Introduction

Firm performance is a crucial factor in how CEOs are evaluated. Strong firm performance is associated with higher compensation (Murphy, 1985) and subsequent CEO appointments at other firms (Fee and Hadlock, 2003) as well as with subsequent board appointments (Brickley, Linck, and Coles, 1999). Likewise, poor performance is associated with a greater likelihood of CEO turnover (Warner, Watts, and Wruck, 1988; Weisbach, 1988; Kaplan and Minton, 2008; Jenter and Kanaan, 2008) and even litigation (Donelson, McInnis, Mergenthaler, and Yu, 2010).

Moreover, anecdotal evidence suggests that greater consideration is given to outcome than process, with extreme performances attracting the most attention. News articles praise or denounce CEOs for their firms' stock-return performances,¹ and Business Week annually publishes the best and worst chief executives based on profitability and changes in shareholder value. Similarly, Forbes ranks the best and worst CEO performances, declaring that "some [executives] are so bad they should be paying the shareholders" (Forbes, 2005).

However, a CEO can be lucky or unlucky (and repeatedly so), adding noise to even long-run performance outcomes as a measure of managerial ability and soundness of judgment. That is, even accounting for systematic economic events over which the CEO has no control, the remaining firm-specific "skill" portion of his performance may not reflect skill after all. Chance events may accumulate to produce sustained profitability in a firm (Denrell, 2004), and extreme success may actually reflect poor skill and judgment (Denrell, 2005; Denrell and Fang, 2010; Denrell and Liu, 2012). In addition to the false positives, luck can also produce false negatives. For instance, Khanna and Poulsen (1995) argue that the actions taken by CEOs of firms filing for Chapter 11 bankruptcy

¹ "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)

are similar to those taken by CEOs of otherwise comparable, matched firms that did not experience such financial distress

More generally, the large cross-section of CEOs guarantees the occurrence of repeatedly lucky and repeatedly unlucky CEOs, thereby generating extreme differences in long-run performance outcomes even if everyone were equally skilled and put forth the same amount of effort. Thus, a natural question arises as to how much of the actual observed cross-sectional dispersion in outcomes can be attributed to differences in luck as opposed to differences in skill. That is, how much variation in long-run performances can we expect in an economy where all CEOs are equally skilled, and in comparison, what does the actual empirical distribution of outcomes tell us about the true underlying differences in ability across CEOs?

To explore this question, we employ a unique cross-sectional bootstrapping method, simulating a time-series of stock returns for each separate CEO-firm pair under the assumption that all CEOs are equally capable. That is, we impose the null that the true benchmark-adjusted long-run performances are the same across all CEO-firm pairs. Thus, any performance differential we observe in a simulated cross-section of CEOs is entirely due to differences in luck rather than differences in skill, providing a benchmark of the extent to which CEOs are expected to substantially over- or underperform by chance alone.

We use stock returns, as opposed to alternative performance metrics, because stock returns provide more frequent data points, which is necessary for reliable bootstrapping. Moreover, insofar as the market can only estimate, but not fully know, the value added or destroyed by an incoming CEO, stock returns provide timely, ongoing assessments of the CEO's accomplishments or failures, the average of which provides a metric of the total value created or destroyed by the CEO's actions

during his tenure at a given firm. Thus, changes in shareholder value reflect the extent to which the CEO has exceeded or failed ex-ante performance expectations.

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,² 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 attributable to luck. That is, although the best-performing CEOs perform too well relative to the median to be completely explained by differences in luck alone, the simulations operating under the (extreme) assumption that all CEOs are equally skilled approximate the observed actual data quite well.

A similar observation applies when we extend the analysis to poor performance outcomes. We observe that, in actuality, 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. Notably, the simulations don't fire repeatedly unlucky CEOs, while in reality, a CEO who repeatedly underperforms must typically tender his resignation.

In additional analyses, we examine performance changes surrounding CEO turnovers. Changes in performance are often used to gauge the success of managerial replacement decisions. However, when evaluating CEOs by the extent to which they improve or impair performance, the

² Specifically, these benchmark-adjusted long-run performances are the alphas obtained from regressing excess monthly returns on excess market and excess industry returns.

question remains as to the performance changes we expect to observe even if the departing CEO is no better or worse than his replacement. To explore this question, we form simulated cross-sections of turnaround performances based on the 36 months prior to through 36 months following the arrival of the new CEO, assuming that there is no difference in true skill between the predecessor and his replacement. We then compare the simulated changes in performances with the actual changes observed in the data. In contrast to the previous simulations, this analysis allows for differing levels of skill across CEO-firm pairs; the only assumption is that an outgoing CEO is of equal skill as his direct replacement.

Empirically, 194 CEOs outperformed their negative-alpha predecessors by at least 3% per month, and 175 underperformed their positive-alpha predecessors by at least 3% per month. In comparison, the bootstrap simulations indicate that, under the assumption that all incoming CEOs have the same ability as their predecessors (i.e., true $\Delta\alpha=0$), 289 incoming CEOs are expected to enjoy such positive improvements in monthly performance by luck alone; 303 are expected to suffer such declines. That is, a substantial number of CEO replacements are expected to result in extreme performance changes, even in an economy where the incoming and departing CEOs are equally skilled and put forth the same amount of effort. Similar observations apply when we examine the frequency of extreme t -statistics (of the difference in outperformance). The results further 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.

Overall, the results show that we can expect substantial differences in realized performance outcomes even if all CEOs in a given sample are equally skilled, suggesting that the true underlying differences in CEO skill are substantially smaller than implied by simply looking at the raw difference

in performance outcomes. Thus, this study highlights 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, since even accounting for industry- and market-wide movements, the remaining “skill” portion may still reflect a substantial amount of luck. Consequently, the measure of luck introduced here is distinct from and complementary to that considered by studies exploring whether executives are rewarded for system-wide economic events (Bertrand and Mullainathan, 2001; Garvey and Milbourn, 2006), and thereby adds to the vast literature on relative performance evaluation (e.g., Holmstrom, 1982; Gibbons and Murphy, 1990; Murphy, 1999; Jenter and Kanaan, 2008).

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 an economy where the departing CEOs are truly no better or worse than their replacements). Thus, this paper also complements the studies exploring whether industry- or firm-specific factors dominate performance outcomes, and whether corporate strategy matters (e.g., Brush, Bromiley, and Hendrick, 1999; Bowman and Helfat, 2001; Ruefli and Wiggins, 2003). Similarly, this paper pertains to studies examining the effect of firm performance on CEO labor-market outcomes (e.g., Coughlan and Schmidt, 1985; Warner, Watts, and Wruck, 1988; Weisbach, 1988; Fee and Hadlock, 2003; Kaplan and Minton, 2008; Jenter and Kanaan, 2008), and more broadly, to any study pertaining to the theory and use of performance measurement (Bourne, 2013; Micheli and Mari, 2013).

This paper is organized as follows. Section 2 describes the data and provides summary statistics; Section 3 describes the bootstrap method used to evaluate CEO performance; Section 4 presents the results; and Section 5 provides concluding remarks.

2. Performance-Evaluation Methodology and Simulation Design

2.1. Variables

Our key dependent variable is *Excess Returns*, denoted $r_{i,t}$, which is measured 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 . 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 to evaluate performance across firms and evaluate changes in performance across CEOs within the same firm.

We use stock returns rather than accounting profitability metrics because stock returns provide more frequent data points, which are necessary for reliable bootstrapping. Specifically, the use of even quarterly accounting figures would require a *minimum* CEO tenure of six years for sample selection,³ 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 executive team's accomplishments or failures.

We have two independent variables against which we assess a CEO's benchmark-adjusted performance. First, to account for the market-wide economic conditions in which a firm is embedded, we control for the monthly excess market return at time t , denoted $MKTRF_t$ (i.e., the average excess stock return across all publicly traded firms). Second, because prior research has

³ 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).

found that the economic cycles and conditions across industries account for a substantial portion of observed performance variation (McGahan, 1999), we also control for the monthly excess industry return for firm i 's industry at time t , denoted $INDRF_{i,t}$ (i.e., the average excess stock return across all publicly traded firms in the same industry as firm i).

2.2. Measuring Long-Run Outperformance

We begin by establishing a baseline model for return performances which we then use to form the basis of our bootstrap simulations. To obtain each CEO's actual, long-run benchmark-adjusted performance accounting for general market- and industry-wide price movements, we estimate the following time-series regression for each CEO-firm pair, i :

$$r_{i,t} = \alpha_i + \beta_{i,MKT} \cdot MKTRF_t + \beta_{i,IND} \cdot INDRF_{i,t} + \varepsilon_{i,t}, \quad (1)$$

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.⁴ That is, we estimate 4,080 separate time-series regressions for each CEO-firm pair in our sample, and we obtain 4,080 $\hat{\alpha}_i$'s (i.e., "alphas"), which represent the actual long-run historical outperformances for each CEO-firm pair i .

By using the benchmark-adjusted performance (alpha) from a time-series regression of monthly stock returns across the CEO's entire tenure, we note that we are comparing differences in sustained performance across CEO-firm pairs. Specifically, to achieve a high alpha, a CEO must persistently outperform his benchmarks, since a few ad-hoc months of stellar performance will be

⁴ Obtained from Ken French's website:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We observe very similar results using the 49-industry classification. Similar observations apply when we use a three-factor model.

washed out if most months during the CEO's tenure are lackluster. Thus, the long-run outperformance throughout the duration of a CEO's tenure can be calculated by multiplying the CEO's alpha by the number of months in the entire time series of that particular CEO-firm pair, or the CEO's alpha can be simply annualized by multiplying by 12.

To be clear, skill is not a uniquely defined concept, and in removing the market and industry component of returns, we may remove aspects of CEO skill. For instance, part of a CEO's job is to determine the firm's optimal exposure to external factors (Gopalan, Milbourn, and Song, 2010). Furthermore, CEOs can collectively influence the aggregate performance measures used as benchmarks (Aggarwal and Samwick, 1999). 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 versus difference in skill. To this end, we study the best and worst (out)performances, irrespective of the overall level of skill in the economy and irrespective of whether CEOs have market-timing skill.

2.3. Simulation Method

Next, for each CEO-firm pair i , we use the OLS parameter estimates from regression equation (1), $\hat{\beta}_{i,MKT}$ and $\hat{\beta}_{i,IND}$, with corresponding time series of residuals, $\hat{\varepsilon}_{i,1}, \dots, \hat{\varepsilon}_{i,T_i}$, to form the basis of the simulated time series of returns for each CEO. That is, for each CEO-firm pair i , we begin by grounding the CEO's simulated performance using his baseline performance based on the historical market and industry returns and his firm's performance sensitivity to these macro factors (i.e., $\hat{\beta}_{i,MKT} \cdot MKTRF_t + \hat{\beta}_{i,IND} \cdot INDRF_t$). We then sample with replacement from the residual vector (i.e., $\hat{\varepsilon}_{i,1}, \dots, \hat{\varepsilon}_{i,T_i}$), assigning equal probability to each $\hat{\varepsilon}_{i,t}$, and we construct a simulated time-series

of error terms, $\varepsilon_{i,1}^*, \dots, \varepsilon_{i,T_i}^*$, which represents a series of lucky and unlucky surprises experienced by this simulated CEO.

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 $\hat{\varepsilon}_{i,t}$'s may be selected multiple times and others may not be selected at all.⁵ Using this nonparametric bootstrap procedure, we simulate a time-series of returns for each CEO-firm pair i on whom we impose the constraint that his true ability to repeatedly outperform market and industry benchmarks is equal to some pre-specified magnitude (i.e., true $\alpha_i = \alpha_{NULL}$).⁶ Nonetheless, a bootstrap sample may yield an estimated alpha that is substantially different from α_{NULL} since, by chance, it may have drawn more of the positive (or more of the negative) $\hat{\varepsilon}_{i,t}$'s.

That is, the bootstrap data-generating process, which produces a separate T -element vector of simulated returns for each CEO-firm pair i , is specified as:

$$\begin{bmatrix} r_{i,1}^* \\ r_{i,2}^* \\ \vdots \\ r_{i,T}^* \end{bmatrix} = \begin{bmatrix} \alpha_{NULL} + \hat{\beta}_{i,MKT} \cdot MKTRF_1 + \hat{\beta}_{i,IND} \cdot INDRF_1 + \varepsilon_{i,1}^* \\ \alpha_{NULL} + \hat{\beta}_{i,MKT} \cdot MKTRF_2 + \hat{\beta}_{i,IND} \cdot INDRF_2 + \varepsilon_{i,2}^* \\ \vdots \\ \alpha_{NULL} + \hat{\beta}_{i,MKT} \cdot MKTRF_T + \hat{\beta}_{i,IND} \cdot INDRF_T + \varepsilon_{i,T}^* \end{bmatrix}, \quad \varepsilon_{i,t}^* \sim EDF(\hat{\varepsilon}_i) \quad (2)$$

in which $\hat{\beta}_{i,MKT}$ and $\hat{\beta}_{i,IND}$ are the OLS parameter estimates for CEO-firm pair i obtained from regression equation (1), $MKTRF_t$ is the monthly excess market return at time t , $INDRF_{i,t}$ is the monthly excess industry portfolio return for firm i 's industry at time t based on the Fama-French (1997) 30-industry classification, and $EDF(\hat{\varepsilon}_i)$ is the empirical distribution function that assigns

⁵ 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.

⁶ Without loss of generality, set $\alpha_i = \alpha_{NULL} = 0$, since the purpose of this study is to examine the cross-sectional differences in outperformance that occur from sampling variation.

equal probability, T_i^{-1} , to each $\hat{\varepsilon}_{i,t}$ in the residual vector $\hat{\varepsilon}_i = [\hat{\varepsilon}_{i,1} \quad \hat{\varepsilon}_{i,2} \quad \cdots \quad \hat{\varepsilon}_{i,T}]'$.⁷ Then, using this simulated time series of returns, we estimate the following regression equation:

$$r_{i,t}^* = a_i + b_{i,MKT} \cdot MKTRF_t + b_{i,IND} \cdot INDRF_{i,t} + e_{i,t}, \quad (3)$$

where $r_{i,t}^*$ is the simulated 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.⁸

Finally, we implement a cross-sectional bootstrap procedure to evaluate CEO performances accounting for the large cross-section to which they belong.⁹ Specifically, we follow the above process for all CEO-firm pairs, $i = 1, \dots, N$, and we repeat this process $B=999$ times to form 999 cross-sections of $N=4,080$ 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 sustained luck as opposed to differences in skill, providing a benchmark of the extent to which CEOs are expected to achieve repeatedly positive or repeatedly negative benchmark-adjusted performances purely by chance.

⁷ 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 $T_i^{-1} \sum \hat{\varepsilon}_{i,t}^2 = T_i^{-1} (T_i - K) \hat{s}^2$, in which K equals the number of regressors and $\hat{s}^2$ is the unbiased estimator of σ_ε^2 .

⁸ Obtained from Ken French's website:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We observe very similar results using the 49-industry classification. Similar observations apply when we use a three-factor model.

⁹ Other studies using this procedure include Kosowski, Timmerman, Wermers, and White (2006) and Kosowski, Naik, and Teo (2007).

¹⁰ B is chosen such that $\lambda \cdot (B+1)$, in which λ is the size of the test, is an integer. Otherwise, the probability of Type I error cannot be exactly λ (though this becomes increasingly irrelevant for large B).

3. Data

3.1. Sources

Our sample period spans 1992 to 2015 and consists of the S&P 1500 CEO-firm pairs, which we obtain from the Standard & Poors Executive Compensation database (hereafter, Execucomp).¹¹ 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. We exclude those CEO-firm pairs whose stock prices dip below \$5.00 during that CEO's term to avoid estimation problems arising from microstructure issues inherent in small, illiquid stocks (such as the bid-ask bounce).¹²

Our final sample consists of 22,617 firm-years, with 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.

3.2. 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, and 29% of CEOs are 60 years of age or older. The average CEO tenure is 7.1 years, and the average firm has \$17.1 billion in total assets, with an average market capitalization of \$15.3 billion.

¹¹ Prior to 1994, the Execucomp database does not provide compensation data for the entire S&P 1500.

¹² Other studies using this \$5.00 filter include Jegadeesh and Titman (2001) and Chan, Chan, Jegadeesh, and Lakonishok (2006).

4. Results

4.1. Differences in CEO performances

We begin by examining the actual best and worst benchmark-adjusted, long-run CEO performance relative to the average performance outcome. We then compare these observed differences with the difference 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, show that the difference in benchmark-adjusted performances between the 90th percentile and median CEOs (Table 2, Panel A) 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 90th 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% differential that is attributable to greater skill. 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 attributable to greater luck instead of greater skill.

The bootstrapped p -value of 0.09 signifies that the simulated performance differential between the 90th percentile and median CEOs exceeds the actual observed difference of 1.39% in 9% of the 999 bootstrapped cross-sections. Thus, while the performance differential observed in actuality exceeds the simulated differences in performance (under the assumption that all CEOs are equally skilled), the 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. However, by sampling variation alone, we can expect the 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 not attributable to greater luck (bootstrapped p -value = 0.01). Furthermore, although the best performer outperformed the 99th 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 99th 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).

Similarly, we extend our analysis to poor performance outcomes (Table 2, Panel B). We observe that the actual benchmark-adjusted performance differential between the median and 10th percentile CEOs was 1.16% per month. In comparison, the simulation results indicate that, by poor luck alone, we would expect the 10th percentile CEO to underperform the median by 1.33% per month, and in all of the 999 bootstrapped cross-sections, the performance differential between

the 10th percentile and median CEOs was at least as extreme as the observed difference in reality (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 sustained 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. Overall, the results highlight the importance of gauging the extent to which CEOs are expected to outperform or underperform benchmarks and peers by sampling variation alone; this particularly applies to large samples, which guarantee the occurrence of seemingly abnormal performance outcomes even if all CEOs are truly, equally skilled.

Most notably, the bootstrapped cross-sections produce repeatedly poor performers who, in our simulations, are allowed to be repeatedly unlucky over long horizons. However, in reality, CEOs are likely to be replaced following repeatedly poor performance even when the poor performance is credibly attributable to poor luck. Thus, another benefit of the simulations is that we can create a counterfactual world to estimate the extent to which very unlucky CEOs are expected to underperform if allowed to continue in their positions despite the poor performance.

4.2. Frequency of extreme CEO performances

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% (untabulated). Evaluated in isolation, the probability of a 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 by chance. 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 in actuality.

To expound 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 sustained good or bad 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 (Table 3, 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.

With regard to extreme poor performances (Table 3, Panel B), we observe that in actuality, 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. Overall, the bootstrapped cross-sections yield more repeatedly poor performers than observed in the actual data. Similar observations apply when we extend our analysis to the frequency of extreme t -statistics (of alphas).

Overall, the results show that in a large sample, many unskilled performers are sure to be sufficiently and persistently lucky (or unlucky) so as to produce outcomes that are statistically significant at conventional cutoffs. That is, a substantial number of CEOs are expected to achieve extreme benchmark-adjusted performance outcomes based on extreme luck (as opposed to extreme skill), which further highlights the importance of evaluating even benchmark-adjusted performances within context of the large cross-section to which they belong.

4.3. Extreme 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. Thus, in our final set of analyses, we assess changes in benchmark-adjusted performances surrounding CEO turnovers.

We begin by examining actual 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% (Table 4, 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% (Table 4, 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 truly 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 is due to differences in luck versus differences in skill 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 new CEO's arrival (skipping the month of the new CEO's arrival / prior CEO's departure), and we estimate the following regression:

$$\begin{bmatrix} r_{prior,1} \\ \vdots \\ r_{prior,T_{old}} \\ r_{new,1} \\ \vdots \\ r_{new,T_{new}} \end{bmatrix} = \alpha \cdot i + \alpha_{\Delta} \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_{MKT} \cdot MKTRF + \beta_{IND} \cdot INDRF + \varepsilon,^{13} \quad (4)$$

in which i is a vector of ones. Using the OLS parameter estimates, $\{\hat{\beta}_{MKT}, \hat{\beta}_{IND}\}$, and residual vector, $\hat{\varepsilon}$, from equation (4), our new bootstrap DGP is then:

$$\begin{bmatrix} r_{prior,1}^* \\ \vdots \\ r_{prior,T_{old}}^* \\ r_{new,1}^* \\ \vdots \\ r_{new,T_{new}}^* \end{bmatrix} = \hat{\beta}_{MKT} \cdot MKTRF + \hat{\beta}_{IND} \cdot INDRF + \varepsilon^*, \quad \varepsilon_t^* \sim EDF(\hat{\varepsilon}) \quad (5)$$

¹³ Results are robust to allowing for differences in market and industry loadings between the new and old CEOs' terms.

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 $\Delta\alpha = \alpha_{new} - \alpha_{prior} = 0$).¹⁴ Nonetheless, a bootstrap sample may yield an estimated change that is substantially different from zero when regressed on the augmented regression equation (4) since, by chance, it may have drawn more of the positive (or more of the negative) $\hat{\varepsilon}_{i,t}$'s for the new CEO's term than for the previous CEO's term.

Because the bootstrap data-generating process in equation (5) 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 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 (Table 5, Panel A), and 175 underperformed their positive-alpha predecessors by at least 3% per month (Table 5, 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 in a world where all incoming CEOs have the same ability as their predecessors (i.e., true $\Delta\alpha=0$). Similar observations apply when we extend our analysis to the frequency of extreme t -statistics (of the difference in alphas).

¹⁴ The bootstrapped changes in excess performance ($\hat{\alpha}_{\Delta}^*$) remain unaltered for any imposed baseline performance $\alpha_i \in \mathfrak{R}$.

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. To be clear, this is not to say that the CEO replacement decision is an inconsequential one; rather, these results highlight the importance of adequately accounting for luck in large cross-sections when interpreting extreme performance changes between an incoming CEO and his predecessor.

5. Conclusion

Firm performance is a crucial factor in how CEOs are evaluated, with extreme performances attracting the most attention. In this paper, we provide evidence that the true underlying differences in CEO skill are substantially smaller than suggested by the dispersion in actual performance outcomes.

As the simulation results demonstrate, a few random CEOs from a large cross-section are certain to be repeatedly lucky or unlucky, guaranteeing extreme differences in performance outcomes even if everyone is equally skilled and puts forth the same effort. Notably, repeatedly unlucky CEOs from the simulations vastly underperform the worst performances observed empirically, since repeatedly unlucky performers are allowed to continue in our simulated economy whereas, in reality, they are terminated before repeated bad luck has the opportunity to accumulate. The 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 an economy where the departing CEOs are truly no better or worse than their replacements).

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

References

- Aggarwal, R.K., and A.A. Samwick, 1999. Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence. *Journal of Finance* 54, 1999–2043.
- Bertrand, M. and S. Mullainathan, 2001. Are CEOs rewarded for luck? The ones without principals are. *Quarterly Journal of Economics* 116:3, 901–932.
- Bowman, E.H., and C.E. Helfat, 2001. Does corporate strategy matter. *Strategic Management Journal* 22, 1–23.
- Bourne, M., 2013. Emerging issues in performance management. *Management Accounting Research* 25, 117–118.
- Brickley, J.A., J.S. Linck, and J.L. Coles, 1999. What happens to CEOs after they retire? New evidence on career concerns, horizon problems, and CEO incentives. *Journal of Financial Economics* 52, 341–377.
- Brush, T.H., P. Bromiley, and M. Hendrickx, 1999. The relative influence of industry and corporation on business unit performance: An alternative estimate. *Strategic Management Journal* 20, 519–547.
- Chan, K., L.K.C. Chan, N. Jegadeesh, and J. Lakonishok, 2006. Earnings quality and stock returns. *Journal of Business* 79, 1041–1082.
- Coughlan, A.T. and R.M. Schmidt, 1985. Executive compensation, management turnover, and firm performance: An empirical investigation. *Journal of Accounting and Economics* 7, 43–66.
- Denrell, J., 2004. Random walks and sustained competitive advantage. *Management Science* 50, 922–934.

- Denrell, J., 2005. Should we be impressed with high performance. *Journal of Management Inquiry* 14, 292–298.
- Denrell, J., and C. Fang, 2010. Predicting the next big thing: Success as a signal of poor judgment. *Management Science* 56, 1653–1667.
- Denrell, J., and C. Liu, 2012. Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences* 109, 9331–9336.
- Donelson, D.C., J. McInnis, R.D. Mergenthaler, and Y. Yu, 2010. The timeliness of earnings news and litigation risk. *Working paper series*.
- Fama, E. and K. French, 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153–193.
- Fee, C.E. and C.J. Hadlock, 2003. Raids, rewards, and reputations in the market for managerial talent. *Review of Financial Studies* 16, 1315–1357.
- Fitza, M.A., 2014. The Use of Variance Decomposition in the Investigation of CEO Effects: How Large Must the CEO Effect Be to Rule Out Chance? *Strategic Management Journal* 35, 1839–1852.
- Fitza, M.A., 2017. How Much Do CEOs Really Matter? Reaffirming That the CEO Effect is Mostly Due to Chance. *Strategic Management Journal* 38, 802–811.
- Forbes, 2005. The Best and Worst Bosses.
- Garvey, G. and T. Milbourn, 2006. Assymetric benchmarking in compensation: executives are rewarded for good luck but not penalized for bad. *Journal of Financial Economics* 82, 197–226.
- Gibbons, R. and K.J. Murphy, 1990. Relative performance evaluation for chief executive officers. *Industrial and Labor Relations Review* 43:3, 30S–51S
- Gopalan, R., T. Milbourn, and F. Song, 2010. Strategic flexibility and the optimality of pay for sector performance. *Review of Financial Studies* 23, 2060–2098.
- Holmstrom, B., 1982. Moral hazard in teams. *Bell Journal of Economics* 13:2, 324–340.
- Huson, M.R., R. Parrino, and L.T. Starks, 2001. Internal monitoring mechanisms and CEO turnover: A long-term perspective. *Journal of Finance* 56, 2265–2297.

- Jegadeesh, N. and S. Titman, 2001. Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance* 56, 699–720.
- Jenter, D. and F. Kanaan, 2008. CEO turnover and relative performance evaluation. *Working paper series*.
- Jones, M.C., J.D. Marron, and S.J. Sheather, 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91, 401–407.
- Kaplan, S.N. and B.A. Minton, 2008. How has CEO turnover changed? *Working paper series*.
- Khanna, N. and A.B. Poulsen, 1995. Managers of financially distressed firms: Villains or scapegoats? *Journal of Finance* 50, 919–940.
- Kosowski, R., A. Timmermann, R. Wermers, and H. White, 2006. Can mutual fund ‘stars’ really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61, 2551–2595.
- Kosowski, R., N.Y. Naik, and M. Teo, 2007. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics* 84, 229–264.
- Micheli, P., 2013. The theory and practice of performance measurement. *Management Accounting Research* 25, 147-156.
- Murphy K.J., 1985. Corporate performance and managerial remuneration: An empirical analysis. *Journal of Accounting and Economics* 7, 11-42
- Murphy K.J., 1999. Executive compensation. North Holland, Amsterdam.
- Parrino, R., 1997. CEO turnover and outside succession: A cross-sectional analysis. *Journal of Financial Economics* 46, 165-197.
- Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Finance Studies* 22, 435-480.
- Quigley, T.J. and S.D. Graffin. Reaffirming the CEO Effect Is Significant and Much Larger Than Chance: A Comment on Fitza. *Strategic Management Journal* 38, 793-801.
- Ruefli, T.W., and R.R. Wiggins, 2003. Industry, corporate, and segment effects and business performance: A non-parametric approach. *Strategic Management Journal* 24, 861–879.

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

Weisbach, M.S., 1988. Outside directors and CEO turnover. *Journal of Financial Economics* 20, 431-460.

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)* 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: *CEO Turnover* is an indicator that equals one for firm-years in which a CEO turnover occurs, and zero otherwise; *Total Assets* is the book value of total assets in millions; and *MV(Equity)* is the market value of equity in millions.

Variable	Mean	(Std. Dev)
<i>Panel A. CEO characteristics</i>		
<i>CEO Age</i>	55.81	(7.02)
<i>CEO Age (≥ 60)</i>	0.29	(0.45)
<i>CEO Tenure</i>	7.11	(7.10)
<i>Panel B. Firm characteristics</i>		
<i>Total Assets</i> (\$million)	17,132	(96,377)
<i>MV(Equity)</i> (\$million)	15,287	(48,361)
No. of firm-years	22,617	---

Table 2. Differences in CEO performances

This table presents the empirical versus bootstrapped differences in performance 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.

<i>Panel A. Actual versus bootstrapped differences among best performing CEOs</i>				
	Top 10% (avg) minus mean	90 th pctl minus median	99 th pctl minus 90 th pctl	Maximum minus 99 th 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
<i>Bootstrapped p-value</i>	[0.01]	[0.09]	[0.02]	[0.46]
<i>Panel B. Actual versus bootstrapped differences among worst performing CEOs</i>				
	Bottom 10% (avg) minus mean	10 th pctl minus median	1 st pctl minus 10 th pctl	Minimum minus 1 st 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
<i>Bootstrapped p-value</i>	[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.

<i>Panel A. Frequency of strong CEO performances</i>								
	Frequency of alphas $\geq$...					Frequency of t -statistics $\geq$...		
	1.0%	1.5%	2.0%	2.5%	3.0%	1.64	1.96	2.57
Empirical outcome	991	538	313	198	126	416	218	57
Bootstrapped outcome:	554 [521, 589]	338 [312, 366]	201 [180, 222]	127 [109, 144]	84 [70, 98]	209 [186, 233]	108 [91, 126]	28 [20, 37]
<i>Panel B. Frequency of poor CEO performances</i>								
	Frequency of alphas $\leq$...					Frequency of t -statistics $\leq$...		
	-3.0%	-2.5%	-2.0%	-1.5%	-1.0%	-2.57	-1.96	-1.64
Empirical outcome	38	64	104	173	333	6	46	86
Bootstrapped outcome	69 [56, 83]	116 [101, 134]	201 [179, 222]	358 [332, 387]	667 [630, 704]	36 [26, 46]	134 [116, 152]	247 [223, 273]

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 and 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.

	Mean	(Std. Dev)
<i>Panel A. Departing CEO has negative-alpha performance</i>		
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	---
<i>Panel B. Departing CEO has positive-alpha performance</i>		
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.

<i>Panel A. Performance improvements following the departure of a negative alpha CEO (N = 1,212)</i>								
	Frequency of turnaround performances $\geq \dots$					Frequency of t -statistics (of turnarounds) $\geq \dots$		
	1.0%	1.5%	2.0%	3.0%	6.0%	1.64	1.96	2.57
Empirical outcome	643	496	353	194	30	131	61	9
Bootstrapped outcome:	631 [606, 654]	534 [510, 559]	427 [404, 449]	289 [270, 310]	76 [65, 88]	271 [250, 291]	191 [172, 209]	87 [74, 100]
<i>Panel B. Performance declines following the departure of a positive alpha CEO (N = 1,505)</i>								
	Frequency of turnaround performances $\leq \dots$					Frequency of t -statistics (of turnarounds) $\leq \dots$		
	-6.0%	-3.0%	-2.0%	-1.5%	-1.0%	-2.57	-1.96	-1.64
Empirical outcome	23	175	341	501	670	12	48	105
Bootstrapped outcome:	70 [59, 82]	303 [281, 327]	462 [438, 488]	590 [564, 617]	707 [679, 736]	76 [63, 89]	181 [163, 201]	267 [246, 289]