

HOW LONG MUST A FIRM BE GREAT TO RULE OUT CHANCE? BENCHMARKING SUSTAINED SUPERIOR PERFORMANCE WITHOUT BEING FOOLED BY RANDOMNESS

ANDREW D. HENDERSON,^{1*} MICHAEL E. RAYNOR,² and MUMTAZ AHMED³

¹ McCombs School of Business, The University of Texas at Austin, Austin, Texas, U.S.A.

² Deloitte Consulting LLP, Boston, Massachusetts, U.S.A.

³ Deloitte Consulting LLP, San Francisco, California, U.S.A.

Although sustained superior firm performance may arise from skillful management or other valuable, rare, and inimitable resources, it can also result from randomness. Studying U.S. companies from 1965–2008, we benchmark how long a firm must perform at a high level to be confident that it is something other than the outcome of a time-homogeneous stationary Markov chain defined on the state space of percentiles. We find (a) the number of sustained superior performers in Compustat, measured by ROA and Tobin's q , exceeds the number of false positives we would expect to be generated by such a process; yet (b) the occurrence of false positives is often enough to fool many observers, so (c) the identification of sustained superior performers requires particularly stringent benchmarks to enable valid study. Copyright © 2011 John Wiley & Sons, Ltd.

INTRODUCTION

Identifying the sources of consistently superior firm performance is a central theme in strategy research (Powell, 2001). Scholars have argued that sustained superiority may reflect monopoly rents arising from favorable industry structures (Porter, 1980), Ricardian rents flowing to firms that possess rare and difficult to imitate resources (Barney, 1991; Peteraf, 1993), or Schumpeterian rents

earned by innovative companies (Eisenhardt and Martin, 2000; Teece, Pisano, and Shuen, 1997). Yet aside from monopoly, Ricardian, or Schumpeterian rents, there is another source of sustained superiority—randomness (Barney, 1986).

Randomness can mislead us in the study of sustained superior performance because people mistakenly perceive patterns in random data (Loftus, 1979; Tversky and Kahneman, 1974) and concoct fanciful stories to explain historical results (Barney, 1997; Rosenzweig, 2007; Taleb, 2005). Most observers also fail to understand that many stochastic processes readily yield long streaks of seemingly exceptional outcomes that are simple byproducts of chance. This property is embodied, for example, in the Lévy arcsine law of random walks, which is well known to statisticians (Feller,

Keywords: sustained superior performance; competitive advantage; randomness; sustained competitive advantage; resource-based view

*Correspondence to: Andrew D. Henderson, McCombs School of Business, Department of Management, The University of Texas at Austin, 1 University Station, B6000, Austin, TX 78712-1178, U.S.A. E-mail: andy.henderson@mcombs.utexas.edu

1968) and has been explored in studies of firm-level resource accumulation (Denrell, 2004; Levinthal, 1990; 1991). However, stochastic processes have not received adequate attention in the study of sustained superior performance.

Being fooled by randomness is a particular concern when researchers select on the dependent variable to identify top performers for case study (e.g., Collins and Porras, 1994; Joyce, Nohria, and Roberson, 2003). This typically involves (a) setting performance benchmarks, (b) screening a large population of firms to find the few that meet those standards, and then (c) using inductive analyses to compare top performing 'success stories' to more average firms. The rub, of course, is that a top performing firm may meet a benchmark by chance, and the larger the population, the greater that possibility.

Almost all case analyses, whether published in academic journals or practitioner-oriented books such as *In Search of Excellence* (Peters and Waterman, 1982), *Built to Last* (Collins and Porras, 1994), *Good to Great* (Collins 2001), and *What Really Works* (Joyce, Nohria, and Roberson, 2003) implicitly assume that most if not all firms with performance above some specified level have achieved that success by virtue of some form of superior management. Because we do not know how many of the firms in such case studies have track records that are indistinguishable from random processes, those books' prescriptions for practice rest on untested assumptions. Similarly, even pioneering nonparametric analyses that stratify firm performance into statistically different levels (e.g., Ruefli and Wiggins, 2000; Wiggins and Ruefli, 2002) have had to assume that top performers owe their results to something other than good fortune.¹

As a first step in identifying firms with superior resources, we benchmark how often a firm must perform at a high level to believe it is not the sort of false positive that would routinely occur in a large population of identical companies whose performances change over time due to a stochastic process. Random processes take many forms, but we focus on time-homogenous Markov chains

whose state spaces are defined by performance percentiles.² This allows us to make no assumptions about how firm performance is distributed, which we show is vital (see also McKelvey and Andriani, 2005). We define *unexpected sustained superiority* as a firm's ability to achieve a highly ranked *focal outcome* (e.g., top 10% return on assets [ROA]) often enough across the firm's observed life to rule out, as a complete explanation of the firm's performance, a Markov process on the percentile state space operating on a large population of similar companies. Randomness may play some role in such a firm's performance, but it cannot explain its overall track record. In contrast, *false positives* are firms that perform well relatively often, yet have track records that are consistent with expected extremes in a large population of similar firms whose performances change annually through a Markov process defined on the percentile state space. In turn, we ask questions such as: if a firm is observed for 15 years, how often must its ROA be in the top 10–20 percent of the population to be confident that its performance is not the sort of false positive that this type of Markov process would produce?

We note three points about this study. First, there are many stochastic processes aside from the type of Markov chains that we consider, so we make no claims about ruling out all forms of randomness. Instead, we hope that our initial effort encourages future research on stochastic processes and the interplay of randomness, resources, and firm performance. Second, as Denrell (2004) points out, the possibility that a random process could produce results that closely match the outcomes of many top performing companies does not demonstrate that success is without cause or due to luck. Other, non-stochastic models may fit the data equally well, or randomness may emerge from high-dimensional causal interactions in densely connected organizational systems whose outcomes, although deterministic, are chaotic, sensitive to initial conditions, and thus difficult for managers to predict (Rivkin, 2000). Third, in addition to false positives, randomness can produce false negatives

¹ To see that statistical significance does not rule out randomness in this instance, consider a scenario in which firms move unpredictably between two conditions: fortunate and unfortunate. Statistically, performance may be significantly higher in the former condition than the latter, yet randomness could still account for any sustained superior performers that were observed.

² A Markov chain consists of a finite number of discrete states and a transition matrix that gives the probabilities of moving between states *j* and *k* in a single move. Conditional on its present state, a Markov chain's past and future states are independent (Spilerman, 1972; Singer and Spilerman, 1973).

in which firms with exceptional resources experience periods of ill fortune. False negatives are worthy of study but beyond the scope of this study.

It is difficult—and important—to distinguish between randomness and systematic firm-level heterogeneity that produces superior results (Denrell, 2004). The more firms that exceed what would be expected through random processes, the stronger the support for theories of sustained advantage and the better our opportunities for empirical study. Our goals in this study are (1) to identify sustained superior performers, if any, that exceed what a specific stochastic process would produce, and (2) to encourage future study of such firms.

RANDOMNESS AND FIRM PERFORMANCE

Unless we know what randomness might produce in a large population of firms, our ability to understand systematic drivers of sustained superior performance is limited. A number of studies have assessed the degree to which firm performance, both good and bad, persists across time (e.g., McGahan and Porter, 1999; Waring, 1996; see Wiggins and Ruefli, 2005, for a review). But none, to our knowledge, has compared the observed number of sustained superior performers to the number that we would expect by chance. Doing so is vital because randomness plays an important role in a world in which boundedly rational decision makers face strategic choices that involve high levels of uncertainty, causal ambiguity, and chaotic complexity (Barney, 1991; Levinthal, 1997; Rivkin, 2000).

Consider a discrete-time stochastic process that exhibits the Markov property because conditional on performance at $t-1$, performance at time t is independent of past results:

$$\text{performance}_{i,t} = f(\text{performance}_{i,t-1}) + \varepsilon_{i,t} \quad (1a)$$

Here, performance might be measured by firm profitability, and $\varepsilon_{i,t}$ reflects stochastic changes in the i^{th} subject's state from $t-1$ to t . Some processes with the Markov property rest on the idea that randomness can be captured by a Gaussian (i.e., normal) distribution. For instance, if performance was a continuous-time function with $0 \leq s < t$, then

$$\text{performance}_{i,t} = f(\text{performance}_{i,s}) + \varepsilon_{i,t-s} \quad (1b)$$

would reflect Brownian motion (Brown, 1828) or a Wiener process if $\varepsilon_{i,t-s}$ was normally distributed with mean zero and variance equal to $t-s$ (Wiener, 1949; Resnick, 1992).

Models such as Equation 1b assume that the step sizes in firms' stochastic evolution are normally distributed and homoskedastic across performance levels. However, empirical observation suggests otherwise. Consider that a Gaussian ordinary least squares (OLS) regression, $\text{ROA}_{i,t} = \gamma \text{ROA}_{i,t-1} + \varepsilon_{i,t}$, performed on our population of Compustat firms ($n = 243,722$ firm years) yields $R^2 = 0.0000$ and $\gamma = 0.022$ (not significant), indicating that ROA is completely uncorrelated from year to year. In comparison, the nonparametric Spearman rank-order correlation of ROA and its one-year lag is $r = 0.73$, a fairly strong association. Why this stark difference between Gaussian and rank-based statistics? The Gaussian assumptions embedded in Equation 1b are violated to such an extent that they yield misleading results, so it is vital to adopt an approach that does not rely on distributional assumptions.

In line with this, prior research indicates that relative to a normal distribution, performance measures such as ROA have (a) much narrower and sharper central peaks, (b) substantial skewness, and (c) much longer and fatter tails, so extreme performances occur relatively often (Mandelbrot, 1963, 1967; McKelvey and Andriani, 2005). In turn, outliers in skewed, fat-tailed distributions cause distributional statistics such as means, variances, and least-squares regression coefficients to be highly misleading. Many studies discard outliers (e.g., Waring, 1996), but in the study of sustained superior performance—which is an outlier phenomenon—this may eliminate the very data in which we are most interested.

Rank-order statistics involving percentiles are highly robust and provide useful information about relative standing regardless of how a variable is distributed (Kennedy, 2008; McKelvey and Andriani, 2005). Therefore, in the remainder of the paper, we model randomness using time-homogenous Markov chains in which the state space is defined by rank-based performance percentiles. This approach captures stochastic changes in performance across time, yet Markov chains on the percentile state space make no assumptions about how firm performance is distributed—cross-sectionally or longitudinally.

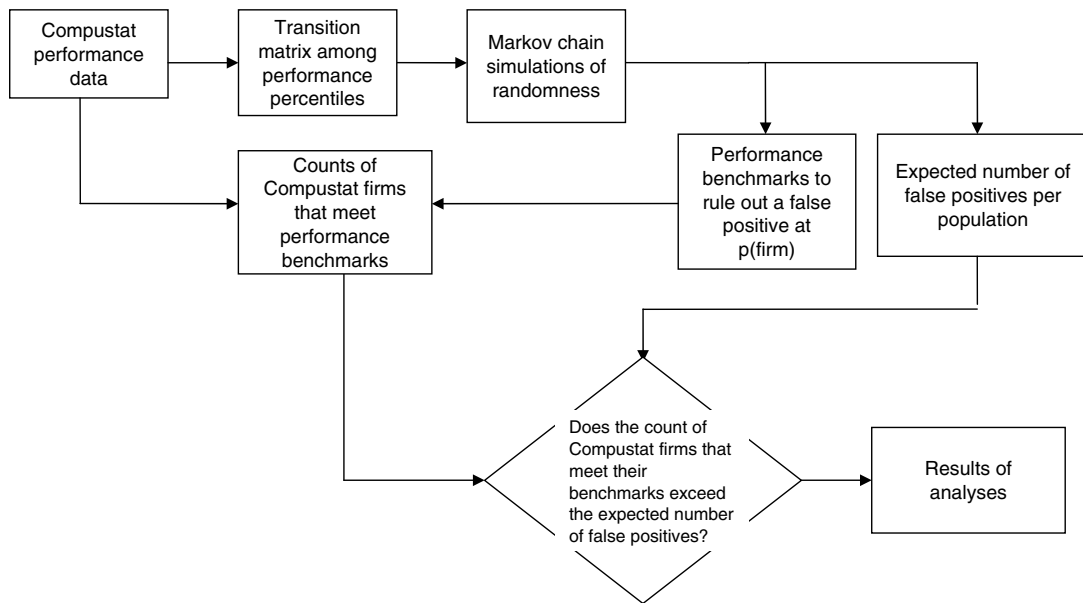


Figure 1. Summary of analysis flow

A MARKOV CHAIN APPROACH

A Markov chain consists of a finite number of discrete states and a transition matrix that gives the probabilities of moving between states j and k in a single move (Singer and Spilerman, 1973; Spilerman, 1972). We define states by performance percentiles and assess the probabilities of moving between percentiles j and k between times $t-1$ and t . We assume that each transition probability $p_{j,k}$ remains constant over time, and we make two further assumptions, which we later relax: (1) firms are homogeneous in their resources, so the same transition matrix applies to all, and (2) all firms reside in a performance state for 1 year before potentially changing state according to the transition matrix. All models are assumed to be first-order Markov, so firms in state j at time t have the same transition probabilities regardless of their past history. Finally, we allow firms to differ in their starting states at $t = 1$.

Using computer simulations and the approach summarized in Figure 1, we can assess how often a firm must perform well, on an annual basis, across the firm's observed life to be confident that its track record is not a false positive, which is an outcome that would be expected due to a Markov process on the percentile state space operating on a large

population.³ This involves setting a *benchmark* that specifies (a) a focal annual outcome (e.g., top 10% of ROA), (b) the confidence level, p_{firm} , at which we want to rule out a false positive for a specific company (e.g., $p_{firm} < 0.01$), and (c) that firm's observed life span (e.g., 30 years). For example, we might find that benchmark (top 10% ROA; $p_{firm} < 0.01$; 30 years) = 12 years, indicating that a firm observed for 30 years would need at least 12 years of top 10 percent ROA performances for us to be confident at $p_{firm} < 0.01$ that its track record is not simply an expected extreme in this type of Markov process. Firms with other observed life spans would have different benchmarks, and benchmarks would be more stringent for smaller values of p_{firm} .

As an input to the simulations that we used to establish these benchmarks, we first constructed a 100×100 population-level transition matrix by calculating the percentage of times across the Compustat population's history that real firms moved between one performance percentile and another between years $t-1$ and t . Table 1 lists two portions of the ROA transition matrix for

³ In this paper, we are interested in how often firms achieve top performances. As a result, we do not penalize firms for poor performances and leave the question of risk-return trade-offs to future study. The discussion section addresses this further. See Raynor (2007) for a broader treatment.

Table 1. Two portions of the 100 × 100 ROA transition matrix showing probabilities of moving among performance percentiles between times t-1 and t

Transition probabilities among percentiles 1–10:

		Percentile location (t)									
		1	2	3	4	5	6	7	8	9	10
Percentile location (t-1)	1	0.465	0.143	0.070	0.060	0.035	0.019	0.026	0.018	0.021	0.010
	2	0.193	0.204	0.112	0.080	0.066	0.058	0.037	0.030	0.028	0.021
	3	0.093	0.140	0.128	0.104	0.078	0.068	0.053	0.043	0.042	0.024
	4	0.047	0.107	0.119	0.110	0.079	0.076	0.053	0.055	0.037	0.037
	5	0.042	0.076	0.094	0.094	0.084	0.089	0.060	0.057	0.040	0.043
	6	0.022	0.059	0.070	0.096	0.072	0.082	0.075	0.062	0.060	0.042
	7	0.021	0.039	0.056	0.060	0.080	0.073	0.070	0.072	0.056	0.052
	8	0.017	0.029	0.047	0.059	0.074	0.075	0.070	0.060	0.065	0.048
	9	0.014	0.028	0.052	0.047	0.059	0.059	0.060	0.065	0.059	0.058
	10	0.012	0.022	0.030	0.039	0.056	0.054	0.051	0.056	0.058	0.060

Transition probabilities among percentiles 91–100:

		Percentile location (t)									
		91	92	93	94	95	96	97	98	99	100
Percentile location (t-1)	91	0.180	0.098	0.041	0.036	0.032	0.017	0.010	0.006	0.006	0.005
	92	0.097	0.153	0.070	0.054	0.047	0.027	0.020	0.011	0.008	0.004
	93	0.044	0.076	0.109	0.093	0.061	0.043	0.031	0.024	0.010	0.007
	94	0.032	0.047	0.082	0.106	0.091	0.074	0.039	0.027	0.016	0.007
	95	0.022	0.042	0.057	0.092	0.109	0.096	0.064	0.037	0.019	0.011
	96	0.017	0.032	0.046	0.064	0.092	0.118	0.110	0.061	0.033	0.010
	97	0.009	0.020	0.034	0.041	0.063	0.108	0.153	0.107	0.054	0.021
	98	0.011	0.018	0.022	0.029	0.040	0.065	0.113	0.174	0.123	0.031
	99	0.010	0.011	0.015	0.024	0.039	0.041	0.066	0.131	0.209	0.091
	100	0.005	0.009	0.016	0.014	0.024	0.023	0.036	0.056	0.118	0.248

Compustat firms, showing transition probabilities among the 10 lowest percentiles and the 10 highest. Cells along the main northwest-southeast diagonal give the probability of a typical firm's remaining in the same percentile from one year to the next. For instance, the probability of a firm's remaining in the top ROA percentile is 24.8 percent, and its chances of remaining in the bottom percentile are 46.5 percent.

By stating performance in percentiles, a transition matrix gives up a small degree of precision relative to continuous measures, yet we are freed from making distributional assumptions that are unjustified in performance data.

Operation of simulation models

Transition matrix in hand, we construct Markov chain simulation models on the percentile state space to reveal outcomes that occur stochastically in a simulated population whose number of firms,

observed life spans, and starting percentile positions exactly match the characteristics of the real Compustat population. To illustrate, a total of 430 firms were observed in Compustat for exactly 43 years, the maximum observed life span. To simulate the life history of one of those firms, we take its actual starting ROA percentile at $t = 1$ and then make 42 random draws, which produces a 43-year simulated history. Each draw is from a uniform distribution on the interval (0,1), which we map to an outcome percentile using the cumulative probabilities derived by summing across the j^{th} row of the transition matrix, where j equals the prior percentile outcome. A similar procedure is used for all other firms in the simulated population.

Using simulation models to set benchmarks

Each run of the simulation replays the population's entire observed history and provides a count of the number of focal outcomes (e.g., top 10% ROA)

attained by each simulated firm. However, randomness causes each run of the simulation to yield somewhat different results. We capture the extent of that variation using a nonparametric bootstrapping method (Efron, 1981, 2000) by rerunning the simulation many times, allowing history and randomness to play out in many possible ways, then using the overall set of results to get a very accurate picture of the distribution of Markov-based outcomes. Here, we simulated the entire histories of 1000 populations, each of which mirrors the entire history of the Compustat population.

To illustrate, consider the 430 firms in Compustat that were observed for exactly 43 years. We simulated each of those firms' lives 1,000 times, resulting in 430,000 simulated firm-life observations. The number of top 10 percent annual outcomes across those firms' simulated lives ranged from zero to 31; the median was three, and the ninety-fifth percentile was 10 such outcomes. Using those simulated results, let us set our benchmark high enough so that there is less than one chance in 100 ($p_{\text{firm}} < 0.01$) that the number of focal outcomes that a real firm achieved was consistent with a Markov process on the percentile state space, thus making it a false positive. That involves picking a benchmark that is strictly above the ninety-ninth percentile of simulated lifetime outcomes, so a real firm had to have at least 15 top 10 percent outcomes in 43 years for us to consider that it was not a false positive. We denote this as benchmark (top 10%; $p_{\text{firm}} < 0.01$; 43 years). In the same way, we set benchmarks (top 10%; $p_{\text{firm}} < \alpha$; τ years) for all combinations of $\alpha \in [0.05, 0.01, 0.002, 0.001]$, and $\tau \in [1 \text{ year}, 43 \text{ years}]$, the latter being the range of observed life spans. Firms are benchmarked against others with identical observed life spans because firms with more observations have more chances to benefit from randomness. Table 2 lists, for each observed life span, benchmarks for the number of top 10 percent ROA and Tobin's q outcomes needed to rule out a false positive at several values of p_{firm} . Similar benchmarks for top 20 percent outcomes are available from the authors.

Using simulations to set population-level confidence intervals

Regardless of how stringent our benchmarks are, some firms may meet them by chance, and the

larger the population, the more such false positives there will be. Therefore, in our next step we ask: for a Compustat-sized population of firms, how many false positives should we expect?

To assess this, we applied the benchmarks in Table 2 to the results of our 1,000 simulation runs. We found, for instance, that for top 10 percent ROA outcomes at $p_{\text{firm}} < 0.01$ there were an average of $\mu = 97.57$ firms per simulated population that met their respective benchmarks due to a Markov process on the percentile state space, with $\sigma = 9.7$ simulated firms, making $\sigma^2 = 94.04$ firms. This overall set of outcomes, within a very small degree of stochastic error, is Poisson distributed, as the nearness of the mean to the variance suggests. Due to the mixing properties of binomial processes in large samples (e.g., Daley and Vere-Jones, 2003; Hiriji, 2005), this Poisson distribution of false positives per simulated population is a general result, one that holds for the collection of outcomes produced by *any* time-stable transition matrix.

Consequently, for a population as a whole, the number of firms expected to meet their benchmarks due to randomness in a time-homogenous Markov process on the percentile state space is Poisson distributed with grand mean μ , *regardless* of how firm performance is distributed, either cross-sectionally or longitudinally. We obtain μ from our simulations and use it to construct population-level confidence intervals. In a Poisson distribution, $\sigma = \sqrt{\mu}$, so the population's 99.8 percent confidence interval is $\mu \pm 3\sqrt{\mu}$. These intervals are bounded on the left at zero, because a population cannot have a negative number of false positives. For convenience, we call this integer-rounded population-level confidence interval $\mu \pm 3\sqrt{\mu}$. Other confidence intervals can, of course, be constructed. Note that such a confidence level is conservative because it applies to the population as a whole, not to individual firms.

Finally, we transition away from simulated firms and take up our Compustat population of real firms. Given each real firm's observed life span of τ years, we compare its count of focal outcomes to the benchmarks in Table 2. If a real firm meets a given benchmark, we label it a sustained superior performer, and we count the number, n_{superior} , of such firms in Compustat. Importantly, some of those firms may be false positives because their lifetime outcomes, although unusual, are what we would expect among the positive extremes

Table 2. Number of top 10% outcomes needed to rule out a false positive at several values of $p(\text{firm})$

Life span (years)	Top 10% ROA				Top 10% Tobin's q			
	$P(\text{firm})$ < 0.05	$p(\text{firm})$ < 0.01	$p(\text{firm})$ < 0.002	$p(\text{firm})$ < 0.001	$p(\text{firm})$ < 0.05	$p(\text{firm})$ < 0.01	$p(\text{firm})$ < 0.002	$p(\text{firm})$ < 0.001
3	3							
4	3				4			
5	4				5			
6	4	6			5			
7	5	7			5	7		
8	5	7			5	8		
9	5	8	9		6	8		
10	6	8	10	10	6	8	10	
11	6	8	10	11	6	8	10	11
12	6	9	11	11	6	8	11	12
13	6	9	11	12	6	9	11	12
14	6	9	11	12	6	9	11	12
15	7	9	12	13	6	9	11	12
16	7	10	12	13	7	9	12	13
17	7	10	12	13	7	10	12	13
18	7	10	13	13	7	10	12	13
19	7	10	13	14	7	10	12	13
20	8	11	13	14	7	10	13	14
21	8	11	14	15	7	11	13	14
22	8	11	14	15	7	10	13	14
23	8	11	14	15	8	11	14	15
24	8	12	14	15	8	11	14	15
25	9	12	15	16	8	11	14	16
26	9	12	15	16	8	12	15	16
27	9	13	16	17	8	12	15	16
28	9	12	15	17	9	12	15	16
29	9	13	15	17	9	12	15	16
30	9	13	16	17	9	12	15	16
31	10	13	16	17	9	12	15	17
32	10	13	16	17	9	12	15	17
33	10	13	16	18	9	13	16	17
34	10	13	17	18	9	13	16	17
35	10	14	17	18	9	13	16	17
36	10	14	17	19	10	13	16	18
37	11	14	18	19	10	13	16	18
38	11	15	18	19	10	13	17	18
39	11	15	18	19	10	14	17	18
40	11	15	19	20	10	14	17	18
41	11	15	18	20	10	14	17	19
42	12	16	19	20	10	14	18	19
43	11	15	19	20	10	14	18	19

Blank cells indicate life spans that are too short to rule out a false positive at p_{firm} . Firms with life spans of less than three years cannot be ruled out as false positives at any of these p -values.

in a similarly large population of homogenous firms whose performance changed according to a Markov process on the percentile state space. Because of this, it is vital to compare the observed number of sustained superior performers among real firms (n_{superior}) to the confidence interval generated by our simulations ($\mu \pm 3\sqrt{\mu}$), which indicates the expected number of false positives.

Potential population-level outcomes

In comparing the number of sustained superior performers in Compustat to the expected number of false positives, three possibilities arise. First, if the number of Compustat firms that met their benchmarks fell within the 99.8 percent population-level confidence interval ($\mu - 3\sqrt{\mu} \leq n_{\text{superior}} \leq$

$\mu + 3\sqrt{\mu}$), then the sustained superiority in Compustat would be indistinguishable from outcomes generated by a simple stochastic process operating on a collection of firms with similar resources.

Second, the number of Compustat firms that met their benchmarks might exceed the upper threshold of the 99.8 percent population-level confidence interval ($n_{\text{superior}} > \mu + 3\sqrt{\mu}$), but not by much. For instance, if we expected 100 false positives (i.e., $\mu = 100$; $\mu \pm 3\sqrt{\mu} = 70$ to 130), yet we found that 160 firms in Compustat that met their respective standards (i.e., $n_{\text{superior}} = 160$), that would be a 6σ outcome ($p < 0.000001$) relative to the expected number of false positives. We would conclude that a homogenous Markov process on the percentile state space did not fully explain the sustained superiority seen in Compustat; however, studying those 160 firms could yield erroneous results because somewhere in the neighborhood of $\mu = 100$ of them *might* owe their success to randomness. As noted earlier, even if a random process produces results that closely match the outcomes of many top performing companies, that does not demonstrate that success is random because other processes may explain the data equally well. Rather, it indicates that caution is warranted (Denrell, 2004).

Importantly, the technique described here provides a way to gauge how cautious we should be. We do this by calculating θ , the ratio of unexpected sustained superior performers in Compustat to the expected number of false positives:

$$\theta = (n_{\text{superior}} - \mu) / \mu. \quad (2)$$

If θ was greater than zero but not much greater than one, then the prevalence of potential false positives would muddy the waters for anyone interested in the drivers of sustained superiority.

In comparison, the final possibility is one in which the observed number of superior Compustat firms is both outside the confidence interval for false positives ($n_{\text{superior}} > \mu + 3\sqrt{\mu}$) and substantially larger than the expected number of false positives ($n_{\text{superior}} \gg \mu$), resulting in a large θ ratio. The higher that ratio, the greater the opportunities to study sustained superior performers while being confident that we are not observing only a stochastic process of the type we consider.

To summarize, we take four steps to discern whether the number of sustained superior performers in Compustat is more than we would expect

due to one form of randomness. First, we construct a transition matrix by observing the typical probabilities among Compustat firms of moving between different performance percentiles. Second, we use the transition matrix to simulate many populations of similar firms that behave according to the matrix's stochastic properties. Third, we use the simulation results to establish (a) benchmarks that a firm must meet to be counted as a sustained superior performer, and (b) the Poisson distributed number of false positives that are expected due to randomness in a Markov process on the percentile state space. Finally, we compare the number of Compustat firms that meet our benchmarks to confidence intervals based on those Poisson distributions.

METHOD

We studied the population of U.S.-domiciled publicly traded firms in Compustat for the period 1965–2008. In all, our sample included firms from 431 industries at the standard industrial classification (SIC) code four-digit level. There were 243,722 firm years of observations in the analyses of ROA, and 20,131 firms were represented. In the analyses of Tobin's q , there were 216,594 firm years of observations and 19,458 firms were represented.⁴

Dependent variables

In keeping with prior studies (McGahan and Porter, 1999; Wiggins and Ruefli, 2005), we measured firm performance on an annual basis using both return on assets and Tobin's q . ROA, an accounting measure of performance, equals net income divided by total assets. Tobin's q , which is a market-based measure of performance, equals the total year-end market value of a firm's stock divided by the book value of its assets.

⁴ With two exceptions, this sample included all firms and industries in Compustat. First, if an industry has very few observations, separating industry-level and firm-level effects is problematic, so we excluded industries at the four-digit SIC level that had fewer than 20 firm-year observations. Second, we excluded firms with SIC codes in the range of 6700–6799, which involve mutual funds, mineral royalty trusts, and charitable trusts because many of these are shell corporations with no employees that serve only as tax shelters. Compustat has additional information for the years 1950–1964, but we excluded that data because it is survivor biased and omits firms that delisted prior to 1965.

We defined top 10 percent ROA and top 10 percent Tobin's q as our focal, superior performance outcomes. To assess robustness, we reran our analyses using top 20 percent ROA and top 20 percent Tobin's q . The use of these focal outcomes is somewhat arbitrary; however, we believe they are reasonable starting points. Future studies can consider other focal outcomes.

Control variables

Our interest is in whether sustained superiority exists, not why it occurs, so we used a limited set of controls designed only to insure comparability across observations. First, to account for differing macroeconomic conditions and allow meaningful comparisons across time, we coded a set of year dummies, one for each year that the sample covered. Second, because industries differ in their asset intensity, and because theories of competitive superiority generally speak to within-industry differences, we included a dummy variable for each SIC four-digit industry. A firm's industry was determined by its primary four-digit SIC code, the category in which it had its greatest sales in a focal year. That SIC code was updated annually to account for firms that changed their primary industry location.

Our results were virtually unchanged in analyses that controlled only for year, only for industry, and for neither industry nor year. Similarly, results were virtually unchanged when we included controls for firm size and market share along with those for industry and year. Size and share had modest effects on *which* firms met their respective benchmarks, but little effect on the number of such firms.

Estimation

To remove the effects of a set of controls, we can run a regression and examine the residuals, $\varepsilon_{i,t}$. For least squares regression, positive residuals indicate above average outcomes, net of the controls (Greene, 1993). We are interested, however, in performance in the top 10 or 20 percent of the distribution, and a control's effect may be quite different at those upper percentiles than at the average (Kennedy, 2008; Koenker, 2005). Furthermore, the fat-tailed and highly skewed distributions of our performance measures severely violate the assumptions of normality and homoskedasticity

necessary for least squares estimation. Waring (1996), for instance, had to drop fully 10 percent of the observations from his sample of Compustat firms to prevent outliers from violating the assumptions of his least squares regression models. That is infeasible here because superior performance is an outlier phenomenon, by definition.

We, therefore, need a regression model that (a) makes no distributional assumptions, (b) preserves the rank-order information conveyed by extreme outcomes, yet (c) is not overly influenced by such data. Semiparametric quantile regression, which predicts the value of the q^{th} percentile of the outcome variable, meets each of these requirements (Koenker, 2005; Koenker and Bassett, 1978; Koenker and Hallock, 2001). As an example, you could estimate the ninetieth quantile, which is the value of the dependent variable's conditional ninetieth percentile:

$$Q_{.90}(y|\mathbf{X}_{i,t}) = \beta_{.90}\mathbf{X}_{i,t}.$$

Here, $\beta_{.90}$ is a row vector of regression coefficients, and \mathbf{X} is a column vector of year and industry dummies. Residuals are calculated in the usual fashion: $\varepsilon_{i,t} = y_{i,t} - \beta_{.90}\mathbf{X}_{i,t}$, which removes the effects of the controls for industry and year. Positive residuals indicate values above the dependent variable's conditional ninetieth percentile; negative residuals fall below that mark.

In the analyses of top 10 percent outcomes, we estimated a dependent variable's ninetieth quantile. We estimated the eightieth quantile in top 20 percent analyses. Once we had run a quantile regression on the Compustat data, we obtained the first to ninety-ninth percentiles of the resultant residuals. We used those percentile values to find the percentile location of each firm-year observation in Compustat. We then constructed a transition matrix by calculating the percentage of times across the population that Compustat firms moved between one percentile and another between year $t-1$ and t . As noted earlier, Table 1 reports portions of the transition matrix for ROA.

We also used the percentile location data to count the number of top 10 percent and top 20 percent focal outcomes that each real firm achieved across its observed life. We then determined whether that count of focal outcomes met the simulation-based benchmarks shown in Table 2. Firms that met those benchmarks were

labeled sustained superior performers, and their total number in the Compustat population (n_{superior}) was recorded for each value of p_{firm} .

RESULTS

Figures 2 and 3 graph the distributions and provide descriptive statistics for ROA and Tobin's q . Due to their fat-tailed and skewed distributions, mean ROA is below the tenth percentile; mean Tobin's q is above the ninetieth percentile, and variances are essentially infinite. This accords with research on performance distributions (Mandelbrot, 1963, 1967; McKelvey and Andriani, 2005) and reinforces the need for analyses that are free of distributional assumptions.

Tests of the number of sustained superior performers

For top 10 percent outcomes at several values of p_{firm} , Table 3 shows (1) the expected number of false positives (μ) in a Compustat-sized population due to a Markov process on the percentile state space, which we obtained from our simulations; (2) confidence intervals around those μ 's; (3) n_{superior} , which is the number of Compustat firms that met a set of benchmarks; and (4) θ , which is the ratio of unexpected sustained superior performers in Compustat to the expected number of false positives.

As Table 3 shows, for both ROA and Tobin's q , each value of n_{superior} was greater than the upper end of its respective 99.8 percent confidence interval.

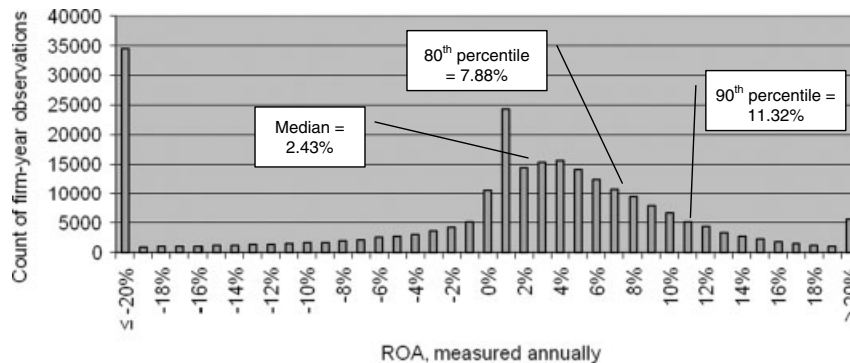


Figure 2. Frequency distribution of ROA, measured annually ($n = 243,722$ firm-years) Mean = -136.84% ; std. dev. = $27,658\%$

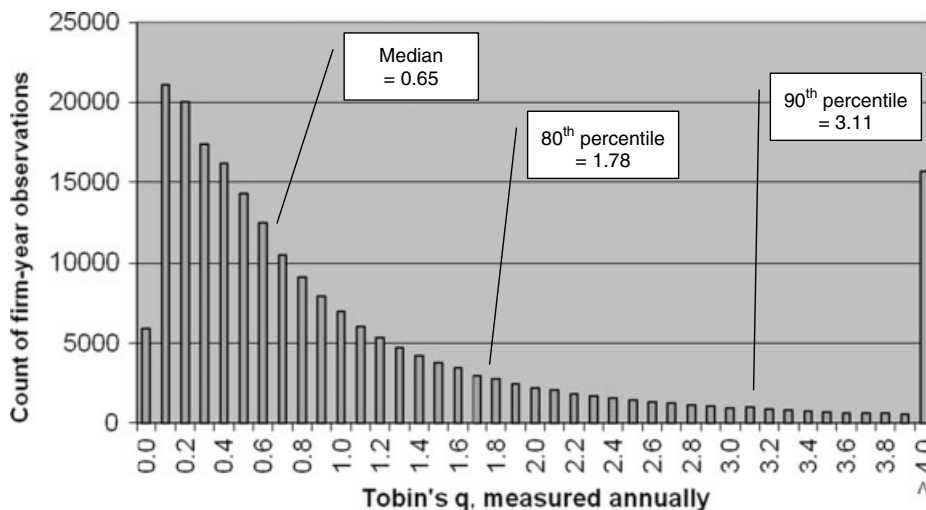


Figure 3. Frequency distribution of Tobin's q , measured annually ($n = 216,594$ firm-years) Mean = 6.27 ; std. dev. = 663.09

Table 3. Counts of sustained superior performers at several values of p_{firm} , top 10% outcomes, benchmarked with simulations using homogenous Markov chains on the percentile state space

p_{firm}	μ = Expected number of false positives due to a homogenous Markov process	99.8% confidence interval around μ	n_{superior} = Observed number of sustained superior performers among Compustat firms	θ = Ratio of unexpected sustained superior performers to expected false positives = $(n_{\text{superior}} - \mu)/\mu$
Top 10% ROA analyses				
0.05	709.98	630 to 790	912	0.28***
0.01	97.57	68 to 127	321	2.29***
0.002	14.65	3 to 26	147	9.03***
0.001	7.00	0 to 15	111	14.86***
Top 10% Tobin's q analyses				
0.05	563.94	493 to 635	861	0.53***
0.01	87.81	60 to 116	293	2.34***
0.002	13.95	3 to 25	107	6.67***
0.001	5.91	0 to 13	69	10.67***

*** $p < .001$

Overall test across all values of p_{firm} that $\theta = 0$ for ROA: $\chi^2 = 3310.76$, 4 d.f. $p < 0.001$

Overall test across all values of p_{firm} that $\theta = 0$ for Tobin's q: $\chi^2 = 1929.79$, 4 d.f. $p < 0.001$

For instance, for top 10 percent ROA and $p_{\text{firm}} < 0.001$, 15 or fewer firms were expected to meet their respective benchmarks due to randomness. In contrast, 111 Compustat firms met those standards. Thus, for top 10 percent outcomes, we can strongly reject the possibility that the number of sustained superior performers in Compustat can be fully explained by randomness in a homogenous Markov process on the percentile state space.

Next, we considered the ratio of unexpected sustained superior performers to expected false positives, which is $\theta = (n_{\text{superior}} - \mu)/\mu$. As seen in Table 3 for $p_{\text{firm}} < 0.05$ and top 10 percent ROA, $\theta = 0.28$, which we can evaluate by a chi-squared statistic: $\chi^2 = (n_{\text{superior}} - \mu)^2/\mu = 57.47$, $p < 0.001$, 1 d.f. Although this θ is significantly greater than zero, its value is small, and approximately $\mu/n_{\text{superior}} = 72$ percent of those 912 firms could be false positives. Using the top 10 percent ROA benchmarks for $p_{\text{firm}} = 0.05$ to find sustained superior performers is therefore ill-advised because the majority of firms identified have track records that would be expected somewhere in Compustat due to a homogenous Markov process of the type we consider. The situation is similar for $p_{\text{firm}} < 0.05$ and top 10 percent Tobin's q outcomes. There, $\theta = 0.53$, and $\mu/n_{\text{superior}} = 65$ percent of the 861 Compustat firms that you might identify for study could be false positives. Thus, using the benchmarks associated with $p_{\text{firm}} < 0.05$ is ill-advised for both ROA and Tobin's q.

At $p_{\text{firm}} < 0.01$, θ ratios were 2.29 for ROA and 2.34 for Tobin's q, indicating a bit more than two cases of unexpected sustained superiority for every potential false positive. If you used those benchmarks and studied a relatively large number of firms (e.g., $n > 30$), then by the central limit theorem you could be fairly certain of having at least twice as many unexpected as expected superior performers. However, many case studies sample a small number of success stories. For instance, Chandler (1962) studied four high performing organizations. Given the formula for θ , the probability of sampling a false positive among a group of firms that has met their respective benchmarks is $p_{\text{false}} = 1/(1 + \theta)$, and the number of false positives follows a binomial distribution. If we sample four success stories, then we need a θ ratio of about 7:1 to yield a 90 percent chance that at least three of those four cases are not false positives. If we relax our standards and accept a θ ratio of 4:1, then there is only an 80 percent chance that three of four cases are something other than what a Markov process on the percentile state space would produce. At $\theta = 2.34$, which is the ratio for top 10 percent Tobin's q outcomes at $p_{\text{firm}} < 0.01$, there is a 42 percent chance of having two or more false positives out of four success stories, which is a good deal higher than we should accept. This indicates that for top 10 percent ROA and Tobin's q outcomes at $p_{\text{firm}} < 0.01$, the rate of potential false positives is high enough that studies of a relatively small number of firms is ill-advised.

In comparison, the situation is more promising for more stringent values of p_{firm} , which reduce the odds of false positives. At both $p_{\text{firm}} < 0.002$ and $p_{\text{firm}} < 0.001$, for both ROA and Tobin's q , θ ratios are relatively high, ranging from 6.67 to 14.86. At those ratios, the number of sustained superior performers in Compustat is sufficiently high, relative to the number of false positives, for valid studies of even a small number of firms.

In summary, we can reject the idea that the number of sustained superior performers in Compustat is fully explained by randomness in a homogeneous Markov process on the percentile state space. In comparison, at both $p_{\text{firm}} < 0.05$ and $p_{\text{firm}} < 0.01$, we found that although the number of sustained superior performers is more than we would expect, there are too few such firms, relative to the expected number of false positives, to enable valid study. It was only at the very stringent benchmarks associated with $p_{\text{firm}} < 0.002$ and $p_{\text{firm}} < 0.001$ that the ratio of unexpected to expected sustained superior performers is high enough for valid study of a relatively small number of firms, possibly through case analysis. To insure that, a firm observed for 20 years, as an example, would need at least 13 years of top 10 percent ROA or Tobin's q outcomes.

Robustness checks using top 20 percent outcomes

To assess robustness, we reran our analyses by counting top 20 percent ROA and Tobin's q outcomes among Compustat firms and then benchmarking those counts against our simulations. We then calculated θ , which is the ratio of unexpected sustained superior performers in Compustat to the expected number of false positives. Those results, which are available by request from the authors, revealed θ ratios for top 20 percent outcomes that were similar in magnitude and significance to those in Table 3 for top 10 percent outcomes. Thus, results and implications are similar for top 10 percent and top 20 percent outcomes.

Extension: potential forms of firm heterogeneity

To now, we have modeled randomness using homogeneous Markov chains. However, because we have found many more sustained superior performers in Compustat than we would expect in such

a process, we now ask: Is there a simple form of firm-level heterogeneity that is consistent with the rate at which unexpected sustained superior performers, whose number equals $n_{\text{superior}} - \mu$, are observed in Compustat? There are many forms of heterogeneity (Denrell, 2004), but we consider two that map well onto theories of sustained advantage.

Heterogeneous Markov chains

Researchers have developed a variety of 'mover-stayer' models that introduce heterogeneity across subjects into Markov chains (e.g., Spilerman, 1972). Stayers never change states (e.g., geographic locations or jobs), so an identity matrix captures their behavior. In contrast, movers change position relatively often and have transition matrices with nonzero values off the main diagonal. Such models introduce heterogeneity by allowing different groups of subjects to move according to different transition matrices.

State changes such as geographic moves contain strong elements of individual choice. Some people can simply decide to stay put. In contrast, firms are always at some risk of changing performance, so we alter the mover-stayer metaphor to fit our needs. One possibility is that isolating mechanisms allow initial differences in firms' resource stocks to persist across time (Denrell, 2004). These mechanisms, which appear relatively early in a firm's life and then persist, include property rights and patents, gifted organizational founders, first-mover advantage, and technological lock-in (Arthur, 1990; Lippman and Rumelt, 1982; Peteraf, 1993; Rumelt, 1984). To capture this, we not only allow firms to begin their simulated lives from different starting percentiles but also have firms with different start points move according to different transition matrices.

To do this, we assigned each Compustat firm, based on its performance decile at $t = 1$, to one of 10 groups. For each group, we followed firms across their lives and constructed a separate 100×100 transition matrix. Simulated firms in the top 10 percent at $t = 1$ drew from one matrix for their entire lives, firms in the next 10 percent at $t = 1$ drew from a different matrix, and so forth.

If a simulation of heterogeneous Markov chains on percentile state spaces closely matches what occurs among real firms, it will account for all systematic variation and leave only random error. That would happen if the simulation produced

benchmarks just stringent enough so that n_{superior} , the number of sustained superiors in Compustat, was very close to μ , which is the expected number of false positives. In turn, if n_{superior} is close to μ , then $\theta = (n_{\text{superior}} - \mu)/\mu$ will be close to zero, which we test by $\chi^2 = (n_{\text{superior}} - \mu)^2/\mu$ at 1 d.f. Furthermore, because benchmarks are set independently for each value of p_{firm} , we can add those four χ^2 statistics to attained another that we test at 4 d.f. The latter, aggregate χ^2 statistic assesses the overall fit of the simulation results to what occurs in Compustat. The smaller that χ^2 value, the better the fit because more systematic variation will have been accounted for.

Table 4 reports results for top 10 percent ROA and Tobin's q . θ ratios were all positive and significant, so the number of sustained superior performers in Compustat was more than we would expect due to Markov processes on percentile state spaces with heterogeneous transition matrices. The θ s in Table 4 are about 30 percent lower than the corresponding ones in Table 3, which were generated by a single transition matrix. Heterogeneity in transition probabilities imprinted at $t = 1$ therefore has a potentially significant yet limited ability to account for the number of sustained superior performers observed in Compustat.

Semi-Markov process

Another variant of the mover-stayer model is a semi-Markov process in which all subjects move

according to the same transition matrix, but mobility rates differ across individuals (Singer and Spilerman, 1973; Spilerman, 1972). Most firms, therefore, make a random draw and potentially change their performance every year. However, other firms, once they land in a performance state, then 'dwell' there for a number of years before making another random draw that potentially changes their performance. Longer dwell times may stem from quasi-irreversible commitments such as long-term supplier contracts or specialized research and development programs (Cool & Henderson, 1998; Williamson, 1983), or from unique resources assembled to create temporary advantages (D'Aveni, 1994; Wiggins and Ruefli, 2005). If such a commitment pays off, a firm may enjoy strong returns for several years. Or things may go wrong and lead to a series of poor returns.

To this point, all simulated firms potentially change performance every period, so their dwell time equals one year. Here, however, we modified that assumption using two parameters, which allow for heterogeneous dwell times. First, pct_{dwell} = the percentage of simulated firms, randomly chosen, with dwell times that may extend beyond a year. Second, μ_{dwell} = the mean of a Poisson distribution of extended dwell times. If a simulated firm is randomly chosen, then it makes a random draw at its birth from a Poisson distribution with mean μ_{dwell} to determine how long it remains in a given performance state before moving according to the transition matrix. That extended dwell time then

Table 4. Counts of sustained superior performers at several values of p_{firm} , top 10% outcomes, benchmarked with simulations using heterogeneous Markov chains on percentile state spaces involving 10 transition matrices

p_{firm}	μ = Expected number of false positives due to a heterogeneous Markov process	99.8% confidence interval around μ	n_{superior} = Observed number of sustained superior performers in Compustat	θ = Ratio of unexpected sustained superior performers to false positives = $(n_{\text{superior}} - \mu)/\mu$
Top 10% ROA analyses				
0.05	674.46	597 to 752	788	0.17***
0.01	99.40	69 to 129	263	1.64***
0.002	14.18	3 to 25	107	6.54***
0.001	6.55	0 to 14	67	9.22***
Top 10% Tobin's q analyses				
0.05	583.56	511 to 656	797	0.36***
0.01	87.90	60 to 116	239	1.72***
0.002	12.13	2 to 23	82	5.75***
0.001	4.99	0 to 12	55	10.18***

*** $p < .001$

Overall test across all values of p_{firm} that $\theta = 0$ for ROA: $\chi^2 = 1453.40$, 4 d.f. $p < 0.001$

Overall test across all values of p_{firm} that $\theta = 0$ for Tobin's q : $\chi^2 = 1241.13$, 4 d.f. $p < 0.001$

Table 5. Counts of sustained superior performers at several values of p_{firm} , top 10% ROA outcomes, benchmarked with simulations using a semi-Markov process on the percentile state space with heterogeneous dwell times

p_{firm}	μ = Expected number of false positives due to a semi-Markov process	99.8% confidence interval around μ	n_{superior} = Observed number of sustained superior performers in Compustat	θ = Ratio of unexpected sustained superior performers to false positives = $(n_{\text{superior}} - \mu)/\mu$	χ^2	$p(\chi^2)$
0.05	697.84	619 to 777	694	-0.00	0.02	$p > 0.88$
0.01	81.98	55 to 109	95	0.16	2.07	$p > 0.15$
0.002	3.15	0 to 8	4	0.27	0.23	$p > 0.63$
0.001	1.04	0 to 4	1	-0.04	0.00	$p > 0.96$

Overall test across all values of p_{firm} that $\theta = 0$ for ROA: $\chi^2 = 2.32$, 4 d.f., $p > 0.67$

remains constant throughout that firm's simulated life. In comparison, $1 - \text{pct}_{\text{dwell}}$ of firms potentially move every year.

A semi-Markov process on the percentile state space would fit well if we found a parameter pair ($\text{pct}_{\text{dwell}}$, μ_{dwell}) that yielded an expected number of false positives per simulated population (μ) that closely matched the actual number of sustained superior performers observed in Compustat (n_{superior}). We grid searched the parameter space $0 \leq \text{pct}_{\text{dwell}} \leq 50\%$; $0 \leq \mu_{\text{dwell}} \leq 10$ years as follows. First, we selected the next pair of ($\text{pct}_{\text{dwell}}$, μ_{dwell}) values using increments of one percent for $\text{pct}_{\text{dwell}}$ and 0.1 years for μ_{dwell} . Second, using those values, we simulated the entire history of the Compustat population 200 times. Third, we used those simulation results to calculate θ ratios and the χ^2 goodness of fit statistics described earlier. Fourth, we identified the ($\text{pct}_{\text{dwell}}$, μ_{dwell}) pair that offered the best fit, which proved to be $\text{pct}_{\text{dwell}} = 16\%$; $\mu_{\text{dwell}} = 4.3$ years for top 10 percent ROA. Note that *smaller* χ^2 values and *larger* p -values indicate better fit because we are trying to find simulation parameters that produce results that differ insignificantly from what we observe in Compustat. (Given the very lengthy computing time involved, we did not assess Tobin's q or top 20% outcomes.)

Table 5 shows that heterogeneity parameters of $\text{pct}_{\text{dwell}} = 16\%$; $\mu_{\text{dwell}} = 4.3$ years produced close matches between μ , the expected number of firms meeting their benchmarks through a semi-Markov process on the percentile state space, and n_{superior} , the observed number of sustained superior performers in Compustat. This yields θ ratios that are all insignificantly different from zero, and a nonsignificant value of $p > 0.67$ for the overall test that $\theta = 0$ across all values of p_{firm} . This indicates that instances of sustained superiority

in Compustat are consistent with a semi-Markov process on the percentile state space with these heterogeneity parameters. This finding has several implications, which we address in the discussion section.

DISCUSSION

Understanding the drivers of sustained superior performance is central to both strategy research and the way that strategy is taught in most business schools (Porter, 1980, 1985). Yet as Denrell (2004: 933) notes, this emphasis on sustained superiority raises a question:

Do the most profitable firms differ systematically ex ante from less profitable firms, or is profitability mainly the result of historical accidents and chance? Addressing this question is a central challenge in strategic management research.

To our knowledge, prior studies have not benchmarked how often a firm must perform at a high level to discount randomness as a sufficient explanation for what is observed. For one type of randomness, a time-homogenous Markov process on the percentile state space, we can now offer such benchmarks. If a firm meets such a standard, randomness may still be at work, but it is unlikely to explain the firm's entire track record. Several points are worth highlighting.

First, for both ROA and Tobin's q , and for both top 10 percent and top 20 percent outcomes, there were many more sustained superior performers than we would expect through a time-homogenous Markov process on the percentile state space. This lends encouragement to theories of sustained

advantage, such as the resource-based view and research on dynamic capabilities, which seek systematic explanations for enduring success (Barney, 1991; Eisenhardt and Martin, 2000; Teece *et al.*, 1997). Forms of randomness other than those we consider may account for what we see in Compustat, so our results are merely a first step in establishing that nonrandom sustained superior performers exist in sufficient numbers to enable valid empirical study. Future research on other stochastic processes is needed to see what roles other sorts of randomness may play.

Second, our results indicate that benchmarks associated with $p_{\text{firm}} < 0.002$ are sufficiently stringent to insure that one is not observing mostly false positives, as we define them, even if one samples a relatively small number of successful firms. At that p -value, a few false positives may sneak in, but their numbers are apt to be small enough, relative to the observed number of sustained superior performers, to encourage empirical study, possibly through case analyses. To promote future research on sustained superiority, Table 6 lists all Compustat firms in our sample that meet the $p_{\text{firm}} < 0.002$ benchmarks for top 10 percent ROA.

Many of us who teach strategy cases have sometimes thought: ‘I wonder if the firm we’re talking about today—often using terms like “sustained competitive advantage”—is truly skilled or just fortunate?’ Such a distinction matters when we claim that a firm is worthy of study because its results are sufficiently impressive to justify systematic examination of its behavior. We now have benchmarks that let us assess whether a firm’s track record can be fully explained by one sort of randomness. Importantly, we believe that those benchmarks are more stringent than many authors imagine. We have subjected the firms in several bestseller success studies to our benchmarks (e.g., *Blueprint to a Billion* [Thomson, 2005] and *Big Winners and Big Losers* [Marcus, 2006], plus the books mentioned earlier). These studies attempt to infer the causes of superior performance by studying allegedly great companies, however, we found that many of the companies therein had track records *at the time of publication* that were indistinguishable from randomness, as we have defined it. We reach this conclusion using the same data as those books’ authors, not post-publication results that were unavailable to them. To the extent that these books contain useful insights, that may reflect the inherent ingenuity of their authors rather

than the validity of the case study data that they employed. We can now do better in our case selection, and we have work underway that employs requisitely demanding benchmarks.

Our third point involves two forms of firm-level heterogeneity that might explain the number of sustained superior performers in Compustat. We found a degree of support for the idea that some firms are endowed early in their lives with resources, such as a talented founder or first-mover advantage that increase their odds of long-term success. Such heterogeneity could account for about 30 percent of the unexpected sustained superior performers observed in Compustat. In comparison, a semi-Markov process on the percentile state space in which (a) most firms potentially change performance every year, while (b) about 16 percent of firms dwell at a given performance level for Poisson distributed periods averaging 4.3 years, provided a sufficient explanation for the unexpected sustained superiority that we observed. This *suggests* (and we emphasize that word because other forms of heterogeneity may fit the data equally well) that although unexpected sustained superiority is rare—we find it in about 100–300 firms among the 20,000 or so in Compustat—the factors underlying it may be shared by a much larger group of organizations, 16 percent of 20,000 being 3,200 firms. Future research is needed to see if superior performers strongly resemble a significant chunk of the population in terms of making longer-term commitments than is the norm. Good fortune may be the major factor that sets superior performers apart from other high-commitment firms, but future research is needed to assess that.

Our last point is that we have barely scratched the surface in terms of the study of what sustained superiority looks like, what sorts of firm-level heterogeneity might explain it, and what forms randomness might take. Future studies might allow the stickiness of firm performance to vary across industries, so stochastic processes could be modeled using a separate transition matrix for each sector. That approach could also be combined with a mover-stayer model to distinguish more and less inertial firms within each industry. Slight differences in within-industry transition matrices might be sufficient to account for observed outcomes. It would also be interesting to consider random walks in which stochastic changes in firm resources from $t-1$ to t were restricted to neighboring states (cf. Denrell, 2004), yet there was

Table 6. Companies meeting $p(\text{firm}) < 0.002$ for top 10% ROA

Compustat gvkey	Company name	Years	Number of top 10% ROA outcomes
12540	ADOBE SYSTEMS INC	24	18
1495	AMERICAN LIST CORP	24	15
14565	AMERICAN POWER CONVERSION CP	19	14
1602	AMGEN INC	27	19
1682	APCO ARGENTINA INC	27	22
142748	ARBITRON INC	10	10
1766	ARNOLD INDUSTRIES INC	24	18
1234	ATRION CORP	43	27
23809	AUTOZONE INC	19	15
1910	AVEMCO CORP	18	13
1966	BACARDI CORP -CL A	18	13
2008	BANDAG INC	40	22
16679	BANK OF GRANITE CORP	16	12
25338	BED BATH & BEYOND INC	18	17
2230	BIOMET INC	26	20
2259	BLACK HILLS CORP	43	19
7853	BLAIR CORP	41	22
2282	BOB EVANS FARMS	42	27
12898	BOSTON ACOUSTICS INC	20	14
2393	BRIGGS & STRATTON	43	21
62686	CARBO CERAMICS INC	14	13
29613	CASCADE BANCORP	16	14
28320	CDW CORP	15	12
2825	CEDAR POINT	16	15
3030	CHURCH'S FRIED CHICKEN INC	20	13
2293	CIRCA PHARMACEUTICALS INC	20	14
3078	CITIZENS COMMUN -SER B OLD	27	19
66503	CITYBANK	11	10
3093	CLARCOR INC	43	22
3186	COLONIAL COS INC -CL B	25	23
3188	COLONIAL PENN GROUP INC	13	11
3271	COMMUNICATIONS INDS INC	19	18
147849	COMPUTER PROGRAMS & SYSTEMS	9	9
17160	CORUS BANKSHARES INC	21	18
64162	COVANCE INC	14	11
3568	COX COMMUNICATIONS INC -OLD	19	13
20019	CVB FINANCIAL CORP	28	15
3898	DETROIT INTL BRIDGE CO	13	12
3917	DIAGNOSTIC PRODUCTS CORP	25	22
3971	DIONEX CORP	27	18
4062	DOW JONES & CO INC	41	20
4011	DR PEPPER CO-OLD	17	16
4301	ELECTROSPACE SYSTEMS INC	14	12
4321	EMERSON ELECTRIC CO	43	25
4404	ENTEX INC	17	12
63172	FACTSET RESEARCH SYSTEMS INC	14	14
4560	FAMILY DOLLAR STORES	39	19
14225	FASTENAL CO	22	19
4640	FIFTH THIRD BANCORP	29	22
16764	FINANCIAL TRUST CORP	9	9
109584	FRONTIER FINANCIAL CORP/WA	15	13
5116	GENTEX CORP	28	17
5125	GENUINE PARTS CO	43	24
5224	GOODHEART-WILLCOX CO INC	31	17
5351	GROSS TELECASTING	19	13
12338	HEALTH MANAGEMENT ASSOC	24	14

Table 6. (Continued)

Compustat gvkey	Company name	Years	Number of top 10% ROA outcomes
12840	HEARTLAND EXPRESS INC	24	23
63763	HIBBETT SPORTS INC	14	12
5680	HOME DEPOT INC	29	17
5797	HYATT INTL CORP -CL A	10	10
5878	ILLINOIS TOOL WORKS	43	20
5892	IMPELL CORP	11	11
6074	INTL DAIRY QUEEN -CL A	24	15
6297	JUNO LIGHTING INC	23	14
6375	KELLOGG CO	43	29
6379	KELLY SERVICES INC -CL A	41	20
6450	KING WORLD PRODUCTIONS INC	15	14
25283	KOHL'S CORP	18	13
14169	LANDAUER INC	22	22
6617	LAWSON PRODUCTS	39	26
6618	LAWTER INTERNATIONAL INC	33	20
12592	LIFE TECHNOLOGIES INC	16	16
6737	LINCOLN ELECTRIC HLDGS INC	35	18
14954	LINDSAY CORP	21	14
12216	LINEAR TECHNOLOGY CORP	24	15
6756	LIQUI-BOX CORP	25	18
7040	MARION MERRELL DOW INC	29	18
23592	MBNA CORP	14	14
7219	MEDCHEM PRODUCTS INC	11	10
7228	MEDTRONIC INC	43	23
7253	MERCANTILE BANKSHARES CORP	29	19
7254	MERCANTILE STORES CO INC	32	16
7257	MERCK & CO	43	26
7275	MESA LABORATORIES INC	26	16
24379	MGIC INVESTMENT CORP/WI	19	15
12141	MICROSOFT CORP	24	21
65069	MID-STATE BANCSHARES	9	9
7409	MILLIPORE CORP	41	18
7481	MOCON INC	30	18
139665	MOODY'S CORP	11	11
7551	MOORE (BENJAMIN) & CO	22	19
7637	MYLAN INC	36	18
19124	NASB FINANCIAL INC	16	12
7799	NATURES SUNSHINE PRODS INC	31	16
7863	NEW ULM TELECOM INC	33	17
7878	NEWHALL LAND & FARM -LP	37	18
7964	NORTH PITTSBURGH SYSTEMS	31	17
64028	NU SKIN ENTERPRISES -CL A	14	12
24186	OSI RESTAURANT PARTNERS INC	17	13
19318	PARK NATIONAL CORP	19	15
25880	PATTERSON COMPANIES INC	18	13
8434	PENN ENGR & MFG CORP	39	20
19159	PENNS WOODS BANCORP INC	10	10
25333	PERPETUAL FEDERAL SAVINGS BK	15	12
8589	PINKERTONS INC -CL B	14	12
8596	PIONEER HI-BRED INTERNATIONL	26	16
8633	PLANTRONICS INC	39	22
8642	PLENUM PUBLISHING CORP	30	16
8726	PREMIER INDUSTRIAL CP	29	23
13733	PSYCHEMEDICS CORP	22	15
8815	PUBLIX SUPER MARKETS INC	35	23
8126	RE CAPITAL CORP	29	15

Table 6. (Continued)

Compustat gvkey	Company name	Years	Number of top 10% ROA outcomes
65474	RENAISSANCE LEARNING INC	14	11
17168	ROYAL BANCSHARES/PA -CL A	16	14
9294	RUSSELL STOVER CANDIES INC	15	12
19570	S & T BANCORP INC	16	14
9459	SCHERING-PLOUGH	43	21
9526	SCRIPPS HOWARD BROADCASTING	26	18
9676	SHOP & GO INC	13	11
9682	SHOWBOAT INC	30	17
9699	SIGMA-ALDRICH CORP	34	30
30260	SIMPSON MANUFACTURING INC	16	13
63338	STRAYER EDUCATION INC	14	12
10225	SWISS CHALET INC	27	16
13041	SYNOVUS FINANCIAL CORP	28	17
65270	SYNTEL INC	14	12
10243	SYNTEX CORP	28	15
10326	TAMBRANDS INC	31	29
10362	TECHNALYSIS CORP	16	13
15414	TECHNE CORP	20	19
24965	TERRA NITROGEN CO -LP	18	13
10631	TOTAL SYSTEM SERVICES INC	26	25
10920	UNITED PARCEL SERVICE INC	35	17
11018	UNIVERSAL MANUFACTURING CO	30	24
63863	USANA HEALTH SCIENCES INC	14	12
11051	UTAH MEDICAL PRODUCTS INC	27	21
11861	VALLEY NATIONAL BANCORP	21	15
66599	WADDELL&REED FINL INC -CL A	13	11
17145	WASHINGTON FED INC	19	18
11234	WD-40 CO	36	28
11343	WEIS MARKETS INC	43	28
11513	WILMINGTON TRUST CORP	27	18
11535	WINN-DIXIE STORES INC	43	21
24725	WORLD ACCEPTANCE CORP/DE	19	14
1478	WYETH	43	28
12189	X-RITE INC	24	14
24405	ZEBRA TECHNOLOGIES CP -CL A	19	15

also a second stochastic process that mapped a firm's resources at time t to its performance. This would capture not only path dependence in the accumulation and depreciation of firm resources (Barney, 1991) but also instances in which firms (a) assembled substantial resource stocks yet stumbled during strategy implementation, or (b) were resource poor yet benefitted from fortunate breaks. Finally, in this study we have not penalized firms for poor performances, but future work might do so. This may shed further light on the 'Bowman paradox' in which firms with rare competencies perform well on a consistent basis with little downside risk (Andersen, Denrell and Bettis, 2007; Bowman, 1980).

As this research suggests, there is much to do in the study of sustained superior performance. More generally, from what we have seen of the performance numbers among U.S. publicly traded companies, we call for much greater use of robust analyses that are free of distributional assumptions because firm performance has fat-tailed, skewed distributions that violate the assumptions of the Gaussian models that pervade the strategy field. Techniques such as Markov chain modeling and quantile regression will not only give us better analyses but also—and more importantly—enable us to study what occurs at the upper end of the performance distribution, where our theories of competitive advantage make their home.

ACKNOWLEDGEMENTS

We are indebted to Tamara Fossey, James Guszczka, Bruce Rudy, Lige Shao, Jeff Schulz, and James Wappler for research assistance, and to Craig Crossland, Don Hambrick, George Huber, Francisco Polidoro, Gerry Sanders, and Robert Wiggins for their comments and suggestions on an earlier draft. We also thank Editor Rich Bettis and two anonymous reviewers for their invaluable guidance.

REFERENCES

Andersen TJ, Denrell J, Bettis RA. 2007. Strategic responsiveness and Bowman's risk-return paradox. *Strategic Management Journal* **28**(4): 407–429.

Arthur WB. 1990. Positive feedbacks in the economy. *Scientific American* **262**: 92–99.

Barney JB. 1986. Strategic factor markets: expectations, luck, and business strategy. *Management Science* **32**: 1231–1241.

Barney J. 1991. Firm resources and sustained competitive advantage. *Journal of Management*, **17**: 99–120.

Barney J. 1997. On flipping coins and making technology choices: luck as an explanation of technological foresight and oversight. In *Technological Innovation: Oversights and Foresights*, Garud R, Nayyar PR, Shapira ZB (eds). Cambridge University Press: Cambridge, U.K.; 13–19.

Bowman EH. 1980. A risk/return paradox for strategic management. *Sloan Management Review* **21**: 17–31.

Brown R. 1828. A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Journal* **4**: 161–173.

Chandler AD. 1962. *Strategy and Structure*. MIT Press: Cambridge, MA.

Collins J. 2001. *Good to Great*. HarperCollins: New York.

Collins J, Porras JI. 1994. *Built to Last*. HarperCollins: New York.

Cool K, Henderson J. 1998. Power and firm profitability in supply chains: French manufacturing industry in 1993. *Strategic Management Journal* **19**(10): 909–926.

Daley DJ, Vere-Jones D. 2003. *An Introduction to the Theory of Point Processes* (2nd edn). Springer: New York.

D'Aveni RA. 1994. *Hypercompetition: Managing the Dynamics of Strategic Maneuvering*. Free Press: New York.

Denrell J. 2004. Random walks and sustained competitive advantage. *Management Science* **50**: 922–934.

Efron B. 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* **68**: 589–599.

Efron B. 2000. The bootstrap and modern statistics. *Journal of the American Statistical Association* **95**: 1293–1296.

Eisenhardt KM, Martin JA. 2000. Dynamic capabilities: what are they? *Strategic Management Journal*, October–November Special Issue **21**: 1105–1121.

Feller W. 1968. *An Introduction to Probability Theory and its Applications* (3rd edn). Wiley: New York.

Greene WH. 1993. *Econometric Analysis* (2nd edn). Macmillan: New York.

Hiriji KF. 2005. *Exact Analysis of Discrete Data*. Chapman & Hall: Boca Raton, FL.

Joyce W, Nohria N, Roberson B. 2003. *What Really Works*. HarperCollins: New York.

Kennedy P. 2008. *A Guide to Econometrics* (6th edn). Wiley-Blackwell: New York.

Koenker R. 2005. *Quantile Regression*. Cambridge University Press: New York.

Koenker R, Bassett G. 1978. Regression quantiles. *Econometrica* **46**: 33–50.

Koenker R, Hallock KR. 2001. Quantile regression. *Journal of Economic Perspectives* **15**: 143–156.

Levinthal DA. 1990. Organizational adaptation, environmental selection and random walks. In *Organizational Evolution: New Directions*, Singh J (ed). Sage: Newbury Park, CA; 201–223.

Levinthal DA. 1991. Random walks and organizational mortality. *Administrative Science Quarterly* **36**: 397–420.

Levinthal DA. 1997. Adaptation on rugged landscapes. *Management Science* **43**: 934–950.

Lippman SA, Rumelt RP. 1982. Uncertain imitability: an analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* **13**: 418–438.

Loftus EG. 1979. The malleability of human memory. *American Scientist* **67**: 312–320.

Mandelbrot B. 1963. The variation of certain speculative prices. *Journal of Business* **36**: 394–419.

Mandelbrot B. 1967. The variation of some other speculative prices. *Journal of Business* **40**: 393–413.

Marcus AA. 2006. *Big Winners and Big Losers: The 4 Secrets of Long-Term Business Success and Failure*. Prentice Hall: Upper Saddle River, NJ.

McGahan AM, Porter ME. 1999. The persistence of shocks to profitability. *Review of Economics and Statistics* **81**: 143–152.

McKelvey B, Andriani P. 2005. Why Gaussian statistics are mostly wrong for strategic organization. *Strategic Organization* **3**: 219–228.

Peteraf MA. 1993. The cornerstones of competitive advantage: a resource-based view. *Strategic Management Journal* **14**(3): 179–191.

Peters TJ, Waterman RH. 1982. *In Search of Excellence: Lessons from America's Best-Run Companies*. HarperCollins: New York.

Porter ME. 1980. *Competitive Strategy*. Free Press: New York.

Porter ME. 1985. *Competitive Advantage*. Free Press: New York.

Powell TC. 2001. Competitive advantage: logical and philosophical considerations. *Strategic Management Journal* **22**(9): 875–888.

- Raynor ME. 2007. *The Strategy Paradox: Why Committing to Success Leads to Failure (and What to Do About It)*. Currency/Doubleday: New York.
- Resnick SI. 1992. *Adventures in Stochastic Processes*. Birkhauser: Boston, MA.
- Rivkin JW. 2000. Imitation of complex strategies. *Management Science* **46**: 824–844.
- Rosenzweig P. 2007. *The Halo Effect*. Simon & Schuster: New York.
- Ruefli TW, Wiggins RR. 2000. Longitudinal performance stratification: an iterative Kolmogorov-Smirnov approach. *Management Science* **46**: 685–692.
- Rumelt RP. 1984. Towards a strategic theory of the firm. In *Competitive Strategic Management*, Lamb RB (ed). Prentice-Hall: Englewood Cliffs, NJ; 556–570.
- Singer B, Spilerman S. 1973. Social mobility models for heterogeneous populations. *Sociological Methodology* **5**: 356–401.
- Spilerman S. 1972. The analysis of mobility processes by the introduction of independent variables into a Markov chain. *American Sociological Review* **37**: 277–294.
- Taleb NN. 2005. *Fooled by Randomness* (2nd edn). Random House: New York.
- Teece DJ, Pisano G, Shuen A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* **18**(7): 509–553.
- Thomson DG. 2005. *Blueprint to a Billion: 7 Essentials to Achieve Exponential Growth*. Wiley: Hoboken, NJ.
- Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* **185**: 1124–1131.
- Waring GF. 1996. Industry differences in the persistence of firm-specific returns. *American Economic Review* **86**: 1253–1265.
- Wiener N. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. MIT Press: Cambridge, MA.
- Wiggins RR, Ruefli TW. 2002. Sustained competitive advantage: temporal dynamics and the incidence and persistence of superior economic performance. *Organization Science* **13**: 82–105.
- Wiggins RR, Ruefli TW. 2005. Schumpeter's ghost: is hypercompetition making the best of times shorter? *Strategic Management Journal* **26**(10): 887–911.
- Williamson OE. 1983. Credible commitments: using hostages to support exchange. *American Economic Review* **73**: 519–540.