Great Companies: The Secrets of Success are Still a Secret

Gabrielle Baum  
Department of Economics  
Pomona College  
Claremont CA 91711

Gary Smith  
Department of Economics  
Pomona College  
Claremont CA 91711

contact author: Gary Smith, phone: 909-607-3135; fax: 909-621-8576; gsmith@pomona.edu
Great Companies: The Secrets of Success are Still a Secret

Abstract

Many management books ransack a list of thriving companies for common characteristics. The problem is that any group of companies (bad, good, or great) will inevitably share some common characteristics. Finding such traits only confirms that we looked, and tells us nothing about whether these found characteristics are responsible for past successes or are reliable predictors of future success. Nor does it tell us whether these companies’ stocks will be profitable investments.

We illustrate these arguments by considering the companies identified in the best-selling books *Good to Great* and *In Search of Excellence*. 
**Great Companies: The Secrets of Success are Still a Secret**

Why are some companies more successful than others? Why do some companies grow and prosper while others languish and fail? Why are some great while others are good, mediocre, or bad? These questions are repeatedly asked by business executives, analysts and consultants, but the answers remain elusive.

In this paper, we contrast the companies identified as successful by the best-selling books *Good to Great* (Collins, 2001) and *In Search of Excellence* (Peters and Waterman, 1982) with the procedure used by *Fortune* magazine’s most-admired list. First, we look at their selection criteria. Then we look at how well these companies’ stocks did after the publication of these books.

The history of corporate America has witnessed the rise and fall of many companies and even entire industries. In the early 1990s, having a dot-com at the end of a company’s name was assumed to guarantee success, and stock prices soared to levels that turned out to be completely unjustified. A dot-com looked liked the formula for success, but was not.

Stock prices are often pumped up by dreams of reaping riches from being part of the latest fad, the next new thing, the next IBM, Microsoft, or Google. Are there any objective criteria for predicting which companies will be successful and which stocks will be profitable investments? There are three major pitfalls: data mining, the efficient market hypothesis, and regression toward the mean.

**Data Mining**

Statistical tests assume that the researcher starts with a theory, collects data to test the theory, and reports the results—whether statistically significant or not.
Many people work in the other direction, scrutinizing the data until they find a pattern and then making up a theory that fits the pattern. Ransacking data for patterns is fun and exciting—like playing Sudoku or solving a murder mystery. Examine the data from every angle. Look for something—anything—that is interesting. After a pattern is discovered, start thinking about reasons.

This pillaging is known as *data mining* (aka data grubbing, data dredging, fishing expeditions). The problem is that even random coin flips contain patterns that appear to be meaningful, but are, in fact, meaningless. When a fair coin is flipped ten times, a streak of four heads in a row seems too remarkable to be explained by chance, even though streaks this long, or longer, can be expected 47 percent of the time. We think something must be unusual about the coin or the person flipping the coin.

This is one version of the Texas sharpshooter fallacy, where a person with no talent fires shots randomly at the side of a barn and, afterward, paints a bullseye around a cluster of bullet holes. Even randomly generated data usually contain clusters and, if we look for an explanation for a data cluster, we inevitably will find one. In a cancer study, we might discover that several cancer victims happened to live near power lines, a Little League field, or a water tower.

Data mining demonstrates little more than a researcher’s endurance. We cannot tell whether a data mining marathon demonstrates the validity of a useful theory or the perseverance of a determined researcher. Data without theory is treacherous, and we should be deeply skeptical.

**The Efficient Market Hypothesis**

In an efficient market, all available information is taken into account by investors and is therefore fully reflected in market prices. If it is well known that a company is great, then its
stock will trade at a price that gives investors an appropriate anticipated return, taking into account risk and other characteristics that are relevant for their investment decisions.

There is, however, a difference between possessing and processing information. Warren Buffett did not beat the market for decades by having access to information that was not available to other investment professionals, but by thinking more clearly about information available to everyone.

The question here is whether best-selling books like *Good to Great* and *In Search of Excellence* look at companies differently and uncover strengths that are not appreciated fully by security analysts and other investors. Do these successful companies turn out to be great investments?

**Regression to the Mean**

A statistics professor named Horace Secrest wrote a book he titled, *The Triumph of Mediocrity in Business* (1933). In every industry that he examined, the most successful companies tended to become less successful over time, while the least successful tended to become more successful. Secrist concluded that all companies would soon be mediocre. The president of the American Statistical Association wrote an enthusiastic review, but both he and Secrist were fooled by regression toward the mean.

To understand regression, suppose that 100 people are asked 20 questions about the stock market. Each person’s “ability” is his or her average score on an infinite number of such tests. Some people have an ability of 90, some 80, and some near zero.

Someone with an ability of, say, 80 will not score 80 on every test. So, what can we infer from a person’s score on one test? A person who scores in the 90th percentile on a test could be
someone of more modest ability who did unusually well, or could be someone of higher ability who did poorly. The former is more likely because there are more people below the 90th percentile in ability than above it. If this person’s ability is, in fact, below, the 90th percentile, then when he or she takes another test, the score will probably also be below the 90th percentile. Similarly, a person who scores far below average is likely to have had an off day and will probably score somewhat higher on later tests. This tendency of people who score far from the mean to score closer to the mean on a second test is an example of regression toward the mean.

Regression occurs in many contexts. School children who are given special tutoring because of their low test scores can be expected to do better on subsequent tests even if the tutor does nothing more than snap his fingers and say, “Improve!” Patients who are given treatment because of a worrisome medical test result can be expected to improve even if the treatment is worthless. In the 1800s, Sir Francis Galton observed that unusually tall parents tend to have somewhat shorter children, while the reverse is true of unusually short parents. The erroneous conclusion is that heights are regressing to mediocrity—a conclusion encouraged by Galton titling his study “Regression Towards Mediocrity in Hereditary Stature.”

Regression toward the mean does not imply that everyone will soon be the same height any more than it implies that everyone will soon get the same score on tests. What regression does imply is that when an unobserved trait is measured imperfectly, measurements that are far from the mean overstate how far the underlying trait is from the mean.

This framework is directly applicable to any imperfect measure of a company’s success. For example, earnings growth that is high relative to a group of companies is also likely to be high relative to that company’s “ability.” We can consequently anticipate the company’s earnings
growth to regress toward the mean. Freeman and Tse (1992) and Fama and French (2000) found such earnings regression, although they attributed it to competitive forces rather than the purely statistical explanation that the companies with relatively high earnings are more likely to have experienced good luck than bad.

There is well-established evidence that regression to the mean is a pervasive but subtle statistical principle that is often misunderstood or insufficiently appreciated (Kahneman and Tversky 1973). In the stock market, Keynes (1936) observed that “day-to-day fluctuations in the profits of existing investments, which are obviously of an ephemeral and nonsignificant character, tend to have an altogether excessive, and even absurd, influence on the market.”

Lakonishok, Shliefer, and Vishny (1994) found that companies that have doing well as gauged by such metrics as earnings growth rates have turned out to be relatively poor investments compared to companies doing poorly by such metrics. La Porta (1996) found that companies that analysts predict to have high earnings growth rates have underperformed the stocks of companies that analysts are pessimistic about.

Smith, Keil, and Smith (2004) similarly found that earnings forecasts are systematically too extreme—too optimistic for companies predicted to do well and too pessimistic for those predicted to do poorly. The accuracy of these forecasts can be improved consistently and substantially by shrinking them toward the mean forecast. They also found that a portfolio of pessimistic company stocks systematically outperformed an optimistic portfolio.

The same is true of virtually any measure of a company’s success. Companies that have been the most successful are more likely to have had good luck than bad and will consequently regress to the mean.
Fortune’s Most-Admired List

Since 1983, Fortune has been releasing an annual list of “America’s Most Admired Companies.” The differences between this list and the lists appearing in many management books is instructive.

The Fortune list is based on a survey of thousands of business executives, directors, and analysts who rate the largest companies in their industry on eight key attributes that Fortune’s editors believe are crucial for success: innovation, people management, use of assets, social responsibility, management quality, financial soundness, long-term investment, and product quality. In recent years, a ninth attribute has been added to the list: global competitiveness.

One virtue of the Fortune list is that it is based on a set of ex ante criteria. The same measures of successful companies are used year after year. Instead of looking for traits that are common to admired companies, it looks for companies that have these predefined traits.

Antunovich, Laster, and Mitnick (2000) found that, over the period 1983 through 1995, the stocks of the companies in the top decile of the Fortune ratings did better than the stocks in the bottom decile. Using data for 1983 through 2004, Anderson and Smith (2006) found that a portfolio consisting of the ten most-admired stocks outperformed the market, whether the stocks were purchased on the publication date, or 5, 10, 15, or 20 trading days later. This is a clear challenge to the efficient market hypothesis since the Fortune list is readily available public information. Anderson and Smith conclude that,

We have no compelling explanation for this anomaly. Perhaps Philip Fisher was right: the way to beat the market is to focus on scuttlebutt—those intangibles that don’t show up in a
company’s balance sheets—and *Fortune*’s most-admired survey is the ultimate scuttlebutt.

Using data for 1983 through 2007, Anginer and Statman (2010) confirmed that a portfolio consisting of the ten companies with the highest *Fortune* ratings did well relative to other admired stocks and relative to the ten least admired stocks but they also found that, overall, the stock performance was better for the companies in the bottom half of the ratings than for companies in the top half—which is consistent with regression to the mean. The superior performance of the top 10 companies is an anomaly.

**Good to Great**

In 2001 Jim Collins published his best-selling management book, *Good to Great: Why Some Companies Make the Leap... And Others Don’t*. It has sold more than 4 million copies and appeared on several lists of the best management books of all time.

Collins wrote that his book is about “our search for timeless, universal answers that can be applied by any organization” His conclusion? “We believe that almost any organization can substantially improve its stature and performance, perhaps even become great, if it conscientiously applies the framework of ideas we’ve uncovered.” (p. 5)

Collins and his research team spent five years looking at the 40-year history of 1,435 companies and identified eleven stocks that outperformed the overall market and were still improving 15 years after they made the leap from good to great:

- Abbott Laboratories
- Circuit City
- Fannie Mae
- Gillette
- Kimberly-Clark
- Kroger
- Nucor
- Pitney Bowes
- Walgreens
- Wells Fargo
- Philip Morris

Collins scrutinized these eleven great companies and identified several common themes that
he gave catchy names:

1. Level 5 Leadership: Leaders who are personally humble, but professionally driven to make a company great.

2. First Who, Then What: Hiring the right people is more important than having a good business plan.

3. Confront the Brutal Facts: Good decisions take into account all the facts.

4. Hedgehog Concept: It is better to be a master of one trade than a jack of all trades.

5. Build Your Company’s Vision: Adapt operating practices and strategies, but do not abandon the company’s core values.

These characteristics are plausible and have memorable names. The problem is that this is a backward-looking study that is undermined by data mining. Collins wrote that we developed all of the concepts in this book by making empirical deductions directly from the data. We did not begin this project with a theory to test or prove. We sought to build a theory from the ground up, derived directly from the evidence (p. 10).

Collins evidently thought this statement made his study sound unbiased and professional. He didn’t just make this stuff up. He went wherever the data took him.

In reality, Collins was admitting that he had no idea why some companies do better than others. And he was revealing that he was blissfully unaware of the perils of deriving theories from data. When we look back in time at any group of companies, the best or the worst, we can always find some common characteristics. Every one of those eleven companies selected by Collins has either an i or an r in its name, and several have both an i and an r. Is the key for going from good to great to make sure that your company’s name has an i or r in it? Of course
Finding an \( i \) and \( r \) pattern is an obvious example of data mining. So is examining a sequence of coin flips and finding that two heads happened to be followed a tail more than 50 percent of the time. Collins’ data mining is less obvious, because his found theory sounds plausible. It is nonetheless data mining because, as he freely admits, Collins made up his theory after looking at the data.

To buttress the statistical legitimacy of his theory, Collins talked to two professors at the University of Colorado. One said that, “the probabilities that the concepts in your framework appear by random chance are essentially zero.” (p. 212) The other professor was more specific. He asked, “What is the probability of finding by chance a group of 11 companies, all of whose members display the primary traits you discovered while the direct comparisons do not possess those traits?” He calculated this probability to be less than 1 in 17 million. Collins concludes, “There is virtually no chance that we simply found 11 random events that just happened to show the good-to-great pattern we were looking for. We can conclude with confidence that the traits we found are strongly associated with transformations from good to great.” (p. 212)

It is not clear how this probability of 1 in 17 million was calculated. (I contacted the professor and he couldn’t remember.) What is clear is that it is incorrect.

In statistics, this kind of reasoning is sometimes called the Feynman Trap, a reference to the Nobel Laureate Richard Feynman. Feynman asked his Cal Tech students to calculate the probability that, if he walked outside the classroom, the first car in the parking lot would have a specific license plate, say 8NSR26. Cal Tech students are very smart and they quickly calculated a probability by assuming each number and letter were independently determined. This answer is
less than 1 in 17 million. When they finished, Feynman revealed that the correct probability was 1 because he had seen this license plate on his way to class. Something extremely unlikely is not unlikely at all if it has already happened.

The calculations made by the Colorado professors and the Cal Tech students assume that the five traits and the license plate number were specified before looking at the company data or the cars in the parking lot. They were not, and the calculations are irrelevant.

Collins does not provide any evidence that the five characteristics he describes were responsible for these companies’ success. To do that, he would have had to provide a theoretical justification for these characteristics, select companies beforehand that did and did not have these characteristics, and monitor their success according to some metric established beforehand. He did none of this.

After the publication of Good to Great, success evaporated for some of these companies. Fannie Mae stock went from above $80 a share in 2001 to less than $1 a share in 2008 and delisting in 2010. Circuit City went bankrupt in 2009.

**In Search of Excellence**

Twenty years earlier, another best-selling business book did something very similar and had exactly the same problems. Two McKinsey consultants, Tom Peters and Robert Waterman, were asked to study several successful companies. They talked to other McKinsey consultants and came up with a list of 62 leading companies.

In order to make their analysis appear scientific, they looked at six measures of long-term success, three related to growth and three measuring returns on capital and assets. In order to stay in the sample, a company had to rank in the top half of its industry for four of the six measures
during the period of 1961 through 1980. As a final screen, they asked industry experts to rate the companies’ twenty-year record of innovation. The final 43 firms included the 35 publicly traded companies shown in Table 1 and eight companies that were privately held or subsidiaries of other companies (Allen-Bradley, Atari, Bechtel, Chesebrough-Pond’s, Frito-Lay, Hughes Aircraft, Mars, and Tupperware).

Table 1 The 35 Publicly Traded Companies Identified in *In Search of Excellence*

<table>
<thead>
<tr>
<th>3M</th>
<th>Disney Productions</th>
<th>Marriott</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amdahl</td>
<td>Dow Chemical</td>
<td>Maytag</td>
</tr>
<tr>
<td>Amoco</td>
<td>Du Pont</td>
<td>McDonald’s</td>
</tr>
<tr>
<td>Avon</td>
<td>Eastman Kodak</td>
<td>Merck</td>
</tr>
<tr>
<td>Boeing</td>
<td>Emerson Electric</td>
<td>National Semiconductor</td>
</tr>
<tr>
<td>Bristol-Myers</td>
<td>Fluor</td>
<td>Proctor &amp; Gamble</td>
</tr>
<tr>
<td>Caterpillar Tractor</td>
<td>Hewlett-Packard</td>
<td>Raychem</td>
</tr>
<tr>
<td>Chesebrough-Pond’s</td>
<td>IBM</td>
<td>Revlon</td>
</tr>
<tr>
<td>Dana Corporation</td>
<td>Intel</td>
<td>Schlumberger</td>
</tr>
<tr>
<td>Data General</td>
<td>Johnson &amp; Johnson</td>
<td>Texas Instruments</td>
</tr>
<tr>
<td>Delta Airlines</td>
<td>K Mart</td>
<td>Walmart</td>
</tr>
<tr>
<td>Digital Equipment</td>
<td>Levi Strauss</td>
<td>Wang Labs</td>
</tr>
</tbody>
</table>
Peters and Waterman then spoke to managers and read magazine stories about these companies. They identified eight common traits; for example, a bias for action and being close to the consumer. The book they wrote about their efforts, *In Search of Excellence* (1982), was again a backward-looking study undermined by data mining. There is no way of knowing whether companies with a “bias for action,” whatever that means, were more successful than other companies, or whether companies that had been excellent in the past would be excellent in the future.

Clayman [1987, 1994] conducted two studies of the stock returns of companies labeled as excellent by Peters and Waterman. In the first study, Clayman looked at 29 of the 36 publicly traded companies over the period 1981 through 1985. Eighteen underperformed the stock market, but some of the 11 companies that outperformed the market (including Walmart, Maytag, and McDonald’s) did very well, so that an equally weighted portfolio beat the market by 1.1 percent per year. In the second study, Clayman (1994) concluded that, for the years 1988 through 1992, the investment performance of the 29 excellent companies as a whole was indistinguishable from that of “unexcellent” companies.

Her explanation is an incorrect interpretation of regression to the mean:

Over time, company results have a tendency to regress to the mean as underlying economic forces attract new entrants to attractive markets and encourage participants to leave low-return businesses. Because of this tendency, companies that have been “good” performers in the past may prove to be inferior investments, while “poor” companies frequently provide superior investment returns in the future. (Clayman, 1987, p. 63)

Regression to the mean does not assume that “economic forces tend to move things towards
equilibrium.” (Clayman, 1987, p. 62) Economic forces may exist, but regression to the mean is a purely statistical phenomenon that occurs when observed data measure unobserved latent traits imperfectly; for example, using earnings growth rates to measure a company’s greatness.

**Methods**

We looked at the performance of the eleven *Good to Great* stocks (“Great”) and the 35 publicly traded *In Search of Excellence* stocks (“Excellent”) from the first trading day in January following the publication of each book through December 31, 2012, using the total daily returns in the CRSP database. We also looked at a portfolio consisting of an equal investment in each Great stock and a similar portfolio for the Excellent stocks. If a company was acquired or went private, the proceeds were evenly invested in the remaining stocks in the portfolio. For example, after Gillette merged with Proctor & Gamble on October 1, 2005, the proceeds were invested equally in the ten stocks remaining in the Great portfolio.

The returns for the individual stocks and the two portfolios were compared to the NYSE/AMEX/NASDAQ market index portfolio constructed by CRSP and used in the Fama-French factor model.

We used a matched-pair t-test to gauge the statistical significance of the observed daily differences between the returns on these stocks and the market index. The null hypothesis is that the expected value of the daily difference is zero: $H_0: \mu = 0$. The t-statistic is

$$ t = \frac{\bar{X} - 0}{s / \sqrt{n}} $$

where $\bar{X}$ is the mean of the daily differences, $s$ is the standard deviation of the daily differences, and $n$ is the number of trading days. We report the two-sided p-value because we cannot *a priori*
rule out the possibility that these stocks will do better or worse than the market.

We also looked at the risk-adjusted returns for the individual stocks and for each portfolio using the capital asset pricing model (CAPM),

\[ R = \alpha + \beta \text{MKT} + \varepsilon \]

where

- \( R \) = return on an individual stock or portfolio minus return on Treasury bills
- \( \text{MKT} \) = market index return minus return on Treasury bills

In theory, \( \alpha \) should be zero, so finding a substantial, statistically significant value for \( \alpha \) indicates an excess risk-adjusted return.

CAPM is appealing because it is based on a theoretical model of asset prices. On the other hand, the model’s assumptions are fairly restrictive and unrealistic. Fama and French (1993), among others, have found empirical regularities in the distribution of stock returns that cannot be explained by CAPM, in particular a small-stock effect and a value effect.

The Fama-French three-factor model, augmented by a momentum factor, is

\[ R = \alpha + \beta_1 \text{MKT} + \beta_2 \text{SMB} + \beta_3 \text{HML} + \beta_4 \text{UMD} + \varepsilon \]

where

- \( R \) = return on an individual stock or portfolio minus return on Treasury bills
- \( \text{MKT} \) = market index return minus return on Treasury bills
- \( \text{SMB} \) = average return on three small portfolios minus the average return on three big portfolios (size factor)
- \( \text{HML} \) = the average return on two value portfolios minus the average return on two growth portfolios (book-to-market factor)
UMD = average return on two high prior return portfolios minus the average return on two low prior return portfolios (momentum factor)

The Fama-French three-factor model accounts for the fact that, in general, small stocks tend to outperform big stocks (Banz 1981; Reinganum 1981) and value stocks with high book-to-market ratios tend to outperform growth stocks (Rosenberg, Reid and Lanstein, 1985). A momentum factor is included because of empirical evidence that stocks that have been doing well tend to outperform those doing poorly (Jegadeesh and Titman 1993).

It is unsettled whether these factors reflect risks that matter to investors (Chan 1988; Fama and French, 1992) or are evidence of market inefficiencies (Lakonishok, Shliefer, and Vishny, 1994). Either way, the question here is whether the performance of the Great and Excellent stocks can be explained by these four factors. As with CAPM, a substantial, statistically significant value for \( \alpha \) indicates an unexplained excess return.

The data for estimating CAPM and the Fama-French model are from CRSP and Ken French’s web site (2013).

Results

Table 2 shows the annualized returns for each Great stock compared to the index. (The index returns are not the same for each stock because the holding periods differed due to mergers, acquisitions, buyouts, and bankruptcies.) Five of the Great stocks did better than the market index, six did worse. The only stock with a matched-pair p value less than 0.05 is Circuit City, which underperformed the market (in fact, went bankrupt in 2008). The Great portfolio did worse than the market index but the observed difference is not close to be statistically persuasive (p = 0.707).
Table 2 Annualized Returns (%) and Estimated CAPM and Fama-French Alphas,
Good to Great Stocks, January 2, 2002 - December 31, 2012

<table>
<thead>
<tr>
<th>Stock</th>
<th>Annualized Returns</th>
<th>CAPM</th>
<th>Fama-French</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stock</td>
<td>Index</td>
<td>P value</td>
</tr>
<tr>
<td>Abbott Laboratories</td>
<td>4.88</td>
<td>5.23</td>
<td>0.991</td>
</tr>
<tr>
<td>Circuit City</td>
<td>-100.00</td>
<td>-0.03</td>
<td>0.037</td>
</tr>
<tr>
<td>Fannie Mae</td>
<td>-100.00</td>
<td>2.80</td>
<td>0.360</td>
</tr>
<tr>
<td>Gillette</td>
<td>17.89</td>
<td>6.03</td>
<td>0.306</td>
</tr>
<tr>
<td>Kimberly-Clark</td>
<td>6.76</td>
<td>5.23</td>
<td>0.880</td>
</tr>
<tr>
<td>Kroger</td>
<td>2.99</td>
<td>5.23</td>
<td>0.909</td>
</tr>
<tr>
<td>Nucor</td>
<td>14.62</td>
<td>5.23</td>
<td>0.107</td>
</tr>
<tr>
<td>Philip Morris</td>
<td>17.09</td>
<td>5.23</td>
<td>0.103</td>
</tr>
<tr>
<td>Pitney Bowes</td>
<td>-6.43</td>
<td>5.23</td>
<td>0.075</td>
</tr>
<tr>
<td>Walgreens</td>
<td>2.08</td>
<td>5.23</td>
<td>0.815</td>
</tr>
<tr>
<td>Wells Fargo</td>
<td>7.04</td>
<td>5.23</td>
<td>0.360</td>
</tr>
<tr>
<td>Portfolio</td>
<td>5.20</td>
<td>5.23</td>
<td>0.707</td>
</tr>
</tbody>
</table>

For CAPM and the Fama-French factor model, the only alphas with p values less than 0.05 were for Philip Morris (with a positive alpha) and Circuit City (with a negative alpha). The Great portfolio has small positive alphas, but both p values are 0.58.

Table 3 shows the results for the Excellent stocks. Fifteen of the Excellent stocks did better than the market, 20 did worse. The only stock with a matched-pair p value less than 0.05 is Dana, which underperformed the market and went bankrupt in 2006. The Excellent portfolio did somewhat worse than the market, but the p value is 0.986.
In Search of Excellence Stocks and Portfolio, January 3, 1983 - December 31, 2012

<table>
<thead>
<tr>
<th>Stock</th>
<th>Annualized Returns</th>
<th>CAPM</th>
<th>Fama-French</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Index</td>
<td>P value</td>
<td>Alpha</td>
</tr>
<tr>
<td>3M</td>
<td>11.16</td>
<td>10.49</td>
<td>0.575</td>
</tr>
<tr>
<td>Amdahl</td>
<td>4.21</td>
<td>16.27</td>
<td>0.987</td>
</tr>
<tr>
<td>Amoco</td>
<td>16.80</td>
<td>16.44</td>
<td>0.680</td>
</tr>
<tr>
<td>Avon</td>
<td>8.86</td>
<td>10.49</td>
<td>0.641</td>
</tr>
<tr>
<td>Boeing</td>
<td>11.70</td>
<td>10.49</td>
<td>0.371</td>
</tr>
<tr>
<td>Bristol-Myers</td>
<td>11.18</td>
<td>10.49</td>
<td>0.500</td>
</tr>
<tr>
<td>Caterpillar Tractor</td>
<td>12.37</td>
<td>10.49</td>
<td>0.248</td>
</tr>
<tr>
<td>Dana Corporation</td>
<td>-100.00</td>
<td>12.40</td>
<td>0.044</td>
</tr>
<tr>
<td>Data General</td>
<td>0.67</td>
<td>16.13</td>
<td>0.972</td>
</tr>
<tr>
<td>Delta Airlines</td>
<td>-100.00</td>
<td>12.10</td>
<td>0.024</td>
</tr>
<tr>
<td>Digital Equipment</td>
<td>0.84</td>
<td>16.38</td>
<td>0.386</td>
</tr>
<tr>
<td>Disney Productions</td>
<td>13.88</td>
<td>10.49</td>
<td>0.151</td>
</tr>
<tr>
<td>Dow Chemical</td>
<td>10.13</td>
<td>10.49</td>
<td>0.479</td>
</tr>
<tr>
<td>Du Pont</td>
<td>10.83</td>
<td>10.49</td>
<td>0.496</td>
</tr>
<tr>
<td>Eastman Kodak</td>
<td>-100.00</td>
<td>10.47</td>
<td>0.083</td>
</tr>
<tr>
<td>Emerson Electric</td>
<td>11.17</td>
<td>10.49</td>
<td>0.466</td>
</tr>
<tr>
<td>Fluor</td>
<td>8.58</td>
<td>10.49</td>
<td>0.470</td>
</tr>
<tr>
<td>Hewlett-Packard</td>
<td>5.66</td>
<td>10.49</td>
<td>0.816</td>
</tr>
<tr>
<td>IBM</td>
<td>9.39</td>
<td>10.49</td>
<td>0.747</td>
</tr>
<tr>
<td>Intel</td>
<td>14.75</td>
<td>10.49</td>
<td>0.171</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>13.31</td>
<td>10.49</td>
<td>0.328</td>
</tr>
<tr>
<td>K Mart</td>
<td>-100.00</td>
<td>11.65</td>
<td>0.112</td>
</tr>
<tr>
<td>Levi Strauss</td>
<td>15.69</td>
<td>15.72</td>
<td>0.819</td>
</tr>
<tr>
<td>Marriott</td>
<td>11.05</td>
<td>14.64</td>
<td>0.877</td>
</tr>
<tr>
<td>Maytag</td>
<td>7.17</td>
<td>12.45</td>
<td>0.916</td>
</tr>
<tr>
<td>McDonald’s</td>
<td>14.62</td>
<td>10.49</td>
<td>0.198</td>
</tr>
<tr>
<td>Merck</td>
<td>13.36</td>
<td>10.49</td>
<td>0.269</td>
</tr>
<tr>
<td>National Semi</td>
<td>7.18</td>
<td>10.02</td>
<td>0.255</td>
</tr>
<tr>
<td>Procter &amp; Gamble</td>
<td>12.99</td>
<td>10.49</td>
<td>0.361</td>
</tr>
<tr>
<td>Raychem</td>
<td>8.54</td>
<td>16.17</td>
<td>0.840</td>
</tr>
<tr>
<td>Revlon</td>
<td>30.04</td>
<td>17.52</td>
<td>0.354</td>
</tr>
<tr>
<td>Schlumberger</td>
<td>8.41</td>
<td>10.49</td>
<td>0.603</td>
</tr>
<tr>
<td>Texas Instruments</td>
<td>9.53</td>
<td>10.49</td>
<td>0.303</td>
</tr>
<tr>
<td>Walmart</td>
<td>17.17</td>
<td>10.49</td>
<td>0.175</td>
</tr>
<tr>
<td>Wang Labs</td>
<td>-100.00</td>
<td>14.55</td>
<td>0.453</td>
</tr>
<tr>
<td>Portfolio</td>
<td>10.29</td>
<td>10.49</td>
<td>0.986</td>
</tr>
</tbody>
</table>
For CAPM, the only alphas with p values less than 0.05 were Delta Airlines and Dana Corporation (both negative alpha). For the Fama-French model, the only alphas with p values less than 0.05 were Walmart and Intel (both positive) and Delta and Dana (both negative). The p values for the Excellent portfolio alphas are 0.194 and 0.121.

Conclusion

The secrets for building a successful company are still a secret, which is unsurprising. If the characteristics enumerated in *Good to Great*, *In Search of Excellence*, and similar books really worked, every company would be great.

The real lesson from the enduring popularity of such advice is that the authors who write these books and the millions of people who buy them do not realize that the books are fundamentally flawed. This problem plagues the entire genre of books on formulas/secrets/recipes for a successful business, a lasting marriage, living to be 100, so on, and so forth, that are based on backward-looking studies of successful businesses, marriages, and lives.

If we believe that “a bias for action” predicts success, a valid way to test this theory would be to identify companies that have a bias for action and companies that do not, and then see which companies do better over, say, the next ten years. The same is true of secrets for a successful marriage and a long life. Otherwise, we are just staring at the past instead of predicting the future.

Looking backward, we will always find commonalities among already great companies. The interesting question is which characteristics will lead to future success, and you can seldom see the future by looking backward.

The stock performance of the Great and Excellent companies after their identification in
these best-selling books has been decidedly mediocre at best. These successful companies have not been successful investments.

The contrast with the Most-Admired stocks is striking. The Most-Admired stocks did better than the market, evidently because the *Fortune* survey captured intangible factors—scuttlebutt—that investors had not taken into account fully. Information about which companies possessed such traits was useful because the traits were identified in advance, before the companies were selected.

For the Great and Excellent companies, the traits were not helpful predictors of stock market performance, either because these companies’ past successes were well known or the traits were useless because they were identified after the companies were selected.
References


Niendorf, B., and K. Beck, 2008, Good to great, or just good ?, Academy of Management Perspectives 22 (4), 13-20

