Numbers Games—

Reflections on the Uses and Abuses of "Big Data"

Keynote Address

John C. Bogle, Founder of The Vanguard Group

at the 2017 Seminar of

The Journal of Investment Management[1]

Boston, Massachusetts

September 24, 2017

In its third quarter 2015 issue, the *Journal of Investment Management* (JOIM) published a paper by investment manager Craig William French, called "The Value of Active Investing." Mr. French's intent was to demonstrate that the value of well-constructed active strategies can be worth substantially more than their cost.

His comparison of active investing versus a passive benchmark[2] covered the 225-year period 1789-2014. Over this long span, French assumed that the active strategy that he selected (WBI Dividend Income[3]) would have provided an annual return of 9.38% before costs. The passive equity portfolio provided a return of 7.70%. According to Mr. French, that active advantage of 1.68% per year eclipsed, by a substantial amount, the investment costs of equity mutual funds, 0.67%, as calculated by Dartmouth's Kenneth French in a 2008 paper.

There is no mistaking Craig French's absolute confidence in active management strategies. His short paper concludes with a full-page chart showing the cumulative returns of active investing versus passive investing over that 225-year span. Before citing his astonishing data, I pause in amazement . . . **(Exhibit 1)**

- $1 invested in the active stock strategy in 1789 grew to $572,197,734 in June 2014.

---

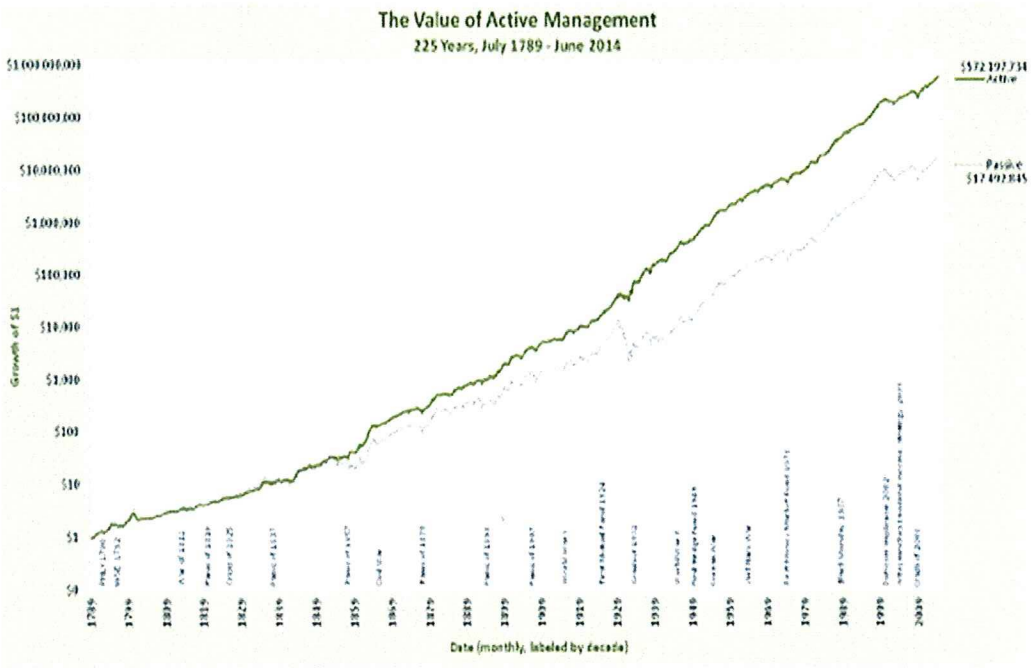
[1] In particular, I appreciate the opportunity to address this audience, in part, because my invitation came from *JOIM* editor Gifford Fong. We first met 33 years ago, in 1984, when I spoke to the Q Group. My speech was titled "Statistics and Suicide." Its theme, "Excessive reliance on data at the expense of judgment [is taking] us down a dangerous path." This address may be thought of as a sequel.

[2] The benchmark data were based on research by Professor William Schwert, University of Rochester, for 1789-1871; the Cowles Commission for 1871-1928; and the S&P 500 for 1928-2014.

[3] Craig French is president of WBI.

- $1 invested in the passive stock portfolio grew to $17,492,845.
- So there we have it: active management created $554,705,879 of additional wealth, 34 times as much as the passive portfolio. It's as easy as that!

Exhibit 1: Active Management versus Passive Investing, 1789-2015 (French, 2015)



The Active Strategy Data – Pie in the Sky?

I can hardly know where to begin to rebut this conclusion. So I start with the phrase that I scrawled across that full page chart when I first saw it: PIE IN THE SKY!

First, in ascending order, I believe that the Kenneth French cost data sharply underestimates the cost of actively managed equity mutual funds, which I estimated to be 2.27% annually in my 2014 paper for the *Financial Analysts Journal*, "The Arithmetic of 'All-In' Investment Expenses."[4] Assuming the accuracy of such "all-in" costs for actively managed funds, the Mr. French's conclusion would be turned upside down—active annual *net* return 7.11%, passive 7.77% return. That $554 million *surplus* for the active strategy is magically transformed into a $12.4 million *deficit.*

[4] *Financial Analysts Journal*, January/February 2014

Second, the implication that there is in fact a known active strategy that could possibly endure for 225 years (let alone a manager who could live that long!) strikes me as absurd. Active management strategies come and go, and so do fund managers.

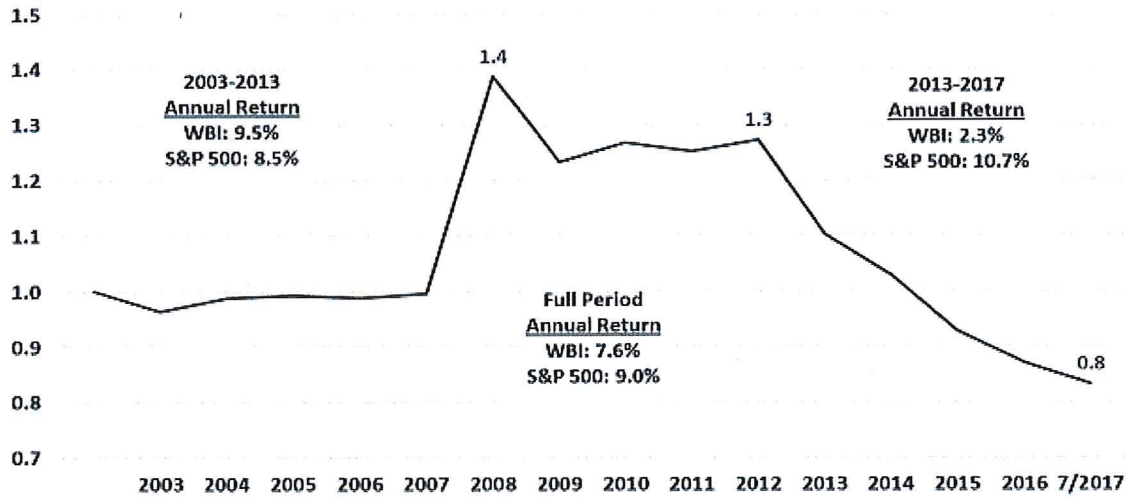
Third, and by far most important, the statistical basis for Craig French's conclusion seems pathetically weak. It rests solely on the "gross performance composite" of a single fund, the WBI Dividend Income Strategy, from its inception on July 1, 2003, through December 2013. *One strategy . . . applied over but a single decade . . . extrapolated to cover 225 years.*

A serious paper with these three flaws provides a sorry defense of "active strategies!"[5]

What Happened Next?

Now let's take a look at the WBI return data, before and after the publication of Mr. French's article.[6] During the period July 2003—December 2013, the WBI Tactical Dividend Income strategy provided an annual return of 9.5% versus 8.5% for the S&P 500.[7] A solid achievement. But for only a single decade.

Exhibit 2: WBI Tactical Dividend Income / S&P 500, 2003-2017



What's happened since then? From December 2013 to July 2017, the annual return on the WBI Dividend Strategy came to 2.3%, a shadow of the S&P 500's return of 10.7% (**Exhibit 2**). For the full 14-year

[5] Mr. French is well aware of my criticism, and unpersuaded by my analysis. Our spirited exchange of letters to the Editor of JOIM was published in the first quarter issue, 2016

[6] The year-by-year record was not presented in the Craig French paper, but is available from Morningstar.

[7] I am unable to explain why this gap of 1.00% does not match the gap of 1.68% cited in the French paper.

period, WBI Strategic Income annual return, 7.6%; cumulative return 182%. S&P 500 Index, 9.0% annually, 239% cumulative. Whatever the case, the mutual fund using the WBI Tactical Dividend Strategy ceased operations in 2017 and the fund was liquidated. Looking forward a mere four years obliterated the pie-in-the-sky of two-plus centuries.

Look, I realize that it is a stretch to combine this anecdotal example of a single strategy with the Big Data (of its day) covering two centuries of U.S. stock returns. But it illustrates my deeply held conviction: *Beware of all data*. Consider their fragility. Remember that the past is rarely prologue. Do your own homework. Take nothing for granted. *Nothing.* And go on "high alert" when a single decade's results are extrapolated to cover two-plus centuries.

Broadening the Data Horizon

This extreme example begins my keynote this evening, "Numbers Games." For I want to express my far broader concerns, not about Big Data itself, but the use of Big Data in today's world, as it permeates professional investing and the academic canon. Yes, the parallel development of Big Data and the availability of cheap computing power has been a blessing beyond imagination for quantitative managers, professors of finance, and financial engineers. These brilliant quants are now able to test every imaginable hypothesis, over an infinite number of periods and sub-periods. Their goal: to discover the Holy Grail of investing—the ultimate algorithm that produces consistently superior portfolio returns.

If the rise of Big Data has been a blessing for these quants. What does it mean to individual investors trying to cut through the fog of investing to achieve their financial goals? While the answer to that question remains to be seen, the early evidence is not encouraging. However, we can thank Big Data for showing us that there have been these two major strategies that have clearly offered an enduring performance edge: (1) small-cap stocks minus big-cap stocks (SMB) and, (2) value stocks (with high book-value-to-market-value ratios) minus growth stocks with low ratios (HML). Eugene Fama and Kenneth French have been preaching this gospel since the early 1990s, and in 2013 Professor Fama won the Nobel Prize in Economic Sciences for his work.[8]

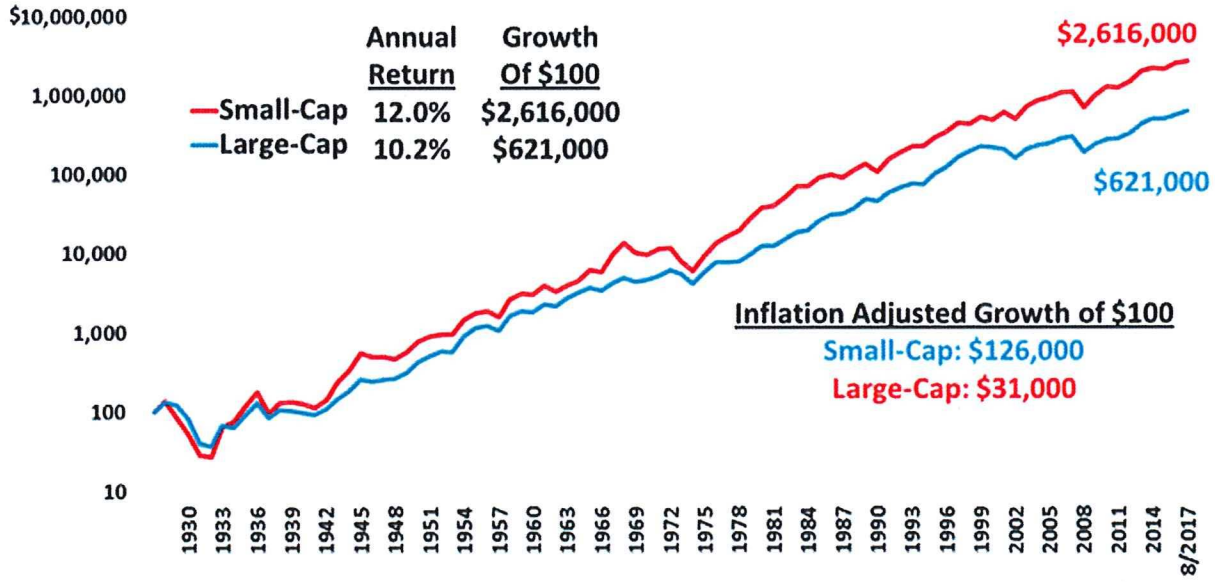
[8] Fama/French defines value stocks and those with high book-to-market ratios.

Small-Cap Stocks versus Large-Cap Stocks

Let's begin with Small versus Large. Few investment principles have been as unchallenged as the now-perennial assertion that over the long-run, Small-cap stocks have outperformed Large-cap stocks. Since 1928, according to the University of Chicago Center for Research in Securities Prices (CRSP), Small has provided an annual return of 12.0%, vs. 10.2% for Large, an annual edge of 1.8 percentage points. Professors Fama and French provided the original analysis on which that assertion rests.

Over that 90-year period, long-term compounding works its magic. Each $100 invested in Small stocks grew to $2,616,000, while each $100 in Large stocks grew to just $621,000—a difference hardly to be sneezed at. **(Exhibit 3)** But if these cumulative returns are adjusted for inflation, Small drops (amazingly!) to $126,000, Large to $31,000. At least as important, bear in mind that Small carried a higher risk (standard deviation of 29% vs. 21%). This difference in risk is significant, and should not be ignored. Result: Sharpe ratio[9] of Small-cap stocks 0.28, Large-cap stocks 0.31.

Exhibit 3: Growth of $100—Small-Cap versus Large-Cap, 1928-2017

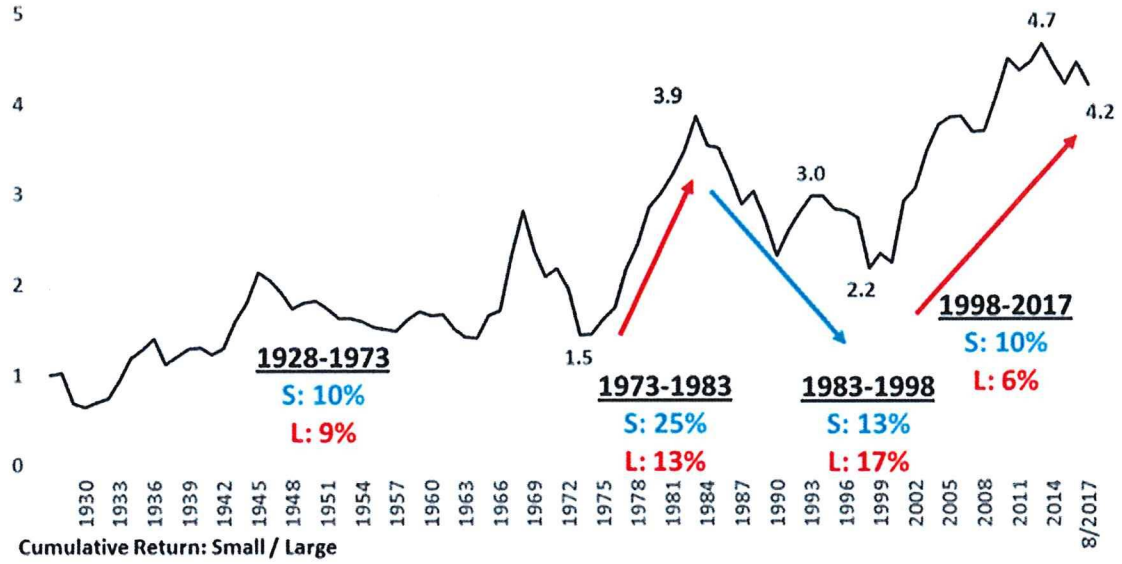


The imposing gap between the cumulative returns conceals more than it reveals. I would strongly urge you not to accept uncritically the conclusion that "Small beats Large" until you consider a chart that I devised decades ago, in which we simply divide the *cumulative* annual returns of one data series into another, year after year. In this case, we divide the year-by-year cumulative returns of Large-cap stocks into the cumulative returns

[9] The Sharpe ratio is now essentially the accepted measure of risk-adjusted return.

of Small-cap stocks. **(Exhibit 4)** Now we see that the relative return over that long period was hardly linear. (Few, if any, of such year-by-year comparisons are.) It was punctuated by a series of ups and downs—great for Small, then great for Large, and so on.

Exhibit 4: Small Stocks versus Large Stocks, 1928-2017



In the hundreds of articles I've read in the investment journals about performance-seeking algorithms, I've never seen such a chart presented. But its message is clear: from 1928 through 1973, there was little to choose from between the annual returns of the two (Small 10%, Large 9%). Then Small wins big through 1983 (25% vs. 13%). Then Large through 1998 (17% vs. 13%). From then on, Small wins again (10% vs. 6%) through mid-2017.

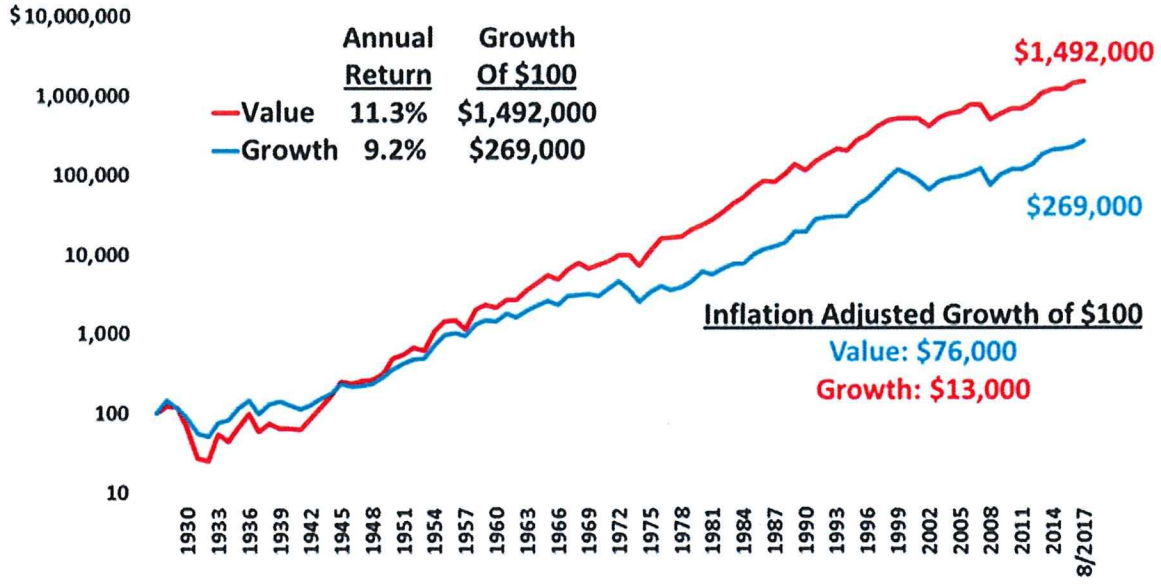
Since 1983, on balance, these to-and-fro reversals have pretty much cancelled each other out. During the past 34-years, the returns of Small and Large have been virtually *identical*--11% annually. So ask yourself whether the evidence that justifies the claim of the enduring superiority of small isn't too fragile a foundation on which to base a long-term strategy for the years ahead.

Value Stocks vs. Growth Stocks

Let's now turn from the Small/Large thesis to the Value/Growth thesis, where we'll see some interesting parallels. Here, the long-term difference between these two factors is even *greater.* **(Exhibit 5)** The annual returns since 1928 have been 11.3% for Value stocks and 9.2% for Growth stocks, a difference of fully 2.1 percentage points. The compounding of those annual returns again results in a stunning difference in cumulative

return—each initial $100 grew to $1,492,000 in Value stocks and to only $269,000 in Growth stocks. (In real dollars, *$76,000* and *$13,000*. Wow!—again.)

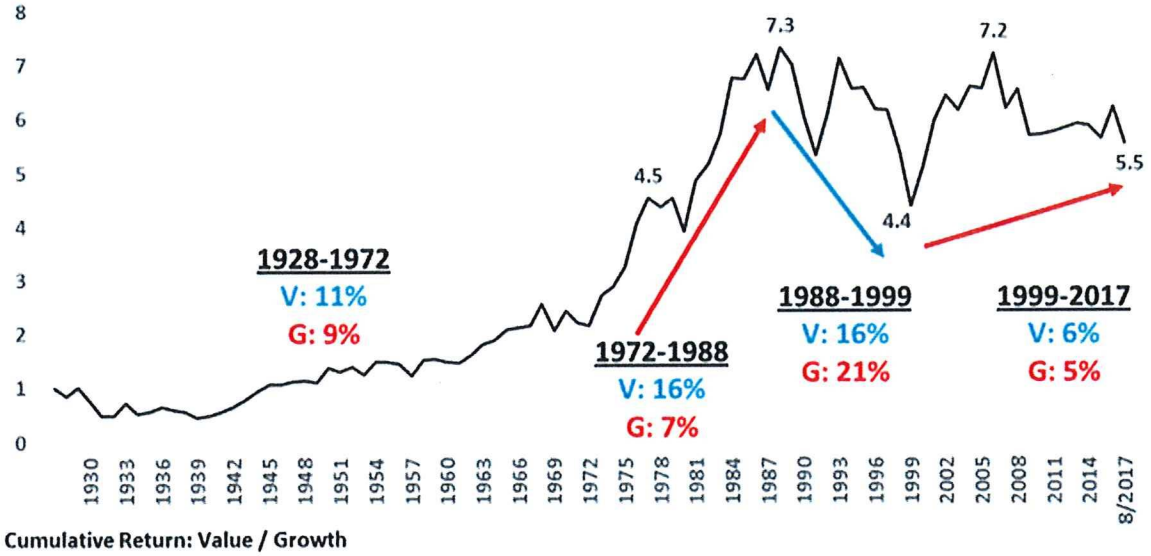
Exhibit 5: Growth of $100—Value versus Growth, 1928-2017



But wait a minute. Let's turn again to our cumulative return format and carefully examine the record of Value and Growth. **(Exhibit 6)** During the first 50 years (!) through 1977, yes, the annual Value advantage was an impressive 2 (Value 11%, Growth 9%). In the next 15 years through 1988, it surges (Value 16%, Growth 7%). Then it reverses. During the next decade-plus, through 1999, Growth 21% annually, Value 16%. Then Value surges through 2006, quickly followed by a surge in Growth. On balance, since 1999, the returns were almost equal—Value 6%, Growth 5%.

Looking to the full period since 1928, higher risk (standard deviation) of Value was 26%, vs. 20% for Growth. So extra risk may account for some of the excess returns earned by Value. Taking that risk proxy into account, the Sharpe ratio 0.30 for Value exceeds only slightly the 0.28 Sharpe ratio for Growth. But despite those extra risks and the remarkable cycles of relative returns between Value and Growth, the historical data showing the powerful relative cumulative return on Value stocks over the full 90-year period is so impressive that one is tempted to say: *Case closed*! I distance myself from that group. *The case is never closed*.

Exhibit 6: Growth versus Value, 1928-2017



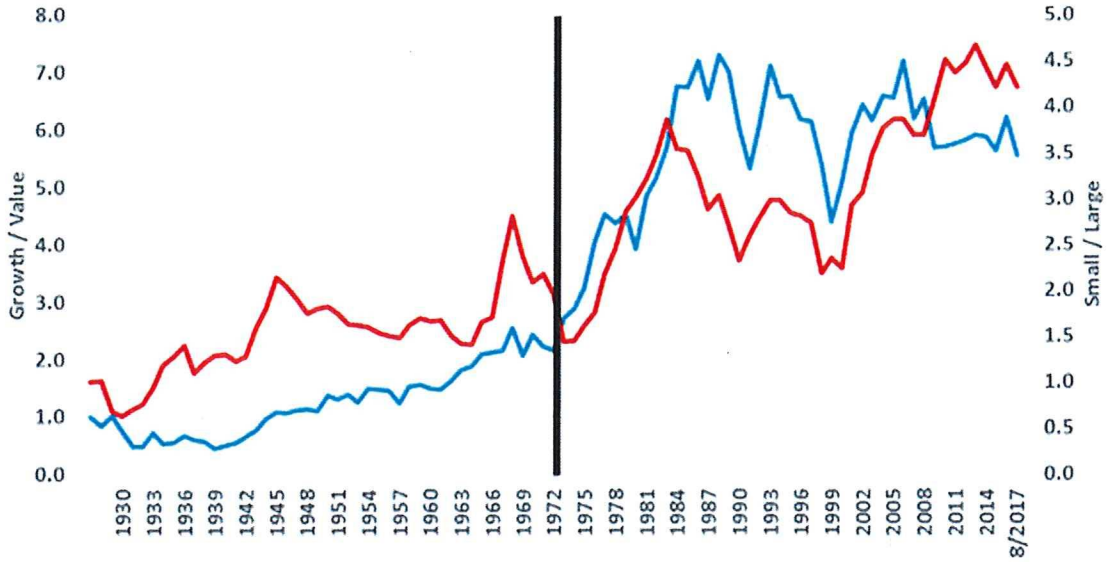
Small/Large and Value/Growth—Some Parallels

Perhaps you've already noticed some significant parallels between the Small/Large and the Value/Growth return patterns. **(Exhibit 7)** Both the Small edge and the Value edge—if you'll give me a little sea room here—reflect a sort of slow and steady accumulation of advantage during roughly the first 45 years, 1928-1972. (The left half of the chart.) Then, during the *next* 45 years, each reflected large swings from advantage to disadvantage and back again. (The right half of the chart.) There were six major swings for Small/Large and six for Value/Growth.

We can't be sure what factors are responsible for this change from the slow but steady increments of the first 45 years to a pattern of reversals during the 45 years that followed. (Truth told, I noticed it only when I was preparing these remarks.) But if recent history is more relevant than early history (as I believe it generally is), we might rethink the sanctity of Small over Large and Value over Growth.

Why? Maybe because the market is now driven by a growing cadre of institutional investors. Maybe the rise in the market's focus on short-term relative returns. Maybe the proliferation of Big Data. Indeed, perhaps there's a bit of the Heisenberg principle here and the sophisticated monitoring of a once steady differential following the Fama-French papers on those trends in the early 1990s. Of course I can't prove any of this, but it seems obvious from the patterns of relative returns of the different types of stocks suggest that the rules of the game have changed.

Exhibit 7: Small / Large and Value / Growth, 1928-2017



If market participants, driven by these relatively new concepts of SMB and HML, behave differently in the face of such a different environment, wouldn't it be likely that we would observe more frequent valuation changes? I can only wonder, so I rely on you quantitative acolytes to help me out. (But not tonight!) Inspired by Aesop, I believe that the shift from the slow but steady increases in the relative returns of Small and Value stocks (the tortoises) through the early 1970s, followed since then by the periodic and often rapid reversals of superiority for both factors (the hares), are worthy of future study.

The World of Reality

Now let's go from the calculated data of the past to the actual data in the world of reality. We need to ask ourselves about the extent to which the findings of the CRSP data—in effect, indexes of market sectors—can be replicated in the real world of investing. Investing costs money, and it is a truism—and increasingly a trite one—that all of the investors in the stock market (and in any discrete market sector) *earn* the market return *before* the costs of financial intermediation, but actually *receive* the return *after* those costs. (It cannot surprise you that I'd mention this subject!)

If the implementation costs of a strategy focused on a particular market factor exceeds significantly the rock-bottom cost of a broad stock market strategy (for example, a total stock market index fund or an S&P 500 index fund), any pre-cost victory for Small over Large and Value over Growth may be pyrrhic.

By 1985, each of these four mutual fund categories had a sufficient population to draw reasonable conclusions. So now let's compare actual mutual fund returns to the essentially cost-free returns presented by the CRSP data.[10] **(Exhibit 8)** While the patterns in each of the comparisons are reasonably parallel, the dimensions are less uniform.

Exhibit 8: Annual Returns, 1985-2017

	Large Cap	Small Cap	Value	Growth Funds	Total/ Average
CRSP	11.1%	11.7%	10.9%	11.5%	--
Mutual Funds*	9.7	11.0	9.7	9.7	--
Difference	1.4%	0.7%	1.2%	1.8%	1.3%
Number of Funds (1985)*	82	10	45	10	147

***Source: Morningstar**

Note: During the 30-year periods 1945-1975 and 1985-2015, the S&P 500 outpaced the average Large-cap fund by 1.6% per year.

During this period, Large-cap mutual fund returns lagged Large stocks by 1.4% per year annually, and Small-cap funds lagged Small stocks by 0.7%. As it happens, during 1985-2017, the CRSP data reflect a small margin in favor of Growth over Value. The average annual return of the Growth mutual funds (9.7%) during this long period trailed CRSP Growth (11.5%), a 1.8% annual shortfall. The Value mutual funds' return of 9.7% all fell below that of CRSP Value return of 10.9% by 1.2%.

These data, fragile though they may be, are discretionally consistent with my thesis that implementation costs matter. The average annual differential comes to 1.3%, close to the 1.6% gap that I calculated for Large-cap mutual funds and the S&P 500 Index during each of the 30-year periods 1945-1975 and 1985-2015 in my 2016 paper for the *Financial Analysts Journal*. This gap is a pretty good approximation of the costs that mutual funds report. So investors should not ignore the obvious but undisclosed costs of implementing a strategy that arises, pristinely, out of academic studies that cannot be precisely replicated in the real world.

In any event, place me squarely in the camp of the contrarians who are reluctant to accept the inevitable superiority of Small-cap strategies over Large-cap strategies and Value strategies over Growth strategies in the

[10] "The Index Mutual Funds: 40 Years of Growth, Challenge, and Change." *Financial Analysts Journal*. January/February 2016.

years ahead. I've been excoriated for my views, but I'm comforted by this reported exchange between Dr. Fama and a participant at an investment conference in the early 2000s: "What do you say to otherwise intelligent people (sic) like Jack Bogle who examine this same data and conclude that there is no size or value premium?" His response: "How far are they from the slide? If I get far enough away, *I don't see it either.* . . . Whether you decide to tilt towards value depends on whether you are willing to bear *the associated risk.* . . . The market portfolio is always efficient. . . . For most people, the market portfolio is the most sensible decision."

RTM - Equity Fund Returns

In the mutual fund field, the Biggest Data of all is past performance. Yet one need not dig very deep into fund performance data to observe that past performance is not prologue. *Reversion to the mean* (RTM) is king. Its powerful magnetism draws winning funds downward to the market's mean return—and beyond—and pulls losing funds upward toward the mean—and beyond. Past fund returns—good, bad, and indifferent alike—have clearly demonstrated a random probability of repeating themselves in the future.

Consider **Exhibit 9,** comparing the returns of all actively managed U.S. equity funds over the decade 2006-2016. We divided that ten-year period into two five-year periods. For the first five-year period (2006-2011), we sorted the returns into quintiles—the top quintile contained the funds with the best performance, and the bottom quintile contained those with the worst performance. We then looked at how the funds fared during the subsequent five-year period, 2011-2016.

Exhibit 9: Reversion to the Mean 2006-2016
First Five Years vs. Following Five Years

2006-2011 Ranking			2011-2016 Ranking				
	Number of Funds		Highest Return	High	Medium	Low	Lowest Return
Highest Return	353	20%	14%	14%	14%	27%	30%
High	352	20	21	17	17	24	21
Medium	353	20	20	21	22	17	20
Low	352	20	20	23	26	21	10
Lowest Return	352	20	24	25	22	14	16
Total	1762	100%	20%	20%	20%	21%	20%

Note-Total number of funds merged or liquidated: 313

If it were easy to select funds that would outperform their peers simply by investing in yesterday's winners, we would expect to see persistence; that is, most funds that ended the first period at the top of the heap would remain there in the next period and those at the bottom would remain there. But no. As it turns out, RTM overpowers persistence.

Consider the funds that ranked in the top quintile during the first period (2006-2011). Over the subsequent five years, only 14% of those funds remained in the top quintile and 14% remained in the second. A remarkable 30% of the winners from the first period ended up in the bottom quintile, and another 27% landed in the next-to-last (fourth) quintile.

At the other end of the spectrum, 24% of the first-period laggards in the bottom quintile ended up in the top quintile in the subsequent period and 25% in the second quintile. Only 16% of the losers in the first five years repeated their dismal performance in the second five years, and only 14% turned up in the fourth quintile. As the Good Book says, "the last shall be first and the first shall be last."

You need not be a statistical wizard to observe the remarkable randomness of returns through each of the three middle quintiles, with steady RTMs centering around 20%. But only 28% of the funds in the top two quintiles remained there in the subsequent period, and 57% tumbled all the way to the bottom two. In the bottom two quintiles during the first period, only 30% remained there and 49% rose to the top two quintiles.[11] Come to think of it, what we see is not pure RTM, but RTM and BTM (Below the Mean). Should contrarians rely on this pattern and bet on losers rather than winners? I wouldn't recommend it.

Data Mining

Of course Big Data has opened the door to almost infinite information about the financial markets and the returns earned by various market factors over an almost infinite variety of time periods, easily divided into an almost infinite variety of sub-periods. (We've gone far beyond the Small/Large and Value/Growth factors I've discussed in my remarks.) The search for that Holy Grail of consistently superior returns--over short-term spans and long-term spans alike--continues to flourish in our academic literature.

[11] You might be wondering if this pattern was just a one-time event, not likely to be repeated. I had the same question. So I looked at the preceding non-overlapping five-year period, 2001–2006, and compared it to 2006–2011. The pattern held. Of the top-quintile winners during the first period, only 15% remained there in the subsequent periods, while 20% fell to the bottom. Fully, 13% of the funds—45 funds—failed to survive and were liquidated.

But in this search, I fear, data mining rears its ugly head. Let's define data mining as poring over back-tested returns on markets, on factors, and stocks to arrive at conclusions about future returns. If only the past were prologue. (Of course it cannot be, and imagine what the world of money management would look like if it were.)

There are numerous problems in the construction of these winning portfolios as I have shown earlier. First, the past is rarely prologue. It is also non-linear, but volatile. Second, correlation is often confused with causation. Third, the costs of implementing these presumably winning strategies are often ignored. As if portfolio transaction costs, operating cost, advisor fees, cash drag, and taxes can be ignored!

Fischer Black, it is said, would often appraise papers that would "torture the data until they confess," and simply respond, "DM." Data Mining. It is pervasive, and (as far as I can tell) the published papers are rarely reapprised after, say, five or ten years. Did it work, or didn't it? *Tell us.* Nor do I know of any examples where the researchers who author these (for the want of a better phrase) Holy Grail articles actually put up their own money to test them, and then report on their success or failure. *Show us.*

Further, experienced professional money managers often disagree with one another on the viability of a given factor. A good example is the spirited debate between Rob Arnott of Research Affiliates and Cliff Asness of AQR Capital Management over value investing. It has reached legendary proportions. Cliff is properly concerned about the incentives to mine history using past data: "Given the rewards to gathering assets, often made easier with a good 'backtest,' the incentive to data mine is great."

In "My Factor Philippic," Cliff Asness delivers 30 full pages that discuss when value works and when it doesn't. (It includes 87 footnotes, usually of an edgy nature. I love Cliff's feistiness almost as much as his brilliance.) His conclusion: "Going forward, I do not know whether value or the non-value factors will do better."[12] Based on his comments that I cited earlier in these remarks, neither, apparently, does Dr. Fama.

M.I.T. professor Andrew Lo joins the skeptics: "The more you search over the past, the more likely it is you are going to find exotic patterns that you happen to like or focus on. Those patterns are least likely to repeat."[13]

---

[12] I've slightly revised Cliff's sentence. But I think that I've captured his meaning.
[13] "Investors Always Think They're Getting Ripped Off. Here's Why They're Right," by Peter Coy, Bloomberg, April 6, 2017.

Then Professor Campbell Harvey of Duke University piles on:

"At least 316 factors have been tested to explain the cross-section of expected returns. . . . Given the plethora of factors, and the inevitable data mining, many of the historically discovered factors would be deemed 'significant' by chance. . . . We believe there are three reasons for tougher criteria [on data mining] today. First, the low-hanging fruit has already been picked. That is, the rate of discovering a true factor has likely decreased. Second, there is a limited amount of data. Indeed, there is only so much you can do with the CRSP database. . . . Third, the cost of data mining has dramatically decreased. In the past, data collection and estimation were time intensive, so it was more likely that only factors with the highest priors—potentially based on economic first principles—were tried."[14]

Perhaps the *coup de grace* on data mining comes from the Notices of the American Mathematical Society, which referred to backtest overfitting as "pseudo-mathematics and financial charlatanism."[15] I rest my case.

Big Data or Little Data?

As academics and investment professionals, our job is to enlighten public investors as to the optimal strategy for investment success over their lifetime. Is it to move from one strategy to another as times charge? As relative value charges? As Big Data enables professionals to create ever larger numbers of ways, to put it bluntly, to "beat the market." For surely that's what we're talking about. But investors as a group do not and cannot beat the market.

So I express again my professional skepticism about the utility of Big Data in the continuing search for the Holy Grail of consistently superior investment returns. Where, then, should we turn? To Little Data! The simple arithmetic of passive investing in the form of the broad market, widely diversified, low-cost Traditional Index Fund (TIF[16]), in which the stock market return is there for the taking.

The rationale for such a TIF—exemplified by the S&P 500 Index Fund, bought and held "forever"—is based on this simple principle, guaranteed to be eternal: Since returns in the stock market are by definition equal to the gross returns earned by investors as a group, then the returns earned by index investors paying 4

---

[14] With Yan Liu, Texas A&M University, and Heqing Zhu, Duke University, "... and the Cross-Section of Expected Returns," working paper, available on SSRN.

[15] Coy, 2017

[16] Establishing TIF as an acronym has not been easy. Please help!

basis points per year are guaranteed to exceed the returns earned by active investors as a group paying, say, 200-plus basis points. *Gross return minus cost equals net return*. Talk about Little Data!

I may be biased—of course I am!—but I challenge today's brilliant academics and quantitative investors to rebut this proposition. RESOLVED: That the probability of relying on one strategy after another, one factor after another, one mutual fund after another, and/or one fund manager after another, has a chance of less than 1/10 of 1% to provide a lifetime wealth accumulation exceeding that of an all-market index fund, bought and held "forever."[17] And from my many talks with academics and quants, that is the way that you have actually placed your own investment bets. And if not yet, there's always tomorrow.

[17] The resolution that I propose comes close to what my sort-of-mentor Paul Samuelson wrote in the first issue of *JPM* (Fall 1974). "Show me the brute evidence that fund managers can win. " His paper, "Challenge to Judgment," helped inspire me to create the world's first index fund in 1975.