

INVESTMENT STRATEGY RESEARCH:  
DON'T FALL INTO THE DATA MINE

By Grant McQueen and  
Steven Thorley

*Grant McQueen and Steven Thorley are both associate professors of finance at the Marriott School of Management at Brigham Young University in Provo, Utah.*

*This article is an updated version of an article that originally appeared in the January 2000 issue of the "Investor Relations Quarterly," a journal published by the National Investor Relations Institute.*

In a hilarious monologue, comedian Bob Newhart reports the results of experiments based on the theory that with enough monkeys, typewriters, and time, the monkeys will eventually re-create all the literary classics. From an imaginary type-written page, the comedian reads, "To be or not to be. That is the ... gazornanplat."

Funny stuff for a comedian—but a sobering lesson for investors. And that lesson is simple: "Don't fall into the data mine."

Data mining is the practice of finding forecasting models by searching through databases for correlations, patterns, or trading rules. The problem with data mining is that after searching over enough variables or rules—say, 100—a researcher will find, just by chance, about ten that are statistically significant. And because they are a statistical fluke of the particular data series, they offer no real potential predictive abilities. Data mining becomes particularly problematic when the final pattern is proclaimed significant without providing the number of mining attempts.

This article revisits our study of data mining, which was originally published in the

*Financial Analysts Journal*,  
March/April 1999.

#### NON-FINANCIAL DATA MINING

Data mining is not solely a financial phenomenon; rather it rears its head in a variety of applications. An unscrupulous lawyer wanting to sue, for example, the local water utility could search through health records for an unusually high incidence of all types of cancer. There are many kinds of cancer and a community will be above the national average in some kinds of cancer and below average in others. And, with high probability, the community will have a significantly higher-than-normal incidence of one or another type of cancer. Such alleged "hot spots" are great for terrorizing residents and negotiating (read blackmailing) with utilities, but in terms of long-term health, the "hot spot" typically turns out to be "much ado about nothing."

The uproar surrounding Michael Drosnin's recent book, *The Bible Code*, also illustrates the data mining problem. Using a computer, Drosnin electronically searched, forward and backward, the letters from the Hebrew Bible using equidistant letter sequences (i.e., every 6<sup>th</sup> letter) for words. Combinations of these words (dinosaur and asteroid, for example) found in proximity to one another allegedly form secret messages. Given the large number of letters in the Bible and the number of letter distances the computer searched, many words and some word clusters should be found by chance. Thus, we were not surprised when *Solving the Bible Code Puzzle*, by McKay, Bar-Natan, Bar-Hillel, and Kalai, refuted the original Code claims or when an Internet site sprung up with a

list of mathematicians unconvinced by the data-mined "evidence" (See *Mathematicians' Statement on the Bible Codes* at [www.math.caltech.edu/code/peyton.html](http://www.math.caltech.edu/code/peyton.html)).

#### FINANCIAL EXAMPLES

With abundant data resources, powerful statistical software, and fast computers, the finance profession has all the tools needed for data mining. Furthermore, the profession has ample motivation to mine. Society rewards successful investors with power, prestige, and the choice of an endowed chair at a well-known business school, a Wall Street bonus big enough to endow such a chair, or the proceeds from a best-selling book.

In our study, we chose the Foolish Four investment strategy to illustrate financial data mining. We chose the Foolish Four because it is very popular, and because it is a product of a subtle form inter-generational mining. Although the Foolish Four promoters themselves searched only four variables (yield, price, number of stocks, and weighting schemes), they built their strategy based on the results of several generations of researchers, each building on the work of prior researchers who searched over a few variables and tried several approaches before publishing their "best" strategy.

The Foolish Four is an investment strategy touted by David and Tom Gardner in their best-selling book, *The Motley Fool Investment Guide: How the Fools Beat Wall Street's Wise Men and How You Can Too*, and on their popular website, [www.fool.com](http://www.fool.com).

In *The Motley Fool Investment Guide*, the Foolish Four strategy is described as a simple way for investors to beat the Dow. The

name “Fools,” with a capital F, derives from Shakespeare’s Fools or court jesters, who, with quick wits, were able to confound conventional wisdom by using plain and simple truths.

The best way to understand the Foolish Four strategy is to track its development:

**1st generation: Dogs of the Dow**-- The Dogs of the Dow approach sorts the Dow Jones industrial average stocks by their dividend yields at the beginning of each year and then buys an equally-weighted portfolio of the ten highest-yielding stocks. This strategy was popularized in the book *Beating the Dow*, by Michael O’Higgins and John Downs, first published in 1991.

**2nd generation: Five Dogs**-- The Five Dogs strategy calls for an additional winnowing of the ten Dogs of the Dow based on price, buying an equally-weighted portfolio of the five lowest-priced stocks within the ten highest-yielding set. This strategy was also introduced in *Beating the Dow* and in Knowles and Petty’s *The Dividend Investor*, first published in 1992.

**3rd generation: The Foolish Four**--The Foolish Four refines the Five Dogs in two dimensions, first by dropping out the lowest-priced stock of the Five, and second by doubling up on the second-to-the-lowest priced stock. This perturbation of the Five Dogs strategy was motivated by O’Higgins’ discussion in *Beating the Dow* of second-to-the-lowest-priced stocks, which he termed Penultimate Profit Prospects. The strategy itself was introduced by the Gardners and Ann Coleman in *The Motley Fool Investment Guide*, first published in 1996.

So, to implement the Foolish Four strategy, a four-stock portfolio is created at the

beginning of January each year from the five lowest-priced of the ten highest-yielding Dow stocks, excluding the very lowest-priced stock. The portfolio is invested 40% in the second-to-the-lowest priced stock, and 20% each among the other three stocks.

In *The Motley Fool Investment Guide*, the Gardners report that for two decades, from 1973 to 1993, the Foolish Four returned an annual average return of 25% and that it “should grant its fans the same 25% annualized returns going forward that it has served up in the past”.

In our study, we confirm the first part of this claim (25% historical returns), but due to the data-mining involved, we contest the latter part (the same 25% annualized returns going forward).

Table 1 reports annual returns and summary statistics on an equally-weighted portfolio of all 30 stocks in the Dow Jones industrial average (the Dow 30), as well as all three generations of the strategy—the Dogs of the Dow, the Five Dogs, and the Foolish Four. (We also include data on our own new strategy, a 4th generation discussed below called the Fractured Four.)

The firms included in the Dow 30 at the beginning of each year are based on the May 28, 1996 Wall Street Journal special report “100 Years of the DJIA” and on the March 13, 1997 *Wall Street Journal* which reports the details of the 1997 change. Dividends, prices, and returns are from the University of Chicago’s Center for Research in Security Prices (CRSP) tapes.

Following *The Motley Fool Investment Guide*, we formed portfolios using the closing price on the first business day of the year and held the portfolios until the close of the first business day of the following year. We calculated dividend

yields by using the last ordinary quarterly dividend times four, divided by the closing price on the first business day of the year. The returns assume immediate reinvestment of all cash dividends and proceeds from selling other distributions (i.e., warrants) in the stock that paid them.

In Table 1, the time period covered is the same 21-year period (1973 to 1993) used by the Gardners in their book documenting the Foolish Four. Over that time period, the Dow 30 portfolio had an average annual return of 14.8%, while the Foolish Four portfolio had a higher return of 28.8%. Our original *Financial Analysts Journal* article also included an analysis of risk. Among other things, we found that the reward-to-risk ratio over the 21 years was higher for the Dogs of the Dow portfolio than either the Dow 30 or the Foolish Four. In this article, we focus on returns, but we do include the average standard deviation for the portfolios. Standard deviation measures the range of actual returns around the average; the greater the standard deviation, the greater the variability of returns and therefore the greater the risk.

You can see from the table that the Dogs of the Dow and Five Dogs strategies’ average returns increased monotonically from the Dow 30 to the Foolish Four, just as if each successive portfolio was a better investment strategy than the preceding portfolio. But this monotonic improvement in historical average returns could also be the result of successive rounds of data mining.

Successful strategy or successive searching? That is the question.

As researchers, we are unable to fully answer this question without knowing how many strategies were explored

unsuccessfully. This lack of knowledge would include unsuccessful strategies tried by the Gardners themselves (which we call intra-generational mining), unsuccessful strategies tried by prior investment advisors such as Michael O'Higgins or Knowles and Petty (which we call inter-generational mining), and even unsuccessful strategies tried by seemingly unrelated researchers (which we call meta-generational mining).

The inability to assess the significance of a trading rule found after extensive collective searching has been dubbed the "file drawer problem" in the financial academic literature. Essentially, the true statistical significance of successful investment strategies can be assessed only after quantifying the number of unreported or unpublished failures gathering dust in the file drawers of stock market analysts, traders, and researchers. Because we do not know the degree of collective data mining that led to this one "successful" rule, we cannot directly assess its significance.

Our concerns about meta-generation mining deserve some clarification. Every day, scores of individuals try out different investment strategies looking for correlations between returns and a wide variety of other variables. Judging from our casual sample of taxi drivers, doormen, and garbage collectors, searching for stock investment strategies may have replaced baseball as our national pastime. In a neo-Darwinist fashion, hundreds of attempts are made but only a few successful strategies survive and make it into a popular book or academic journal. Given this financial frenzy, we expect to hear about many successful strategies, most of which occurred by chance, and we are consequently skeptical about the true significance of any one

strategy. That is, even if we knew that the Gardners only searched over four variables and that Knowles, Petty, and O'Higgins wore blinders and never peeked at variables such as price-earnings ratios, market capitalizations, or market share (no intra-generational mining), we would still be skeptical about the Foolish Four because of meta-generational mining. Because of the untold hundreds of Gardners and O'Higgins, each earnestly searching for a rule, we expect to hear about many apparently successful strategies just by chance.

#### OUR OWN MINE

To illustrate both its ease and absurdity, we took a shot at the data mining game ourselves. We noticed that in even years, the Foolish Four stocks typically beat the Dow 30, but that in odd years, the opposite was true. Thus, after searching over only one variable, the year's last digit, we found a new formula called the "Fractured Four." In our Fractured Four, we proscribe holding the Foolish Four stocks in equal weights in even years and buying only the second-to-the-lowest priced stock in odd years.

The results? Our Fractured Four earned an average annual return of 36.9% (see Table 1) with a 21-year track record! Just as the Foolish Four beat the Dogs of the Dow by about 8% per year, the Fractured Four in turn beat the Foolish Four by 8% per year.

But by now you should see the folly of our claim. The Fractured Four is so obviously the product of inter-generational mining that it drives home the point: "successful" investment strategies, even those that have been "successful" for 21 years and those that search over only

one variable, may turn out to be fool's gold, not a golden chalice.

#### AVOIDING THE PITFALLS

One protection against data mining is to be skeptical about studies that search over a large number of variables or that are built on prior rounds of research in which the exact number of variables is unknown.

David Leinweber, in his satiric study, "Stupid Data Miner Tricks", mines the United Nations' database and finds that butter production in Bangladesh can predict 75% of the variation in the Standard & Poor's 500 stock index.

For two reasons, detecting data mining is never as easy as counting up all the variables in the UN database. First, some researchers reveal the significant, but conceal the insignificant, variables. Second, even researchers who reveal all their own variables may be an unwitting partner to mining. Such researchers may choose their variables after reading the results of prior researchers, yet not report this interdependence nor the list of prior researchers' variables.

A second protection against data mining is to expect a plausible explanation or story as to why the trading rule works. In our research of the Foolish Four strategy, we were troubled by the weakness of the theory. Sorting by dividend yield may well be supported by theory, but a simple sort of raw share prices seems nonsensical—the only difference between a \$20 share and \$10 share is a stock split. In *Beating the Dow*, O'Higgins attempted to develop a story or theory for sorting by price: In essence, stocks with low share prices are alleged to be only temporarily out-of-favor and hence, their price will soon rebound; in contrast, the stock

with the very lowest price is in real, permanent trouble. This Goldilocks theory of stock price--not too low, not too high, but just right--suggests sorting for out-of-favor stocks based on past returns, not share prices. We question a theory that partitions the two lowest-priced stocks, differing by only a few cents in price, where one becomes a perpetual loser not worthy of purchase, and the other a persistent winner, worthy of an extra proportion. This logic seems to create a distinction without a difference.

A third protection against data mining is to test the strategy out-of-sample. If owning high-yield low-price (but not too low) stocks is rewarding, this strategy should also work in the 1950s and 1960s or after 1993, not in just the 1973 to 1993 period reported by *The Motley Fool Investment Guide*. A valid strategy should work on 30 large non-Dow stocks as well as the 30 stocks in the Dow Jones Industrial Average. But such is not always the case.

In our first out-of-sample test, we examined the returns just after the 1973 to 1993 period used to discover the Foolish Four. Between 1994 and 1998, the Dow 30 portfolio had returns of 3.9%, 37.6%, 26.1%, 22.4%, and 16.8%, and over the first 10 months of 1999 the Dow 30 portfolio was up 15.7%. Over the same years the Foolish Four returned -6.1%, 40.1%, 33.3%, 28.1%, 17.0%, and -8.0% for the first 10 months of 1999. Thus, a \$1,000 investment on January of 1994 in the Dow 30 was worth \$2,982 on November 1 if invested in the Dow 30, compared to only \$2,418 if invested in the Foolish Four. With just under six years of post-mining data, strong conclusions cannot be reached; nevertheless, the Foolish Four is not off to a good start.

Followers of the Foolish Four may quibble with our returns for 1998 and the first 10 months of 1999 for the Foolish Four. These numbers are based on a strict following of the *original* strategy (now called the Foolish 4.0) as published in the book. However, the original strategy was "tweaked" once in January 1998 (the Foolish 4.1) and once again in January 1999 (the Foolish 4.2). We focus here on the Foolish 4.0 rather than comparing the Dow 30 to the moving target of generations 4.1 and 4.2.

Table 2 reports the results of our second and third out-of-sample tests. In our second test, in order to match out-of-sample test sizes to the Gardners' in-sample test size, we look at the 21 years of data, 1952 to 1972, just prior to the in-sample period. Over these 21 years, the Dow 30 portfolio returned 12.7% on average, whereas the Foolish Four returned only 11.9%. Thus, in both the 21 years before and the nearly 6 years after the sample period of data used in the Gardners' book, the Dow 30 has beaten the Foolish Four.

In our third out-of-sample test, we use in-sample years, 1973 to 1993, but rank stocks according to dividend yield and price each July (mid-year) rather than each January. Thus, our third test is only quasi-out-of-sample since we use the same years but a different rebalancing anniversary. With mid-year initiation, the Dow returned 13.4% and the Foolish Four 15.7% on average. In this quasi-out-of-sample study, the Foolish Four did beat the Dow, but not by the 14% using January anniversaries.

In our fourth out-of-sample test we created an ersatz Dow 30 by finding a twin for each of the 30 Dow Jones Industrial Average stocks. The twins were

found for each Dow stock in each year by first finding all stocks with matching three-digit SIC codes (similar industries) and second, within the industry, picking a stock closest in size (market capitalization) to the Dow stock. In two cases, we resorted to a two-digit SIC code when no three-digit match was available. One could quibble with our matching algorithm--Some matches seem very good (Exxon with Mobil or Coke with Pepsi); in contrast, one is hard-pressed to find a "twin" for AT&T in the 1960s and 1970s, although our algorithm picked Pacific Telephone & Telegraph. Nevertheless, we do end up with 30 fairly large (typically the second largest company in the industry) companies in a variety of industries not unlike the 30 companies in the Dow.

We then sort the 30 ersatz Dow companies by dividend yield, picking the 10 highest dividend yielders and then resort by price, creating a twin to the Foolish Four. Our ersatz portfolio test runs from the first business day of January 1963 to January 1999. The starting date corresponds to the first year of CRSP daily data. Over the 36 years of ersatz portfolio returns, the Dow 30 Twin averaged 13.9% per year and the Foolish Four Twin averaged 14.9%. Thus, the Foolish Four Twin did beat the Dow 30 Twin, but only by 1%. In the process of generating the one percent premium, the Foolish Four strategy exposed the investor to higher risk, more transaction costs, and earlier taxes caused by more frequent trading.

In summary, the Foolish Four is not a tremendous traveler and is rather timid in new territory. It lost to the Dow 30 in the 21 years before 1973 and the 6 years since 1993. Furthermore, its performance with July starting months and the performance using the

ersatz portfolio was only slightly ahead of the Dow 30.

In case we have overstated our point, not all data mining is bad. Ann Coleman from the Motley Fool notes that the link between smoking and heart disease started in a data mine. However, the smoking/heart disease case is an example of beneficial mining--mining that cleared all three hurdles discussed above. First, researchers readily disclosed the number of variables searched over before finding the link to smoking. Second, subsequent work developed a clear and concise story or theory about the mechanism that linked smoking and heart disease. And third, the link held up time and time again in out-of-sample tests with new subjects not in the original studies.

## CONCLUSIONS

If you torture the data, it will confess. Statistically significant but spurious patterns occur in data simply by chance. Thus, one should not be surprised to find "word clusters" in the Bible after searching over enough equidistant letter sequences, or cancer "hot spots" after searching across enough towns and enough types of cancers. Even finding that stock returns are related to the phases of the moon is not surprising, if the finding came at the end of a long search of unrelated non-financial variables (see Timothy Crack's, "A Classic Case of Data Snooping," in *The Journal of Financial Education*, Fall 1999).

The Motley Fool's Foolish Four portfolio allows us to discuss three forms of data mining: inter-, intra-, and meta-generational, which are hard to avoid and difficult to detect. We measured the Foolish Four against three criteria: the number of variables, the

strength of the story, and the performance out-of-sample. The strategy came up wanting.

Our point, however, is not to pick on this one particular strategy, but to warn against the claims of any strategy that is touted as a "proven approach" leading to market riches. A close examination of the claims with a skeptic's eye should prevent you from falling into the data mine.

**Table 1**  
**Historical Returns for Various Portfolios: 1973 to 1993\***

Year	Dow 30 (%)	Dogs of the Dow (%)	Five Dogs (%)	Foolish Four (%)	Fractured Four (%)
1973	-10.03	4.92	20.97	31.73	46.39
1974	-15.49	1.46	-3.67	3.73	6.28
1975	49.46	60.78	73.69	93.61	154.73
1976	31.53	35.23	41.15	44.57	41.72
1977	-13.30	-1.90	0.49	0.28	6.36
1978	0.92	2.77	3.08	7.43	8.88
1979	12.96	11.57	8.18	8.98	-9.08
1980	24.17	31.04	46.37	48.47	46.10
1981	-2.16	6.11	1.03	2.40	26.59
1982	22.16	23.46	30.91	54.07	44.25
1983	34.62	38.87	35.84	33.74	34.57
1984	-0.46	6.96	10.09	11.86	14.12
1985	31.33	30.49	40.02	38.64	26.20
1986	23.95	33.96	24.46	30.24	29.04
1987	17.89	10.11	12.67	3.41	6.46
1988	19.14	24.23	21.42	21.51	20.83
1989	31.10	29.81	12.20	9.45	14.55
1990	-7.24	-7.82	-14.93	-13.15	-12.65
1991	29.60	35.49	62.58	92.80	186.11
1992	12.72	11.85	31.08	40.44	43.72
1993	18.76	25.39	34.04	41.24	41.69
Average	14.84	19.75	23.41	28.83	36.99
S.D.**	17.93	17.02	22.21	28.37	48.06

\*Portfolios are formed using closing prices and trailing dividend yields on the close of the first business day of each year. The portfolio definitions are:

Dow 30: An equally-weighted portfolio of all 30 stocks in the Dow Jones Industrial Average  
Dogs of the Dow: An equally-weighted portfolio of the ten Dow industrial stocks with the highest dividend yield  
Five Dogs: An equally-weighted portfolio of the five lowest-priced stocks in the Dogs of the Dow Portfolio  
Foolish Four: A four-stock portfolio that eliminates the lowest-priced stock in the Five Dog portfolio and doubles up (40% of portfolio value) on the second to the lowest-priced stock  
Fractured Four: The Foolish Four stocks in equal weights in even years and the lowest-priced stock in the Five Dog portfolio in odd years

\*\* S.D.: Standard deviation measures the variability of actual returns around the average return. The greater than standard deviation, the greater the variability of returns, which implies greater risk.

Table 2.  
Out-of Sample Historical Returns  
for the Dow-30 and Foolish Four Portfolios

<u>1952 to 1972: January Rebalancing</u>			<u>1973 to 1993: July Rebalancing</u>		
	Foolish			Foolish	
	Dow 30	Four		Dow 30	Four
Year	(%)	(%)	Year	(%)	(%)
1952	13.72	15.88	1973	-4.38	2.09
1953	2.17	-14.39	1974	-2.40	-19.32
1954	55.17	68.76	1975	23.65	21.45
1955	24.14	19.08	1976	23.08	23.26
1956	9.20	10.49	1977	1.73	11.76
1957	-7.79	-15.99	1978	-7.94	-8.66
1958	42.45	52.04	1979	7.53	14.73
1959	17.51	11.53	1980	10.75	13.41
1960	-3.70	-5.65	1981	19.58	23.55
1961	23.18	25.07	1982	-14.55	9.11
1962	-7.09	-13.45	1983	60.68	53.99
1963	21.87	14.18	1984	-2.68	4.23
1964	19.41	23.64	1985	26.71	31.42
1965	17.06	28.71	1986	37.75	39.81
1966	-15.42	-24.89	1987	32.57	51.13
1967	22.09	37.11	1988	-5.96	-15.59
1968	10.36	11.00	1989	18.15	24.49
1969	-10.15	-14.52	1990	18.74	5.33
1970	8.00	-11.07	1991	5.80	7.14
1971	7.96	14.35	1992	18.56	5.35
1972	15.87	18.14	1993	13.48	31.37
Average	12.67	11.91	Average	13.37	15.72
S.D.**	16.57	23.09	S.D.**	17.40	18.89

\*\* S.D.: Standard deviation measures the variability of actual returns around the average return. The greater than standard deviation, the greater the variability of returns, which implies greater risk.