Wenzel
Analytics
Inc

Data Mining Research using Fractal Dimension Indicator (FDI) from XLQ

October 19, 2013

I. Background

A couple years ago I became fascinated with fractals after watching an excellent Teaching Company course on chaos theory while on the daily treadmill. Stocks charts are obviously self-similar and fractal, since without a label, no one could differentiate a chart with seconds, minutes, hours, days, or weeks for each period or point on the chart of stock price patterns. Yet, few investors or traders use a fractal methodology. Instead, many use the assumptions of a normal distribution, such as is represented by standard deviations and linear trends.

The search for resources that effectively use a fractal approach to markets lead to many dead ends. Resources I have come to appreciate are daily reports issued by David Nichols. One is called the Fractal Market Report (www.fractalmarketreport.com) and another is his (www.fractalgoldreport.com). Another resource is the utility sold by Matt Trivisonno (www.trivisonno.com). I found it to be very accurate and helpful in foretelling when a trend would end as well as when to expect a breakout up or down. The shortcoming of the utility is that one has to enter the stock symbol for each chart, and can then view the Fractal Dimension Indicator (FDI) chart against the price chart for daily, weekly and monthly data.

The Trivisonno reference material pointed to an article by a biometric researcher in Caracas that had a page of Visual Basic formula for a FDI. I sent this to Leo van Rijswijk who has a product called XLQ that brings a vast amount of financial technical data into any Excel spreadsheet. Leo van Rijswijk sent me a beta of the FDI in XLQ, which enabled doing the quantitative research reported here.

II. Data

Using the XLQ add-on, I assembled monthly data on several variables for a list of more than 4,000 stocks having 30 million or more market capitalization and priced above one dollar going back 112 months to 6/1/2004. After cleaning the data for missing cells, I had 375,671 rows of data with a dependent variable of percentage change going forward one month, three months and twelve months.

Several variables were not predictive, such as tightness of upward trend as measured by standard deviation of rate of change for 30 periods divided by the moving average. Other variables such as the relative strength of price change or volatility relative to the market were not productive and are not reported here.

The same process used for monthly data was applied to weekly data going back 103 weeks to 9/12/2011. After cleaning, this left 404,278 rows of data for analysis.

A relatively small number of rows having inordinate or outlying average returns can greatly distort the average returns, and not be replicated when trying to implement the findings. Therefore the dependent variables (average percent change going forward) were capped at the top and bottom 2.5% of records. So on a sort by percent change, whatever the percent change down from the top and up from the bottom, that rate of change was applied to all the outliers beyond that level.

Lee Wenzel
(952) 944-2699
Lee@WenzelAnalytics.com
www.WenzelAnalytics.com

Wenzel Analytics, Inc.
Registered Investment Advisor
8666 Westwind Circle
Eden Prairie, MN 55344

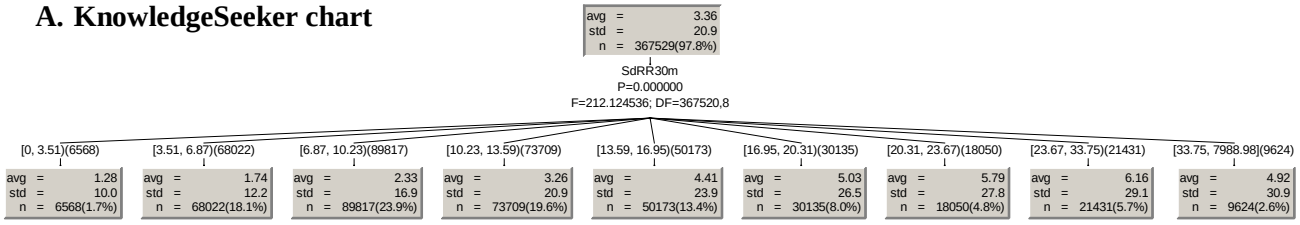
III. Method of Analysis

The data were imported into a data mining tool called KnowledgeSEEKER for exploration and analysis. Combinations of variables and respective values that produced consistent high returns were then compared across time periods going forward (one month, three months and twelve months or one week, four weeks and thirteen weeks), as well as between monthly and weekly data sets.

After finding the best set of screen variables, which is described below, I discovered that I had made a "mistake" in my formulas. The variable is the coefficient of variation (COV) of price (standard deviation over 30 periods divided by the simple moving average). This is not to be confused with the COV of the rate of change. What I later discovered is that for the monthly data, my formulas for the numerator was monthly but for the denominator was weekly. So I "fixed" it to be all monthly and went back and ran the tests again. However, nothing could match the results of using my accidently contrived variable as part of a screen.

IV. Findings

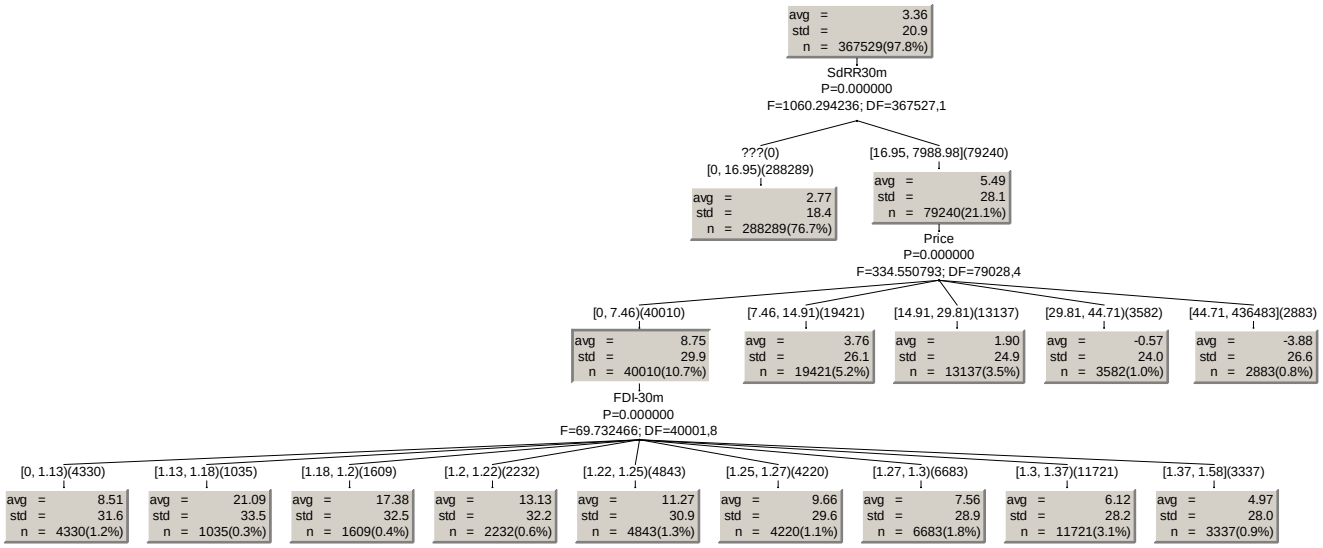
A. KnowledgeSeeker chart



This first tree gives a picture of how KnowledgeSEEKER works. The 3.36 is the average percent change over the next three months, with a standard deviation of 20.9. The reason the 367,529 records is 97.8% of the total is because we don't have returns going forward three months for the last two months. The P is the degree of significance, which in this case is > .000001. The F value and degrees of freedom are given. We have set the software to divide these 367,529 records into 20 clusters sequentially by the value of the variable, in this case the standard deviation of the rate of return for the past thirty months. Adjacent cells with similar statistical characteristics are combined, leaving us with nine cells. The next line gives the value range for the variable and the number of rows in that cell. In the cell you can see the average value which in this case is percent change over the next three months, as well as the standard deviation and the percent of the total falling in that cell. Looking at the values, you can see that there is a linear progression up to the last cell, which is still higher than the average of 3.36%.

Arbitrarily I collapsed everything above and below 16.95, and then expanded the tree as shown on the next page.

A breakdown by price of the root cell > 16.95 yields significantly higher returns for price < $7.46. A breakdown of those records by the Fractal Dimension Index (FDI) yields a very nice linear distribution after the first cluster. A collapse of FDI >1.13 and < 1.22 had an average three-month percent change of **16.22% compared to 3.36% for the database average**. This screen had 4,876 records or 1.3% of the total with a coefficient (standard deviation divided by average) of 2.0.



The FDI was found to be a poor differentiator at the root level, but consistently a good differentiator below other variables with limited interaction effect, meaning that it worked similarly with many different variables.

B. Log Table

The table below is a summary taken from the tree shown above. It also includes comparable results for percent change one month and twelve months forward.

COV and FDI																		
6/1/04 thru 8/1/13 Monthly; 112 months																		
				All - RR1m					**All - RR3m**					**All - RR12m**				
		Range		**Avg**	**StDev**	**N**	**N %**	**COV**	**Avg**	**StDev**	**N**	**N %**	**COV**	**Avg**	**StDev**	**N**	**N %**	**COV**
				1.05	11.4	**375,671**	100%	10.9	**3.36**	20.9	367,529	97.8%	6.2	**12.61**	47.1	331,715	88.3%	3.7
SdRR30m		16.95		1.59	15.3	80,309	21.4%	9.6	5.49	28.1	79,240	21.1%	5.1	20.65	62.6	74,193	19.7%	3.0
	Price		7.46	2.44	16.2	40,670	10.8%	6.6	8.75	29.9	40,010	10.7%	3.4	34.67	68.5	36,621	9.7%	2.0
		FDI-30m	1.18	6.57	18.8	1,062	0.3%	2.9	21.09	33.5	1,035	0.3%	1.6	73.87	76.3	959	0.3%	1.0
			1.18	4.90	17.5	1,654	0.4%	3.6	17.38	32.5	1,609	0.4%	1.9	61.34	70.8	1,495	0.4%	1.2
			1.2	3.70	17	2,278	0.6%	4.6	13.13	32.2	2,232	0.6%	2.5	47.79	70.9	2,083	0.6%	1.5
			1.22	2.96	16.4	9,185	2.4%	5.5	11.27	20.9	4,843	1.3%	1.9	42.78	70	4,490	1.2%	1.6
		Subtotal	1.13	3.60	17.4	9,364	2.5%	4.8	16.22	32.7	4,876	1.3%	2.0	57.77	72.8	4,537	1.2%	1.3

C. Overall Finding

A limitation of the above screen is that when the final cluster is broken down by month, the results were not as consistent as for other screens. The most productive variables were found to be:

1. Price standard deviation (30), monthly or weekly, divided by weekly simple moving average
2. Rate of return standard deviation (30)
3. Rate of return standard deviation (6)
4. Price
5. FDI (30)

Each of these variables had strong linear results. Together a screen of about two percent of the total database produced returns from three to five times the database average for different periods going forward, and for both monthly and weekly data.