Wenzel
Analytics
Inc

Data Mining for High Return Criteria

February 20, 2021

Study Parameters

This analysis was of monthly data pulled from Stock Investor Pro published by AAII starting in February of 2001. The data run through January of 2021. Returns are available for one month, three months, six months and twelve months. The last data with annual returns is prior to January of 2020; returns over shorter time periods draw from more recent data. The returns were capped at the lower and upper one percent extremes as very high or very low returns excessively skew the averages.

The data were analyzed using KnowledgeSEEKER, a decision-tree form of AI which gives hierarchical trees at specified parameters of significance and cluster sizes (example on next page). Unless specified otherwise, the significance required was at the 0.01 level. The minimum cluster size varied depending upon the number of sub-clusters or levels in the hierarchical tree with subordinate screen criteria.

The dataset had 813,191 rows and 86 columns or independent (causative) variables, plus the fields for returns. For some explorations the data were randomly split in half in part because of memory issues and the need for training and test datasets to examine over-training and results which do not replicate.

Results are measured by price change and the coefficient of variation (COV) calculated by standard deviation divided by return (price change). The COV is the inverse of the Sharp Ratio if dropping the comparison to the Treasury yield. Variations test for consistency over time as well as consistency between stocks within a given cluster. Highly variable returns are of less utility in that if we buy ten stocks out of a cluster of 50, our sample may not be representative.

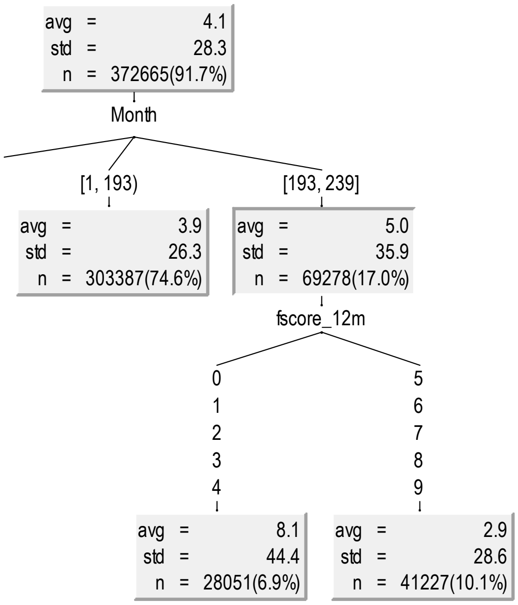
Initial Findings

The initial analysis was for six-month returns covering the full applicable time period. As always, the most significant findings were for differing time periods – which is not useful except to be wary of relying on any finding for any specific time period. Returns vary by sector which may not be predictive. The most significant findings are for price change patterns rather than for fundamental variables such as sales, earnings, debt, f-score, etc. In one analysis, the first five variables by level of significance were all related to price change, followed by sector beta which is also price change, followed by another three of specific price patterns. This points me more to technical analysis than anything based on fundamentals.

To take F-score as an example of a fundamentals-based variable, compared to the overall annual rate of return of 8.6%, the highest F-scores of 8 or 9 give a modest improvement to 11.2%, and score of 7 had returns of 9.4%. However, a 0 or 1 score also gave a 9.0%.

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

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



For the most recent three-plus years, the results are inverse of expectations as shown in the KnowledgeSEEKER tree shown on the right. The average percent change over 26 weeks for the entire period was 4.1% (8.2% annualized. F-scores were not available for the entire period). The standard deviation was 28.3% and the N was 372,665 in the training dataset. Looking at the most recent 41 months under Month (number), the average percent change was 5.0%, the standard deviation was 36% and the N was 69,278. Breaking down the F-score for that time period, low F-scores of 0-4 had an average percent change over the next 26 weeks of 8.1% while F-scores of 5-9 had average precent change over 26 weeks of only 2.9%.

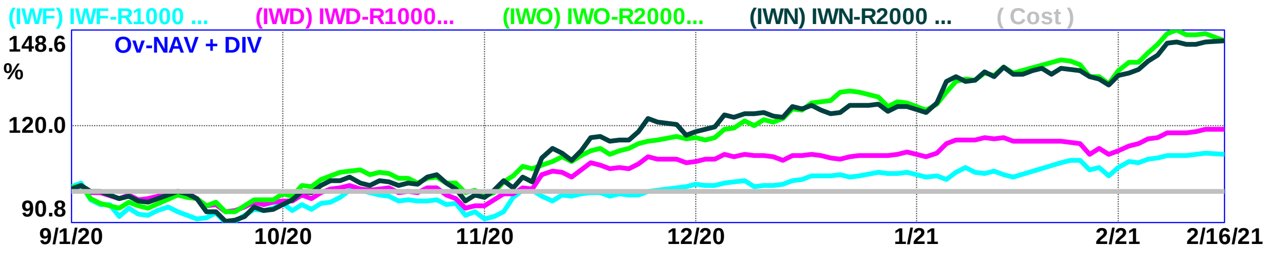
Variables that apply to industry and sector are more predictive than variables that apply to individual stocks. For example, for the entire period combining industry sales growth over the last 12 months < 6.8, sector beta >1.15, and price change over the prior 26 weeks < -55 had an annualized return rate of 62%.

Reversals are far more predictive of positive returns than momentum. Investors in general overweight positive fundamentals and momentum while underweighting low quality and falling prices. An over-priced high-quality stock going up is more likely to be a gotcha.

Criteria for consistent returns over long time periods were not to be found. My first conclusion was that returns projected forward should be related to the duration of the data from which the predictions are based. Analysis of data from twenty years might be relevant to returns over the next ten years after screening for consistency in each of the past ten time periods (could be set to twenty periods). Most investors won't wait ten years through fluctuating returns.

To illustrate, I found a large cluster of high returns with a very low COV but when I reviewed it by ten time periods, the only cluster with any results meeting the criteria encompassed April of 2020. It is very difficult to find a screen that will work over both brief time periods and over long periods.

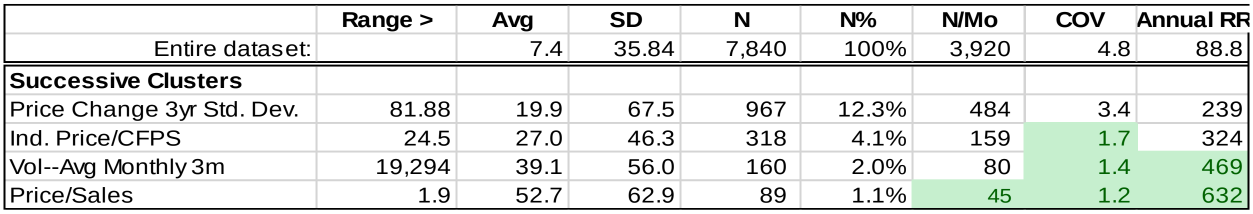
Market dynamics periodically shift. Obviously, April was such a shift. Another occurred about the time of the election as shown on the chart below showing large and small cap growth as well as large and small cap value. Since the election small caps as measured by the respective Russell 2000 ETFs are up 49% while large caps are up less than 20%.



Not only should expectations going forward be related to the past time period from which the data are drawn, but I decided to work only with data since a shift in market dynamics and find what works now. In this case I took the data from only November and December with returns being one month later.

Screen Findings Since the Election

Four-week later average percent change found returns that were both high and consistent by using the following criteria:



	Range >	Avg	SD	N	N%	N/Mo	COV	Annual RR
Entire dataset:		7.4	35.84	7,840	100%	3,920	4.8	88.8
Successive Clusters								
Price Change 3yr Std. Dev.	81.88	19.9	67.5	967	12.3%	484	3.4	239
Ind. Price/CFPS	24.5	27.0	46.3	318	4.1%	159	1.7	324
Vol--Avg Monthly 3m	19,294	39.1	56.0	160	2.0%	80	1.4	469
Price/Sales	1.9	52.7	62.9	89	1.1%	45	1.2	632

Across the top, Avg is the monthly percent change. SD is the standard deviation of the stocks within the cluster. N is the number of stocks. N% is the size of the cluster relative to the entire dataset. N/Mo is the average size of the cluster each month. COV is the coefficient of variation calculated as the standard deviation divided by the percent change. Annual RR is the annualized rate of return (Avg * 12).

Ind. Price/CFPS is average industry price divided by cash flow per share.

These criteria were applied to a current dataset. I reviewed the charts of resulting selections, ran them through the Louis Navellier Grades and selected thirteen positions. (AMRS, ASTC, BLNK, COCP, CEL, IBIO, IGC, IZEA, OPTT, TRXC, TTOO, UAVS, VUZI, XXII)

As one might anticipate, these are highly volatile stocks. Today saw some up more than 10% and some down more than 10%. After the first day the portfolio was down 10%; after the second day it was still down 4%; after four days (today) it is down 6.8%. The Equal Weighted Russell 1000 is down only slightly over the same time period. Seeing it as a rodeo, I'm still riding.