Stock Selection:

A Test of Relative Stock Values Reported over 17 ½ Years

By Charles D. Kirkpatrick II, CMT
Introduction

In the 1960’s and 1970’s, as the ability to use computers became more widespread, a number of experiments were performed on stock market and corporate data to determine the best variables required to select stocks. These experiments were crude by today’s standards, but in their innocence, these analysts discovered many truths and dispelled many myths. For example, one experiment with Price/Earnings ratios (P/E) suggested, ironically, that contrary to beliefs held even today, the best P/E was a high one, not a low one - that stocks with high P/E’s tended to outperform those with low P/E’s\(^1\). Experiments like these allowed analysts to focus on the value of a number of variables that heretofore had been too complicated or time-consuming to pursue.

At that time, when the random walk theory, beta theory, efficient market hypothesis (EMH) and capital asset pricing model (CAPM) were gaining in popularity, Robert A. Levy published a book\(^2\) and an article in the Journal of Finance\(^3\) based on his Ph.D. thesis at American University\(^4\) that showed how well-performing stocks, i.e. those with relative price strength, continued to perform well and that poorly performing stocks likewise continued to perform poorly. Levy’s theory was not original. The theory of relative price strength had been around for a long time.\(^5\) However, Levy, with the new aid of computer power, added some nuances and calculations that had not previously been used and found them to be very successful. Since they tended to refute the then popular theory that the stock price action was entirely random (the Efficient Market Hypothesis), his conclusions were subject to considerable criticism; his calculations and statistical evidence were severely condemned; and finally, his results were left in the dust of academic vitriol.\(^6\) Though long forgotten now by most analysts, his theories nevertheless have been kept alive by a few. For almost twenty years a model run in real time (‘live’) and published every
week for 17 ½ years, based largely on these theories, has show his calculations to have been and continue to be useful in selecting stocks with higher-than-market post performance.

The Test

Most computerized experiments and stock market models are calculated using what is called “optimization.”7 In the attempt to find variables that are important in determining the future of stock prices, most experiments use past data and adjust the variables and their parameters to find a ‘fit’ between those variables and stock post-performance. This is called “forced optimization.” Most discussion then centers around how closely the results fit the data, how sophisticated the statistical methods were, and why the results occurred as they did, forgetting that the results may have no usefulness in the future. Some computerized-trading model builders avoid forced optimization by splitting their data into several parts. They perform their experiments on one or more parts, and then test the results against the other parts. However, the best and most convincing test of any theory is to see if it works by itself using completely unknown data. This is what this study accomplished weekly over 17½ years.

In July 1982, to test variables of relative price strength and relative earnings growth, a selection and deletion criteria was established, a performance measurement determined, and a stock list developed (“List 1”). Later, in 1999, a second
list ("List 2") was established using slightly different criteria. Each list was reported weekly in Kirkpatrick’s Institutional Market Strategist, and periodically performance results were also reported. As of December 31, 2000, List 1 had appreciated 5086.6% versus a S&P 500 gain of 1087.6% and a Value Line Geometric gain of 221.9% (see Chart I on page 2). List 2, during a very difficult and slightly declining stock market, appreciated 137.3% versus a S&P 500 gain of 7.41% and a Value Line Geometric loss of 9.99% (see Chart II above). The second list also outperformed the original list which gained 75.19% over the same two year period. Most performances occurred during a generally rising stock market but none included dividends, which, though small in most cases, would have made the results even more impressive. Transaction costs were not included. Today, at radically discounted levels, commissions are almost negligible costs except in high turnover models.

**Selection criteria**

The first, and longest existing test list, List 1, included relative price strength, relative earnings growth, and a simple chart pattern as variables for stock selection. The selection criteria for List 2 were slightly different. Relative price strength and earnings growth were used but instead of a chart pattern, relative price-to-sales ratio ("PSR") was included to reduce the risk of loss.
The reason for the change in criteria between the second and first test list was that with the general market having risen since 1982 and the strong stocks having become so volatile, the danger existed that a severe correction would exert even more downward pressure on the list’s performance. For example, in the bear markets of the 1960’s and 1970’s, relative price strength performed well as a selection criteria initially until the very end of the general market declines when the strongest stocks tended to decline the sharpest and suffered disproportionately large losses.

To prevent such a loss in an individual stock, in List 1 a simple chart pattern was imposed as a ‘stop’ on negative price action to forcefully delete a stock early and prevent it from being caught in a severe decline. However, later, through tests of stock price patterns alone, no discernible advantage was gained. Therefore, to avoid changing the original selection criteria for List 1 and thereby interrupting its long record of success, List 2 was begun using another approach. Rather than have a price stop to minimize loss, the danger of negative performance was minimized in the beginning by selecting only those stocks trading at low relative price to sales. Presumably these stocks were trading at bargain prices already. As it turned out, three additional advantages arose from this model: (1) portfolio volatility declined rapidly - portfolio beta was consistently below one, whereas List 1 often had a portfolio beta approaching two, (2) turnover declined from an average holding period of 22 weeks in List 1 to well over a year in List 2, and (3) the size of the portfolio was considerably smaller and more manageable - near 5 to 15 stocks in List 2 versus up to 80 in List 1.

Performance measurement

Each week, before making changes to a list, the average percentage gain or loss of each stock in the list was recorded. For example, say the list included only stock A and stock B. If
stock A was up 5% and stock B up 1%, the list was recorded as having risen 3%, the mean of the two stock performances. These list performances were then accumulated each week over the test period. The equivalent in the real world would be for a portfolio manager to equally invest in selected stocks one week, record its combined performance for the next week, and then readjust each stock as well as add new ones and delete old ones such that for the coming week the portfolio would be equally weighted in each stock. Otherwise, the stronger stocks would accumulate over time into a larger relative position in the portfolio and have an unequal effect upon the portfolio’s total performance. Equally weighting of each stock each week, was the best method to reliably measure the criteria used in selecting the stocks.

Specific Selection Criteria

Relative Price Strength: Most measures of relative strength weigh a stock’s performance against a market average or index such as the S&P 500. This is wrong. The addition of a market average only complicates the results. For example, market averages are capital-weighted; individual stocks are not. Furthermore, this kind of measurement makes it difficult to weigh one stock against another, difficult to tell when price strength is changing, difficult to determine comparative periods, and is difficult to quantify for model building. The best calculation for a stock’s relative strength is to measure price performance equally against all other stocks over some specific time period. Until the arrival of computer power, this kind of calculation was very difficult and time-consuming. By the 1960’s it was not.

Several methods of quantitatively weighing price performance have been proposed. More recently, and since List 1 was begun, for example, Jegadeesh and Titman (1993) used six and twelve month returns held for six months during the period 1965-1989. Their results demon-
strated a post-performance excess return of 12.01%. This evidence tends to confirm Levy’s earlier work. However, it was not available when the test model was begun. Instead, both List 1 and List 2 used a derivation of Levy’s original calculations.

Levy originally calculated the ratio of a stock’s 131-day moving average to its latest price. This ratio was calculated for all stocks. The total list of ratios was then sorted. Each stock was allocated a relative price strength percentile between 99 and 0 based on where its ratio fell in the spectrum of ratios. The 0 percentile for the highest ratio (weakest stock) and the 99th percentile for the lowest ratio (strongest stock).

To make the ratio easier to calculate and to understand, the test lists changed several aspects of Levy’s calculation but not the essence. Rather than using the ratio of the moving average to the current price, the inverse was used. The ratio of current price to the moving average made the high percentiles represent the highest relative strength. Thus the 99th percentile represented the strongest stock and the 0 percentile represented the weakest. Second, instead of 131 days of data in the moving average, the test lists used 26 weeks, approximately the same period (131 trading days is 26.2 weeks, not including holidays). In this manner, a large amount of data was not necessary (131 data points per stock versus 26 data points), yet the resulting ratios were equivalent and the effects on the post-performance minimal. The closing price used each week was the Thursday close.

Relative Earnings Growth: Until this point, it would appear that the study was involved solely with technical analysis and price behavior. However, while technical analysis has its weak and strong points, a stock selection method must use all variables that appear to work. Relative earnings growth is one of them.
To a certain extent, “earnings” are a manufactured statistic. They depend on many accounting tricks and are not always truthful measures of a company’s success or failure. Special charges are often later written off against earnings, and depreciation is recalculated, or taxes reassessed. Reported earnings, therefore, are often subject to controversy and exaggeration. No one can argue that a stock closed at a certain price (at least within some small bound), but analysts often disagree on exactly what a company’s actual earnings may be. This becomes even more complicated when earnings are estimated into the future. However, earnings reports are watched, especially for surprises, and are acted upon by investors. Tests have shown that reported relative earnings growth has a positive correlation to the post-performance of a stock. Part of this, of course, is because reported earnings include any earnings surprise.

To be as sensitive as possible without the effect of seasonality, Levy calculated earnings growth by taking the most recent five quarters of reported earnings and measuring the ratio between the latest four quarters total to the first four quarters total. Thus three quarters overlapped, and the seasonal tendency of many quarterly reports was eliminated. The ratio, if positive, showed that earnings were growing and by how much and if negative that earnings growth was negative and by how much. When companies reported losses for any consecutive four quarters, the ratio was not calculated. Growth then was measured over a relatively short period of five quarters. This same calculation was used in determining the earnings growth criteria for both test lists. As with relative price strength, the ratio for each stock was ranked with the same ratio for all other stocks and a percentile ranking determined whereby those stocks with the highest earnings growth were ranked in the highest percentiles, and vice versa for those with the lowest earnings growth.
**Chart Pattern:** As mentioned above, the test required some means to reduce the risk of an individual stock’s failure. List 1, which was the only one to use a chart pattern, by its nature, was very volatile and its selected stocks very high in price and valuation. This is only natural when stocks with high relative strength and high earnings growth are selected. To reduce the danger of individual collapse, the use of a simple chart pattern was thought to be the best method at the time to eliminate those stocks that begin to decline severely and before they collapsed.

Computerizing chart patterns, especially twenty years ago, was and still is a difficult problem. The simplest method was to produce a simple point-and-figure chart, one that shows only price reversal points after a predefined price magnitude has occurred. To do this, only the magnitude of the price move was needed to determine the reversal point. As an example, in many point-and-figure charts, a three point reversal magnitude is required for a price reversal point. If a stock price rises from 50 to 56, then declines to 48, since the three points up and down have been met, the reversal point was 56, the highest point at which the stock price had risen by at least 3 points and reversed by 3 points. This would be called an “upper reversal point” since it marked a top in prices. Had the stock only risen to 52 before declining to 48, no reversal point would have been recorded since the stock had not risen from 50 by the required magnitude of 3 needed to establish a reversal point. Conversely, had the stock then declined to 48 and risen back to 55, the price of 48 would have been a “lower reversal point” since the stock had declined by at least 3 into 48 and then risen more than the required 3 immediately afterward. This combined behavior would then have left us with a history of an upper reversal point at 56 and a lower reversal point at 48. In the chart formula, the last two upper and lower reversal points were recorded each week. When prices rose above two upper reversal points, the chart was said to be “advancing,” and prices declined below two lower reversal points, the chart was said to be “declining.”
In addition, in the chart formula, a sliding scale of reversal magnitudes was established to minimize the effect of absolute price differences. For example, a 3-point reversal in a 100 dollar stock is less significant than a 3-point reversal in a 20 dollar stock. A sliding scale of reversal magnitudes equalized the requirements for a reversal among all stocks.

Rather than be concerned about the actual patterns of the reversal points, List 1 only used the reversal points themselves. Only those advancing stocks were considered for selection, and those stocks in the list that turned down below two lower reversal points were eliminated. This provided the ‘stop’ needed to protect the portfolio from extraordinary negative events.

**Relative Price/Sales Ratio (“PSR”):** Prices are well-known and easily accepted as valid. Annual sales of a company are also well-known and easily accepted as valid, and when combined with prices are an excellent comparative measure of a stock’s value. The higher the price-to-sales ratio, the higher the valuation that investors have placed on the stock’s future, and also the higher the risk of failure. Lower PSRs suggest lower value placed on a stock’s future. Their advantage is that “a small improvement in profit margins can bring a lot to the bottom line, improving the firm’s future P/E. Low PSR stocks are held in low regard by Wall Street. Those with improving profit margins usually catch the Street by surprise.”\(^\text{15}\) PSRs also include stocks with no earnings (and therefore no P/E). Many studies have shown the value of the PSR.\(^\text{16}\) O’Shaughnessy (1998) argues that the PSR is the most reliable method of selecting stock for long term appreciation.\(^\text{17}\) His method of using the PSR, however, requires that an arbitrary level be established, below which a stock is attractive. In List 2 the arbitrary level was disbanded in favor of a relative percentile. First, the ratio was calculated for each stock as the current weekly close price divided by the last reported four-quarters sales. Next, this ratio for all stocks was then sorted and divided...
into percentiles such that the highest was in the 99th percentile and the lowest in the 0 percentile. This way, despite the general market level of valuation, a stock’s PSR could be measured against the PSR of all other stocks at the same time and in the same investment environment.

**Combining Criteria into model - The Parameters**

Each week the entire list of available U.S. stocks (usually around 5,000) was screened for those stocks at or above the 90th percentile in relative price and earnings growth. In List 1 an advancing chart pattern was also required. Any stock not already on the list that met these criteria was added to the list. When relative price strength declined to or below the 30th percentile, relative earnings growth declined to or below the top 80th percentile, or the stock price pattern broke two previous lower reversal points, the stock was eliminated from the list. In List 2, the chart pattern was not used, but relative PSR was. The requirement for addition to the list was a relative PSR at or below the 30th percentile. The deletion criteria in List 2 were the same as in List 1 except they did not include the relative PSR since a high level did not necessarily suggest that a stock was facing an impending decline. Additionally, the deletion requirement for relative earnings growth was reduced to or below the 50th percentile since earlier experience had shown that a high threshold deleted stocks prematurely.

**Specific Results**

Chart III shows the performance of List 1, the S&P 500 and the Value Line Geometric each year since the inception of the study in 1982. Chart IV shows the more recent total history for List 2 versus List 1, the S&P 500 and the Value Line Geometric.
List 1, which began its weekly live trial in July 1982, gained a total of 5086.6% over the 17 ½ years versus a 1087.6% gain in the S&P 500 and a 221.9% gain in the Value Line Geometric. This gain was 4.37 times the gain in the S&P and 16.11 times the performance of the Value Line Geometric. During that 17 ½ year period List 1 had only three down years versus three for the S&P 500 and seven for the Value Line Geometric (see Table A below).

List-2, which began it weekly live trial in January 1999, has had only two years of history to measure. Nevertheless, the results so far have been impressive. Over the two years, the list gained 137.3% versus only a 7.41% gain in the S&P 500 and a 9.99% loss in the Value Line

<table>
<thead>
<tr>
<th></th>
<th>List-1</th>
<th>S&amp;P</th>
<th>VALG</th>
<th>List-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>49.1%</td>
<td>26.5%</td>
<td>29.2%</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>57.6%</td>
<td>17.3%</td>
<td>22.4%</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>-11.4%</td>
<td>0.9%</td>
<td>-9.3%</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>33.3%</td>
<td>26.1%</td>
<td>19.2%</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>20.9%</td>
<td>17.0%</td>
<td>8.1%</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>11.5%</td>
<td>0.1%</td>
<td>-11.6%</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>22.6%</td>
<td>13.1%</td>
<td>14.8%</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>26.5%</td>
<td>25.5%</td>
<td>10.8%</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>-12.1%</td>
<td>-6.4%</td>
<td>-24.0%</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>26.0%</td>
<td>23.3%</td>
<td>23.3%</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>19.4%</td>
<td>7.6%</td>
<td>11.0%</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>25.7%</td>
<td>7.6%</td>
<td>10.4%</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>-1.6%</td>
<td>-1.6%</td>
<td>-6.3%</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>54.5%</td>
<td>33.2%</td>
<td>19.3%</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>24.5%</td>
<td>23.1%</td>
<td>13.8%</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>8.2%</td>
<td>23.4%</td>
<td>17.2%</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>5.5%</td>
<td>31.8%</td>
<td>-0.5%</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>62.4%</td>
<td>19.1%</td>
<td>-2.6%</td>
<td>59.8%</td>
</tr>
<tr>
<td>00</td>
<td>7.9%</td>
<td>-9.6%</td>
<td>-7.6%</td>
<td>48.5%</td>
</tr>
</tbody>
</table>

**TABLE A**
Geometric. It had no down years versus one for the S&P and both for the Value Line, and as mention earlier, its beta and turnover were considerably lower than List-1.

Conclusion

The quantitative analysis of stock selection criteria has diverged in many directions since the relatively recent widespread use of the computer. Most analysis has centered on demonstrating the validity of one or more specific stock market theories and many have shown mediocre results. The method of testing these results has also fallen into the optimization trap whereby the “best fit” between data and performance was not tested with new data and especially with unknown future data.

This study took several variables that had been demonstrated to have value in stock selection and in one list, beginning in July 1982, tested the results “live” each week for 17 ½ years. The test was done through simulating the performance of a hypothetical portfolio, thus adding an element of practicality not seen in most studies of stock selection, and used a combination of technical and fundamental factors without prejudice. These factors measured aspects of a company or its stock on a basis relative to all other stocks and were independent of general market averages except in the demonstration of performance. The results were exceptionally favorable for the methods used and demonstrated the usefulness of the variables employed. Relative price strength and relative reported earnings growth, when calculated in the manner of this study, showed superior results when compared to market averages. Since the period over which the study was done was one of generally rising stock prices, the final test will be completed only after a major stock market decline. However, considering the long period over which the study was
conducted without adjustment for market changes, the presumption is that the relative post-performance results of the methods used will continue to exceed average market returns.

Charles D. Kirkpatrick II, CMT
Footnotes

1 Avanian and Wubbels (1983)

2 Levy (1968b) - lengthy and doesn’t add much more than Levy (1967)

3 Levy (1967) - This article caused quite a stir in academia because it was the first major attempt to refute the efficient market hypothesis.

4 Levy (1966)

5 Bernard (1984), the founder of Value Line, as an example, had successfully utilized the concepts of relative price strength and relative earnings growth since the late 1920’s. “dividing the stock’s latest 10-week average relative performance by its 52-week average relative price” is the price momentum factor used by Value Line. For a recent discussion of the merits of the Value Line system see Choi (2000).

6 The reaction to Levy’s (1967) article was swift. Michael Jensen (1967) of Harvard was the first to publish comments. Initially he criticized Levy’s methodology on the basis that the sample was too small and over too short a period, had a selection bias, and other errors that would overstate the results. His comment was that Levy’s comment of “the theory of random walks has been refuted” was a little too strong. Levy (1968c) then countered with another study including more stocks and a longer time period that produced even better results (31% versus the market 10% for 625 stocks from July 1, 1962 to November 25, 1966). Finally Jensen and Bennington (1970) did their own study, supposedly using Levy’s rules but including transaction costs and adjustments for risk, and using 1962 stocks from 1926 to 1966, and reported that Levy’s rules resulted in a risk-adjusted loss. We never heard of Levy’s relative strength work again.

7 see Murphy (1986) and Kaufman (1978)

8 Kirkpatrick (1978 - 2001)

9 Merrill (1977)

10 The entire concept of past price returns having an effect on future price returns has academia in quandary since it tends to cast severe doubt on the efficient market hypothesis. Many different price return anomalies have been reported, some positive and some negative. Long-term and very short-term results tend to be consistently negative. Chopra, Lokonishok, and Ritter (1992), Cutler, Poterba and Summers (1988), De Bondt and Thaler (1985), and Fama and French (1986) show that for holding periods beyond 3 years, the return is negative. Over periods of a month or less, French and Roll (1986) and Lehmann (1990) found negative returns in individual stocks weekly and daily; Lo and MacKinlay (1990) found positive returns weekly in indices and portfolios but negative returns for individual stocks; and Rosenberg, Reid and Lanstein (1985) found negative reversals after a month. There seems, however, to be a window of about six to twelve
months when returns are consistently positive. This was Levy’s hypothesis and it has now been confirmed by Brush (1983, 1986) and Jegadeesh and Titman (1993). BARRA [see Buckley (1994)] has found the price momentum anomaly in a number of countries, including the US, Japan, the UK, Australia, and France. Explanations for these anomalies are varied but best summed in Chan, Jegadeesh and Lokonishok (1996, 1999).

11 The question of how accurate are reported earnings and especially how accurate are future earnings forecasts has been widely studied. Niederhoffer (1972) and Cragg and Malkiel (1968) suggest that reported earnings are better forecasters of future earnings than analysts forecasts. Indeed, Harris () concludes that analyst forecasting accuracy is extremely poor, biased and inefficient. The inaccuracy is mostly the result of random error and the performance of forecasts vary with both the company characteristics and the forecast itself. A whole series of studies has evolved around “earnings surprises” those frequent events when reported earnings differ markedly from analysts’ expectations. La Porta (1996) has shown that superior results can be gained by exploiting these analyst errors because expectations are too extreme. Investors overweight the past and extrapolate too far into the future. Chan, Jegadeesh and Lokonishok (1996, 1999) speculate that the reason for the relative strength positive anomaly over six to twelve months is that it takes that long for the analysts to adjust. La Porta (1996) suggests that it takes several years.

12 Ramakrishnam and Thomas (1998)

13 Levy and Kripotos (1968a)

14 The best and most recent discussion about analyzing chart patterns is in Lo, Mamaysky, Wang, and Jegadeesh (2000).

15 Fisher (1996)

16 A number of financial ratios have been used and tested. The most common, of course, is the price-to-earnings ratio (PER). More recently the market-to-book ratio has become popular, and even more recently attention has returned to the price-to-sales ratio (PSR). Senchack and Martin (1987) had shown that low PSR stocks tended to outperform high PSR stocks but that low PER stocks dominated low PSR stocks on both an absolute and risk-adjusted basis. But recently Barbee (1995) showed in tests from 1979 to 1991 that price-to-sales and debt-to-equity had greater explanatory power for stock returns than did either market-to-book or market-to-equity. Liao (1995) also showed that low PSR stocks avoid the ambiguities of the CAPM approach and dominate high PSR stocks and the market.

17 O’Shaughnessy (1998) also argues for relative price strength as a selection criterion.
References


Kirkpatrick, Charles D., II, *Kirkpatrick’s Market Strategist* (Post Office Box 699 Chatham MA 02633)


