An Attempted Replication of the Results from: Inducing Stock Screening Rules for Portfolio Construction by Tam, Kiang, and Chi.

John L. Weatherwax

October 18, 2006

1 Introduction

This paper was started as an attempted replication of the results reported on in the paper [1]. In that paper it is claimed that using a classification tree to select *high* growth stocks can yield returns that in many cases statistically outperform the NYSE and the S&P-500 composite. The classification tree is built using publically available fundamental and technical data. A high level definition of the specific fundamental and technical measures used to determine high growth stocks is presented in the paper and will be discussed below. The authors of this paper attempted to download pertinent data from Compustat, compute the required features, build a classification tree, and perform the required backtesting to verify that the technique described in the paper would indeed outperform the indices specified, at least for the time frames claimed in the paper. The results from these exercises were mostly failures. This failure could be the result of several things. Most notably incorrect computation of the key technical and fundamental company indicators. It should be mentioned that before commencing this study, none of the above authors had detailed knowledge of the fields available in the Compustat database or many of the financial definitions required to compute the needed fundamental and technical indicators. The purpose of this paper is to report *explicitly* how the features denoted as important in [1] were calculated from the base fields in the Compustat database. The authors hope that someone with more financial accounting knowledge could either, find errors in the definitions computed below (and offer corrections) or offer reassuring that the measures of company growth are computed correctly.

2 The Important Stock Indicators

Below, I discuss in detail the data extraction performed as I attempted to reproduce the results from the paper [1]. As stated earlier, at a very high level I was unable to reproduce the results presented there, perhaps due to the formulas I used in computing the features that would be selected as distinguishing "good" stocks (ones that doubled in price over a years period) from "bad" stocks. The features selected from this paper for classification of the two stock types are given by the features below. Very simple definitions of these fundamental measurements of a companies growth potentials were given in the paper but no specific information on *how* they were calculated was provided. The features suggested as indicators of high growth companies (and the verbatim description given in the paper) are:

- **Common Shares Outstanding:** No definition given.
- **Price Ratio:** The ratio of the quarter's closing price to the highest price in the previous two years.
- **Return on Equity (ROE):** Net income divided by the stockholders shareholders equity.
- **Net Current Asset Value:** The net current assets per common share outstanding.
- **Price-to-Book Ratio:** The ratio of the market value of an equity to its book value. **Note 1:** The use of the total assets rather than the *current* total assets might make this a positive number, since calculated as we have it, currently it will have the same sign as the feature net current asset value.
- **Relative Strength:** The weighted average of the quarterly price changes during the previous year, where the last quarter has a weight of 40%, and the other three quarters each have a weight 20%.
- **Change in Quarterly Earnings:** The change in quarterly earnings between -2 and -1 quarters, *measured as a percentage*. This is obtained from Compustat field DATA58, retained earnings.
- **Market capitalization:** The number of common shares outstanding multiplied by the stocks closing price at the quarter end.

3 Calculation from Compustat data

Base fields from the Compustat database were then used to compute the quantities above. The specific formula and Compustat fields used are given as:

- **Common Shares Outstanding:** Obtained directly from the Compustat database field DATA61. This feature is in units of millions.

- **Price Ratio:** The ratio of the quarter's closing price to the highest price in the previous two years. Obtained from the a time history of the field DATA14 in the Compustat database. This is a unitless feature.

- **Return on Equity (ROE):** Net income divided by the stockholders shareholders equity. I assumed that the net income would be the pretax income. The pretax income is obtained from the Compustat field DATA23, while the total shareholders equity is obtained from DATA60, their ratio gives the feature desired in units of dollars.

- **Net Current Asset Value:** The net current assets per common share outstanding. The current assets of a company are the current total assets minus current total liabilities. The current total assets is given by Compustat field DATA40, while the current total liabilities is given by Compustat field DATA49. Common shares outstanding is given by field DATA61. Their ratio is the feature net current asset value. This feature has units of dollars. Note I have assumed that we are using *current* total assets in this calculation rather than just total assets (the latter incorporates more long term growth potentials) of the asset this is probably a good assumption since the term current appears in the features name.

- **Price-to-Book Ratio:**

  The ratio of the market value of an equity to its book value. The market value of an equity is its price at quarter close price obtained from the Compustat field DATA14. The book value of an equity requires some calculation. The book value of an equity is calculated as the total current assets minus total current liabilities. This piece of data was required in calculating net current asset. The book value per share is then simply the book value divided by the number of shares outstanding. Since the number of shares outstanding is given by Compustat field DATA61, the book value per share (BVPS) is calculated then as

$$\text{BVPS} = \frac{\text{DATA40} - \text{DATA49}}{\text{DATA61}} \,. \tag{1}$$

  So that the total formula for the price to book ratio (PBR) is given by (in terms of the Compustat database fields)

$$\text{PBR} = \frac{\text{DATA14}\,\text{DATA61}}{\text{DATA40} - \text{DATA49}} \,. \tag{2}$$

  This produces a unitless feature. Note I have assumed that we are using *current* total assets in this calculation rather than just total assets. The latter incorporates more long term growth potentials of the company. This may not be a good assumption if this indicator is to indicate long term growth potential.

  Note 1: The use of the total assets rather than the *current* total assets might make this a positive number, since calculated as we have it, currently it will have the same sign as the feature net current asset value.

  A company's common stock equity as it appears on a balance sheet, equal to total assets minus liabilities, preferred stock, and intangible assets such as goodwill.

  Note: it has recently be brought to my attention that this is defined as "Market to Book" in the Compustat lingo. Page 31 of the third chapter of the Compustat users guide has more information.

- **Relative Strength:** The weighted average of the quarterly price changes a during the previous year, where the last quarter has a weight of 40%, and the other three quarters each have a weight 20%. I'm assuming that this should be *measured as a percentage*. The quarterly prices are obtained from a time series of the field DATA14 from the Compustat database. This is a unitless feature.

- **Change in Quarterly Earnings:** The change in quarterly earnings between -2 and -1 quarters. I'm again assuming that this should be *measured as a percentage*. This is obtained from Compustat field DATA58, retained earnings.

- **Market capitalization:** The number of common shares outstanding multiplied by the stocks closing price at the quarter end. The number of common shares outstanding is given by Compustat field DATA61, while the stocks closing price is given by DATA14.

- **Comp stat Company Index Number:**

Note I am assuming that the feature *relative strength* and *change in quarterly earnings* are measured in a relative fashion, i.e. if the quarter closing price for a given equity at time index i is given by P_i, then the relative strength at time index i is defined as

$$\mathrm{RS}_i = 0.4\left(\frac{P_i - P_{i-1}}{P_{i-1}}\right) + .2\left(\frac{P_{i-1} - P_{i-2}}{P_{i-2}}\right) + .2\left(\frac{P_{i-2} - P_{i-3}}{P_{i-3}}\right) + .2\left(\frac{P_{i-3} - P_{i-4}}{P_{i-4}}\right). \quad (3)$$

As something to try in the future, it might be advantageous to measure each quarters *return* (defined by $(P_i - P_{i-1})/P_{i-1}$) in decibels rather than a linear scale as is done above.

4 Conclusion

When the features suggested in [1] are calculated according to the above rules, the features obtained are do *not* separate the two classes (stocks that doubled v.s. those that did not) in any statistically significant way. In other words, there is no clear way using the information provided to design an algorithm that will provide any "insight" over random chance. Again, as stated above this may in fact be due to errors in my interpretation of the Compustat database fields and or the formulas used to derive the fundamental and technical indicators. I would very much appreciate any comments (at wax@alum.mit.edu) on correctness or incorrectness of my above implementations from anyone with more knowledge of financial accounting than myself. In addition, questions are always welcome.

References

[1] K. Y. Tam, M. Y. Kiang, and R. T. H. Chi. Inducing stock screening rules for portfolio construction. *Journal of the Operational Research Society*, 42:747–756, 1991.