Decision Trees and Financial Variables

Roy Rada, Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA
Hayden Wimmer, Department Information Technology, Georgia Southern University, Statesboro, GA, USA

ABSTRACT

A decision tree program for forecasting stock performance is applied to Compustat’s Global financial statement data augmented with International Monetary Fund data. The hypothesis is that certain Compustat variables will be most used by the decision tree program and will provide insight as to how to make investing decisions. Surprisingly, the authors’ experiments show that the most frequently used variables come from the International Monetary Fund and that variables provided exclusively for Financial Industry stocks were not useful for forecasting financial stock performance. These experiments might be part of a constellation of such experiments that help people map financial forecasting problems to the variables most useful for solving those problems. The research shows the value of using decision tree methodologies as applied to finance.

KEYWORDS

Fundamental Financial Analysis, Global Stocks, ID3 Macroeconomic Data, Splitting Variables

INTRODUCTION

One of the fascinating problems in accounting and finance is how to predict which stocks will garner above average returns. For this problem, what forecasting model should be applied to what financial variables? Since decision trees can generate rules which are understandable to people and can be derived automatically from historical data, decision trees have been chosen as the forecasting model in this study.

The authors apply a decision tree algorithm to hundreds of financial variables and their experiments show that the variables most frequently used by the decision tree are different from those expected. The finance domain could benefit from further mapping of forecasting models (such as decision trees) to financial variables (such as Total Assets on the Balance Sheet) for solving particular stock valuation problems (such as selecting high performing global stocks in the year 2005). The literature shows various financial variables being useful for various purposes. However, more study is needed of which variables to use when. This paper asks whether decision tree methodologies can be helpful in solving this financial problem.

LITERATURE REVIEW

Decision trees in the management sciences are used to model the likely consequences of various decisions for a particular problem domain (Corner & Kirkwood, 1991). Within artificial intelligence, a decision tree algorithm is a classification method where the induction process “constructs a flowchart-like structure where each internal (non-leaf) node denotes a test on an attribute, each branch corresponds to an outcome of the test, and each external (leaf) node denotes a class prediction” (Han & Kamber, 2006). The decision tree models will often “help users to investigate how each feature weighs in the decision-making process and uncover the logical relations between the features and
classified results” (Lin, Wang, & Chung, 2010). For instance, in an effort to improve the efficiency of call centers, researchers have created a decision tree to help identify the factors that were most crucial in influencing the performance of the call center (Kyper, Douglas, & Blake, 2012).

The value of automated decision trees for financial forecasting has been established (Tsang, Yung, & Li, 2004). Roko and Gilli (2008) feed fundamental and technical ratios to a decision tree algorithm to forecast high earning stocks. Given the Adaptive Market Model, one might want decision trees to adapt over time, and research has shown how to evolve decision trees (Fu, Golden, & Lele, 2003). Of course, a robust approach to multi-attribute decision making in finance might use more than decision trees, and other methods, such as fuzzy correlation rule mining (Robinson & Amirtharaj, 2014) or interpolative Boolean algebra (Mandic, Delibasic, & Radojevic, 2015), could be applied.

For decades researchers have been identifying patterns in accounting data and predicting stock returns by exploiting those patterns (Richardson, Tuna, & Wysocki, 2010). Some research shows that investors make false assumptions about companies with extreme book-to-market (BM) values and with large accruals. To earn above average returns, Piotroski (2000) first identifies high BM stocks and then ones with strong financial fundamentals. Sloan (1996) showed that stock prices do not adequately reflect extraordinary items in the accruals and said, “stock prices are found to act as if investors ‘fixate’ on earnings, failing to reflect fully information contained in the accrual and cash flow components of current earnings until that information impacts future earnings.” Galdi and Hermesmeyer (2010) apply Piotroski’s method to stocks in Brazil but find it prudent to put different weights on the independent financial variables. High BM stocks are considered ‘value’ stocks because they are deemed to be undervalued relative to their assets, while low BM stocks are considered ‘growth’ stocks. To successfully forecast which stocks will be high-earners, Mohanram (2005) first screens for ‘growth’ stocks and then identifies which independent financial variables can be used to forecast high earners among the growth stocks.

Beneish et al (2001) emphasize the role of context in the selection of financial independent variables to be fed to a forecasting algorithm. For instance, Piotroski found one set of independent financial variables worked for high BM stocks, but Mohanram found a different set of independent financial variables worked for low BM stocks. Extreme performers can be characterized by different formulas than can non-extreme performers, and, generally speaking, what best models performance of one category of stocks is different from what best models performance of another category of stocks. Beneish et al argue that given a category of stocks, a unique formula might best predict how stocks in that category will perform.

The literature (Green, Hand et al. 2011) suggests that published methods to exploit anomalies are subsequently used by market players until the method ceases to work. Thus, the independent financial variables that are best for forecasting stock values vary across time. The Adaptive Market Model accounts for this evolutionary character of markets (Lo, 2004). In addition to being sensitive to time, the variables that might influence stock prices are many, and another approach to the problem presented here would be to compartmentalize the variables into knowledge marts (Wimmer, Forgionne, Rada, & Yoon, 2012).

**METHOD FOR DATA**

Standard & Poor’s is one of the premiere sources of fundamental financial data for stocks. One of the prominent products of Standard & Poor’s is called Compustat Global (Standard & Poor’s, 2011) and was the product used for this study. Compustat Global can be accessed through various portals and the Wharton Research Data Services (WRDS) was the portal used for this study (Glushkov & Moussawi, 2009). Financial statement data for twenty years for all companies was downloaded from WRDS.

Thirty-six financial statement variables were used (see Appendix “Financial Statement”). The variables were chosen directly from common financial statement attributes that are publically available. They were identified as adequately covering the essential components of the financial statement.
Alignment Of Organizational Strategy With Information Technology Strategy
Tamio Shimizu, Marley Monteiro de Carvalho and Fernando Jose Barbin (2006).
www.igi-global.com/chapter/alignment-organizational-strategy-information-technology/29709?camid=4v1a

Non-Parametric Statistical Analysis of Rare Events in Healthcare: Case of Histological Outcome of Kidney Transplantation
Soheila Nasiri and Bijan Raahemi (2017). Handbook of Research on Data Science for Effective Healthcare Practice and Administration (pp. 198-212).
www.igi-global.com/chapter/non-parametric-statistical-analysis-of-rare-events-in-healthcare/186938?camid=4v1a