Federated Hermes MDT Advisers

Federated Hermes

Classification and regression trees: A deliberate approach to stock selection

- Regression trees are a series of yes/no questions based on explanatory (independent) variables that lead to a prediction.
- The MDT investment team uses regression trees to predict individual stock performance versus the universe.
- We arrive at our predictions by asking questions about fundamental and technical company characteristics.
- By asking a series of questions in sequence, regression trees allow us to focus on the most important characteristics for a particular stock, and downplay those characteristics that are less important.

Overview

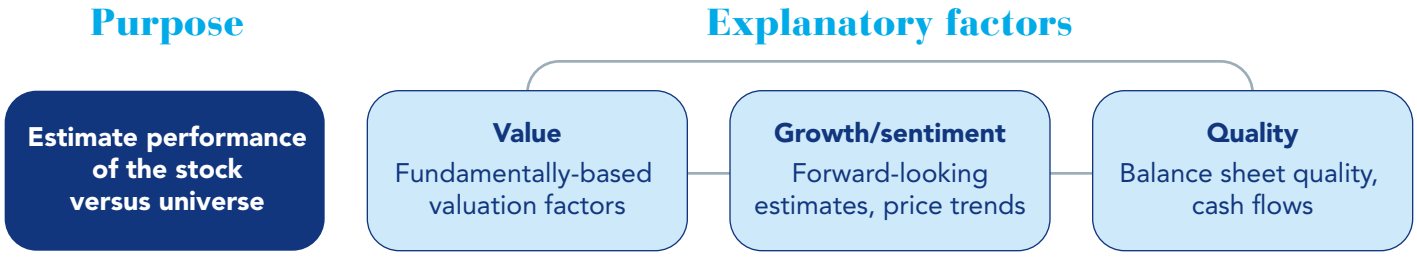
This paper is the first in a new series to review some of the ideas driving our investment process. It focuses on an important differentiator: regression trees.

The formal academic background for classification and regression trees came out of Stanford University research by Professor Leo Breiman (Jerome Friedman, Charles J. Stone, and R.A. Olshen), who published a book entitled "Classification and Regression Trees" in 1984.[1] Their methodology for building regression trees and the software they provided to create them became a standard tool in the insurance industry and the physical sciences. The difference between regression trees and the more commonly known decision trees is simple: regression trees predict a number, while decision trees predict an endpoint state or "classification." We focus on regression trees in this paper, as we use them in an effort to predict how much a stock will outperform or underperform its universe.

Classification and regression trees (CART) background

Regression trees are a series of yes/no questions chosen using an algorithm: given the measure to be determined and a substantial dataset holding a number of possible explanatory variables, what is the best question to ask to arrive at a more accurate estimate of the target value? The algorithm figures out the best questions to ask at each point by rote. It tests every possible question (every explanatory variable, every point at which you can split the dataset into observations with a higher value and lower value) and chooses the one that yields two subgroups with different average values of the target measure and a minimum sum of squared errors versus those new averages. The algorithm uses the same process to choose the next best question for every subgroup.

MDT regression trees



[1] Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8.

Using regression trees in an effort to predict stock alphas*

In 2000, MDT began a program of research based on the CART technique. We believed that CART might work well for selecting stocks, as we liked the non-linear nature of the analysis — the fact that a regression tree didn't allow a characteristic that wasn't important to a specific company to affect the outcome. We evaluated the technique with back tests and saw that the regression trees improved the results, so we added trees to our live strategies in 2001.

We provide an example tree here to show how the regression trees can be used to predict alpha. This is a very small tree, but it illustrates the concepts and advantages of using a tree for this purpose. This illustration does not represent any of the regression trees in our strategies.

In this tree, the first question is whether or not the company has recently been a net issuer of equity or debt. On average, companies that use a substantial amount of external financing tend to underperform. Note that the question chosen doesn't have to split the data 50/50. In the example tree, it is a minority of companies (Group 5) that use a substantial amount of external financing, but it provides a strong signal that (again, on average) those companies are likely to underperform.

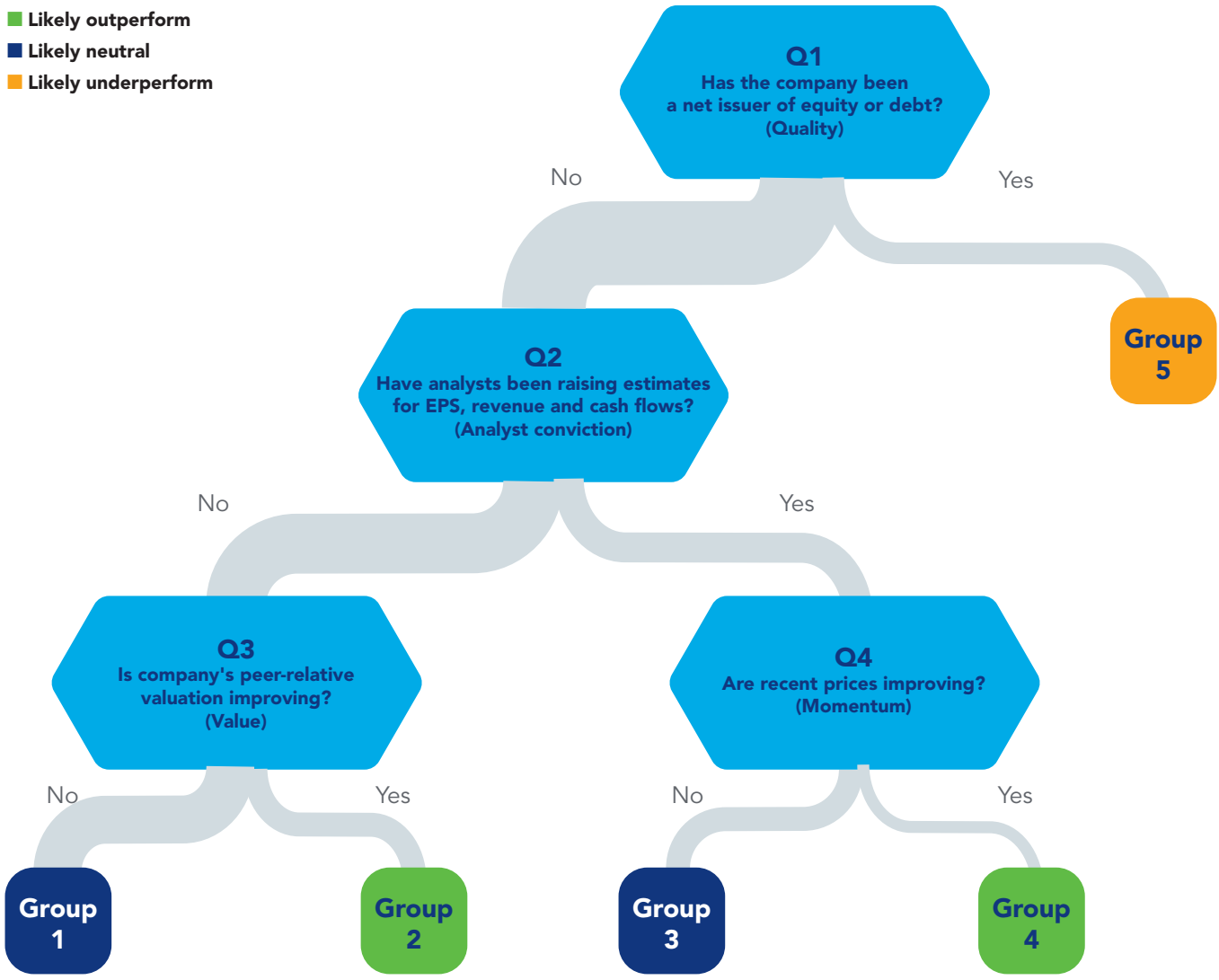
*Alpha in this document refers to a stock's excess return versus a strategy's universe.

The second question in this sample tree is about sell-side analyst conviction. This question is asked of only the companies that answered "No" to the first question, so we already know that these companies aren't relying on excessive financing. Next, the algorithm determines the best question to ask these specific companies, and the question it finds is "Have analysts been raising estimates for earnings-per-share, revenue and cash flows?". That question we have found is most useful for growth-oriented companies. The question divides the remaining data observations into two smaller groups, with somewhat more companies answering that question with "No" than "Yes."

At this point, we know more about the companies going into questions three and four. The companies going into question three did not have high analyst conviction, and the algorithm finds that the best next question for those companies is about a value-oriented variable. The companies going into question four do have high sell-side analyst conviction, and the best next question is about momentum, another factor that tends to do well for growth-oriented companies.

After each company has answered the relevant questions in this simple example tree, we now have five groups of companies with differentiated alpha estimates.

Using CART to select companies



Classification and regression trees

Advantages of using a regression tree to predict stock alpha:

Versus a more traditional/linear approach

- Regression trees sift through a vast amount of data to find companies with combinations of characteristics that have foreshadowed price movements relative to the universe over its history.
- Regression trees use explanatory variables non-linearly. A high value of one explanatory variable may be suitable for one company but bad for another with different characteristics.
- Only the questions relevant to each company are asked. If a company is a value company, the algorithm doesn't waste time asking questions that are more relevant to growth companies.
- Highly scored companies won't all have the same characteristics. In the sample tree shown above, both Groups 2 and 4 are predicted to outperform. Group 2 has more value-oriented names, while Group 4 has more growth-oriented names with substantial price momentum. That makes it easier to build a portfolio with better risk characteristics than if you had high-scored companies from a linear model where the companies with the highest scores had similar characteristics.
- From anecdotal evidence, few investment shops make regression trees a central part of the investment process. We believe that our trades are less crowded than those of other managers.

Versus other machine learning techniques

- Regression trees are transparent. It is easy to understand why companies get high scores by looking at their values of the explanatory variables. That means it is easy for us to understand why or why not a company gets a high score; and easy for us to review the model's daily trades and understand why they are being made — a valuable quality control.
- Regression trees are relatively easy to build; there are a modest number of parameters to be specified and what those parameters do is fairly intuitive.
- Regression trees are robust to input data. Outliers aren't a problem as all that matters is which companies are above or below the split point. The data doesn't need to be normalized.

Key takeaways

The MDT investment team has made regression trees central to the investment process. They provide a daily estimate of every company's likely performance versus the universe based on the most recent market prices, financial reports and sell-side analyst estimates.

Regression trees:

- Sift through a vast amount of data to find companies with combinations of characteristics that have foreshadowed price movements relative to the universe over its history.
- Provide much more information from stock selection characteristics than standard regression models — they are non-linear and can focus on a particular range within a meaningful characteristic and ignore the rest of the values of that characteristic.
- Allow a customized analysis for each stock — without any qualitative input, the regression tree focuses the analysis on the most meaningful characteristics for a particular stock.

MDT key investment team members

Daniel Mahr, CFA
Managing Director, Research

Sarah Stahl, CIPM
Managing Director

Frederick Konopka, CFA
Portfolio and Trading Manager

John Paul Lewicke
Research Manager

Damien Zhang, CFA
Research Manager

Tony Ng, CFA
Research Manager

Kelly Patel, CFA
Senior Analyst

Keeva Walker
Senior Analyst

Katherine Silva
Analyst

Tyler Piazza
Associate Analyst

Michael Bertani
Assistant Portfolio and Trading Manager

David Gomez
Trading and Performance Analyst

There is no guarantee that the use of regression trees will be a successful investment approach.

The quantitative models and analysis used by MDT may perform differently than expected and negatively affect performance.

Investing in equities is speculative and involves substantial risks.

50783 (11/22)
MDT Advisers, a Federated Advisory Company