**Using Supervised Machine Learning to Construct Optimal Corporate High Yield Bond Portfolios**

**Goal:** To build a build a fundamental relative value model using supervised machine learning techniques. The supervised machine learning algorithm will assign a credit risk score to each high yield corporate bond issuer and rank the issuer cohort. All of the issuer rankings and weighted average option adjusted spreads (i.e., the issuers implied credit risk that is being priced in by the market) will be plotted and used to form a liner regression of the two. Issuers whose ranking and option adjusted spread point are above/below the liner regression should be sold/bought.

**Subject Material:**

High Yield corporate bond managers who use a “bottoms-up” approach to investing rely on buying or selling high yield corporate bonds that trade at a price below or above a credit’s intrinsic value based on the issuers credit risk. The yield, intrinsic value, of a high yield bond is determined by discounting its expected future cash flows to its present par value. Assuming the price of a given high yield corporate bond is above or below its intrinsic value, the fundamental analyst can buy or sell the associated bond and, assuming the issuer does not default, profit from the spread between the two. However, as my colleagues Michael Kimble and Jakob Bak astutely pointed out a decade ago, the prevailing model for thinking about corporate bonds has been derived from strategies that have been effective for equity markets, yet it is profoundly true that bonds are different from equities and behave differently. Equities possess unlimited upside potential, so picking just a few winners can offset a significant number of losers. Historically, the majority of trailing twelve month total return distributions for equities have been positively skewed. Therefore, conventional methods have a good chance of success. High Yields bonds present a fundamentally different situation. Upside potential is constrained as bond prices approach par. Rare is the bond that produces a better than 20% total return relative to the rest of its cohort. Losers of that magnitude are a much more common occurrence. Due to its fundamental structure, for any given point in time the trailing twelve month total return distribution is negatively skewed and leptokurtic (as shown in the historical price distribution in figure 1). As such, the bonds you do not own are more important than the ones you do.



The goal of this project will be to develop a new approach to bond investment strategy by training a number of supervised machine learning algorithms to rank the credit risk of a given high yield issuer relative to its cohort. For issuers whose financial statements are not public, I will assume that the market is efficiently capturing the credit risk of the issuer and use its option adjusted spread. I do not believe, however, that using a training set of backward-looking financial ratios will result in optimal credit risk ranking for any particular point in time. Regardless of the algorithm being used, any machine learning technique depends on the depth and usefulness of information in its training set. As such, I will be using a database that I am currently building which has 40 different financial metrics that captures information about a company’s capital structure, growth, profitability, how effective it is in using new “money” (newly issued debt/equity or free cash flow) to generate returns from new projects, leverage, debt service ability, cost of capital, and assuming the size (total assets), income (EBIT), and cost of capital (debt & equity) remain constant over the next twelve months how much new money can be created before it’s cost of capital exceeds its return on capital. I’m gathering this information by using Factset, Bloomberg, & Compustat to pull information on public investment grade and high yield corporate debt issuers from 12/31/1996 – 10/31/2016 on a monthly basis.

Regardless of how effective this technique will be, fundamental analysis remains critically important for any high yield bond strategy. Constantly shifting landscape, accounting rule changes, economic and competitive pressures, and structural adjustments in the market preclude a formulaic approach to bond analysis. However, supervised machine learning can be effectively deployed when the focal point of the process is avoiding losses, rather than hitting big winners. High yield fixed income team resources can thus focus on incorporating private issuer fundamental information into the training set and assess whether new or possibly more forward looking measures can be added to the training set to increase the effectiveness of the issuer credit risk ranking. In a world where the amount of information and statistical learning techniques used to analyze it are growing exponentially it would be wise for any asset manager to focus on how it can use and analyze this information in order to increase productivity and distinguish itself from the rest of the pack.